Design Framework for the ACT® Enhancements

March 2025

Authors

Introduction

Jeff Allen, PhD; Jay Thomas; Stacy Dreyer; Scott Johanningmeier; and Dana Murano, PhD

Chapter 1

Jay Thomas, Stacy Dreyer, and Scott Johanningmeier

Chapter 2

Jay Thomas, Stacy Dreyer, and Scott Johanningmeier

Chapter 3

Jay Thomas, Stacy Dreyer, and Scott Johanningmeier

Chapter 4

Jeff Allen, PhD; Ty Cruce, PhD; Xin Li, PhD; Edgar Sanchez, PhD

Acknowledgments

While the team of authors was responsible for creating this report that synthesizes all findings related to the ACT Enhancements, there are countless team members across Measurement Research and Development who contributed to the original research and content presented in this report. We would like to acknowledge the contributions from individuals on the psychometrics team (Dongmei Li, Shalini Kapoor, Ann Arthur, Chi-Yu Huang, YoungWoo Cho, Chen Qiu, and Hongling Wang), the ELA team (in particular, Kelly Smith), the math team, and the science team (in particular, Greg Hanson). We would also like to thank the many other teams across ACT that facilitated and supported the ACT Enhancements.

Version	Date	Description
2025.1	March 2025	Original version

Table of Contents

Figures

Tables

Foreword

ACT has a long history helping learners achieve college and career success and providing the opportunity for each individual to show what they know. As the world of education continually evolves, so does the ACT, striving to meet the varied and changing needs of learners, educators, and other stakeholders. The enhancements outlined here reflect our commitment to maintaining the validity and predictive value of the ACT while making the testing process more flexible, accessible, and supportive of each learner's journey.

Beginning in April 2025, U.S. learners opting to take an online version of the ACT during a national Saturday event will experience a shorter format and the option to choose whether to take the science section. This will result in a Composite score that reflects learners' achievements in English, mathematics, and reading, while science and writing scores will be reported separately for those who elect to take those sections. This change offers greater flexibility for learners while providing our stakeholders in secondary and higher education with a test that continues to meet their needs. Starting in September 2025, these enhancements will extend to the paper test format for Saturday testing in the United States and to all ACT international testing. From September 2025 onward, all Composite scores will reflect English, mathematics, and reading.

Several additional important changes are being introduced: The overall number of questions has been reduced, the test length has been shortened—allowing students to receive their college-reportable Composite score after 125 minutes of testing—and more time has been allotted per question. These changes give test takers an improved opportunity to show all they know and have learned. These updates ensure that K–12 systems and states have the information they need about their students' college and career readiness and that colleges can continue to make informed decisions about admission, scholarships, placement, and student support. Research thus far supports the claim that these changes result in valid, reliable, and comparable scores that stakeholders can continue to use for these purposes. The updates also make learners' results even more relevant as they take the next steps after high school.

With these enhancements, some aspects of the exam remain unchanged. The ACT scale will remain 1–36, with no modifications to the ACT Benchmarks or state-specific achievement standards. Composite scores from tests taken before September 2025 will not change. Existing research shows that the ability to predict success in first-year credit-bearing college courses remains consistent, and we will continue to conduct and disseminate predictive validity research to further support this claim. Both the paper and online tests will continue to be available so that we can continue to support test takers by giving them the option to choose the format they feel more comfortable with.

We are confident that these enhancements will better support learners in their paths to college and career success, ensuring that their experience with the ACT is beneficial and reflects their academic capabilities.

Adrienne Dieball

Senior Vice President, Measurement Research and Development

Introducing the ACT Test Enhancements

ACT has been dedicated to improving college and career readiness for all students since its inception in 1959. The ACT® test has provided students, educators, and other educational stakeholders with unparalleled measures of college and career readiness. An optional writing test was introduced in 2005, but the general structure of the ACT multiple-choice tests—including the number of test questions and the time allowed per test section—remained the same between 1989 and 2025. ACT has, however, continually modified the test content to ensure that the test continues to measure the knowledge and skills needed for college and career success. In this report, we document the latest enhancements to the ACT, which include changes to both the general test structure and the test content.

In this introductory section, we provide a summary of the enhancements. We begin by describing the core features of the legacy ACT test that provided the foundation for the enhanced ACT test. We then describe the enhancements made to the ACT test, with a special focus on what is changing.

The ACT enhancements are the result of a careful and deliberate process of collecting evidence and stakeholder feedback. This introductory section highlights the enhancements made to the ACT test, and the remaining chapters provide more details on the process of collecting evidence and feedback. To help readers navigate this report, we conclude this introductory section with an overview of each chapter.

Maintaining the Foundations of the ACT Test

Measurement Philosophy

Underlying the ACT test is the belief that students' preparation for college and the workplace is best assessed by measuring, as directly as possible, the skills learned in high school that are required for success in first-year college courses. The required academic skills can be assessed most directly by reproducing the complexity of the work students complete in high school and are expected to complete in college. Therefore, the ACT is designed to determine how skillfully students solve problems, grasp implied meanings, draw inferences, evaluate ideas, and make judgments in subject-matter areas important to success in college. These philosophical foundations of the ACT test are not changing with the enhancements.

The ACT is oriented toward the general content areas of college and high school instructional programs. The ACT emphasizes students' academic preparedness by directly assessing the content domains students must master to achieve college and career readiness—including English, math, reading, science, and writing. The design of the ACT test is informed by several factors, including input from subject matter experts; academic research on skill targets, sequencing of skills, and grade placement; evidence of student understanding collected from the ACT test; the ACT National Curriculum Survey; and a survey of standards frameworks including the ACT College and Career Readiness Standards, common state standards (such as those based on the Common Core State Standards for Math and ELA [National Governors Association for Best Practices, & Council of Chief State School officers, 2010] and the Next

Generation Science Standards [NGSS Lead States, 2010] and NRC Framework for K–12 Science Education [National Research Council, 2012] for science), and other assessment frameworks such as NAEP. The research, standards frameworks, and feedback mechanisms informing ACT test design are not changing with the enhancements.

ACT Score Scales

ACT scores are reported on scales of 1–36. ACT maintains the ACT section test score scales through rigorous equating procedures to ensure that the scores have the same meaning over time. Interpretations of ACT scores have been developed and validated through numerous studies conducted by ACT and other organizations, resulting in important resources for using ACT scores, such as the ACT College Readiness Benchmarks and ACT's College and Career Readiness Standards. Some stakeholders, such as institutions of higher education and scholarship agencies, rely on the stability of the ACT scales and have developed their own interpretations and uses. The ACT score scales are not changing with the enhancements, ensuring continuity in the interpretation and use of ACT scores.

Intended Uses of ACT Scores

The ACT is designed to measure high school academic achievement and college readiness. Interpretations of ACT scores support their use for informing college admissions decisions, awarding college scholarships, placing students into programs and courses, identifying students in need of academic support, and measuring academic achievement at the school and district levels for accountability systems. The validity evidence supporting the interpretations and uses of ACT scores is documented in the *ACT Technical Manual* (ACT, 2024). The intended uses of ACT scores are not changing with the enhancements to the ACT.

Commitment to Fairness

ACT test materials are designed and developed with a focus on fairness to all students, offering them multiple ways to show their knowledge and ensuring accessible formats, particularly for students with visual and physical differences that impact their ability to engage with a computer-based or paper-and-pencil exam. ACT applies the principles of universal design for assessment beginning with item and passage development; as a result, the assessment better allows all students to show what they know and can do while minimizing construct irrelevant variance.

A wide range of highly trained ACT staff and external reviewers representing various backgrounds review every test item and passage multiple times for accuracy and fairness. All materials are reviewed for fair portrayal and balanced representation of groups within society. In addition to conducting these reviews, we perform statistical analyses to ensure test content does not inadvertently disadvantage any group.

The enhanced ACT will maintain the highest standards for fairness, and we will continue to put every test item through multiple reviews by diverse groups of experts to ensure fairness to all learners. The accessibility supports that are available for the legacy ACT, including testing accommodations, will also be available for the enhanced ACT. Further, the ACT fee waiver program will remain available for students needing financial assistance.

Modes of Testing

The ACT test can be taken online or on paper. ACT has provided online testing since 2015 for state and district testing and since 2018 for international testing. Online testing for national testing was launched in February 2024; ACT will continue to expand online testing as a choice, but not a requirement. While the ACT can be taken online, it is not a computer-adaptive test. Test mode options are not changing with the enhancements to the ACT.

The ACT can be taken during the school day through state and district testing programs or on the weekend through national testing programs. The ACT can be taken in the United States or at many other testing centers in other countries around the world. These testing programs are not changing with the enhancements to the ACT.

The Enhanced ACT Test

The ACT test is evolving to better meet the needs of students. The enhancements are designed to help students perform their best on test day, have the flexibility to determine how they will test, demonstrate their true capabilities and potential, and tailor their testing experience to fit their future goals and highlight their strengths. We now describe the enhanced ACT test, highlighting how it is different from the legacy ACT test.

Test Length and Timing

Like the legacy test, the enhanced test includes multiple-choice tests in English, math, reading, and science. The enhanced test is shorter than the legacy test and allows more time per test item. The enhanced test also includes embedded field-test items, which are not counted in students' scores. A comparison of the legacy and enhanced test specifications for English, math, reading, and science is provided in the table below.

Enhanced and Legacy ACT Test Specifications

Test	Section	Scored items	Field-test items	Total items	Minutes allowed	Seconds per item
Enhanced ACT	English	40	10	50	35	42
	Math	41	4	45	50	67
	Reading	27	9	36	40	67
	Science	34	6	40	40	60
Total without science		108	23	131	125	57
Total with science		142	29	171	165	58
Legacy ACT	English	75	0	75	45	36
	Math	60	0	60	60	60
	Reading	40	0	40	35	53
	Science	40	0	40	35	53
Total with science		215	0	215	175	49

Relative to the legacy ACT, the enhanced ACT offers an average of nine seconds more per item. (This difference varies across test sections.) Total testing time is reduced, with Composite scores requiring 125 minutes of testing for the enhanced ACT rather than the 175 minutes required for the legacy ACT. Most of this reduction is because of the science test being optional and not included in the Composite score for the enhanced ACT, two changes that are discussed later in this chapter.

For the legacy ACT, national testing includes stand-alone field-test items (often referred to as the 5th test), which add 20 minutes to the total testing time. Therefore, for students taking all four section tests through national testing, the entire testing experience for the enhanced ACT is 30 minutes shorter than it is for the legacy test. For students taking all four section tests through state and district testing, the entire testing experience for the enhanced ACT is 10 minutes shorter than it is for the legacy test. The ACT writing test is not changing: The design, timing, and scoring procedures are the same for both the legacy and enhanced tests.

Embedded Field-Test Items

The use of embedded field-test items means that students who take the enhanced ACT will respond to test items that do not count toward their scores but will be used to build future ACT test forms.

The use of embedded field-test items eliminates the need for the 5th test (stand-alone field-test items) that is administered with the legacy ACT during national test administrations. Examinees who participated in the legacy ACT's 5th test were aware that the items did not count toward their scores, so some students may not have been motivated to put forth their best effort. For the enhanced ACT, field-test items are placed strategically throughout the section tests, blending seamlessly with operational items and preserving the natural flow of the ACT assessment. With this approach, examinees engage with field-test items as they would with items that count toward their scores, leading to more authentic responses and a more accurate evaluation of item performance.

Embedded field-testing aligns the testing experience for school-day (state or district) and national test administrations and creates a more inclusive sample for testing new items. Embedded field-testing also allows data from each student to contribute to the pretesting pools of multiple content areas. Over time, we anticipate that this change will also enable faster score reporting.

Updates to Test Content

In addition to the shorter test length and use of embedded field-testing, other aspects of ACT test design are changing to improve alignment with assessment industry best practices. For all four section tests, specifications for the percentage of items representing different reporting categories have been updated (refer to the following table). While the reporting categories are the same for the legacy and enhanced ACT, these updated reporting category specifications ensure that the ACT will continue to measure the knowledge and skills considered essential for college and career readiness while aligning to state content standards.

Operational Item Reporting Category Alignment for the Legacy and Enhanced ACT

Section	Reporting category	Legacy ACT %	Enhanced ACT %
English	Production of Writing	29–32	38–43
	Knowledge of Language	15–17	18–23
	Conventions of Standard English	52–55	38–43
Math	Preparing for Higher Math	57–60	80
	Number & Quantity	8–12	10–12
	Algebra	12–15	17–20
	Functions	12–15	17–20
	Geometry	12–15	17–20
	Statistics & Probability	8–12	12–15
	Integrating Essential Skills	40–43	20
	Modeling	≥20	≥20
Reading	Key Ideas & Details	53–60	44–52
	Craft & Structure	25–30	26–33
	Integration of Knowledge & Ideas	15–23	19–26
Science	Interpretation of Data	40–50	38–50
	Scientific Investigation	20–30	18–32
	Evaluation of Models, Inferences, and Experimental Results	25–35	24–38

Note: The Preparing for Higher Math reporting category includes Number & Quantity, Algebra, Functions, Geometry, and Statistics & Probability.

The enhanced ACT also has section-specific improvements to the test content.

English

The enhanced English test will have at least one argumentative essay on each test form, with items designed to target important writing skills related to argumentation. The test features a reduction in the percentage of items in the Conventions of Standard English reporting category and an increase in the percentage of Production of Writing and Knowledge of Language items. Stems have also been added to each English item. The stem is the question or the problem and any instructions that go with it. For the legacy ACT, the examinee receives initial instructions and then is presented with items that have less context and refer to a portion of the passage that is underlined and numbered. Adding stems to these items provides all students with more direction. Examinees will no longer need to refer to generalized instructions but will have everything they need in the items themselves. In addition, this section will now feature a mix of short and long essays, whereas the legacy English test included only long essays.

Math

For the enhanced math test, the number of response options has been reduced from five to four. This reduction brings the ACT test into alignment with other assessments that examinees might be familiar with, reduces the time spent reading and thinking through answer choices, and potentially improves the measurement quality of the math section. Additionally, the number and percentage of math items that are set in a real-world context (and therefore require the student to spend more time reading and processing the contextual information) will be reduced. The enhanced math test also has a smaller percentage of items in the Integrating Essential Skills reporting category and the advanced topics subcategory (higher grade level items).

Reading

The enhanced reading test has a smaller percentage of items that align to the Key Ideas and Details reporting category and a slightly larger percentage of items that align to the Integration of Knowledge and Ideas and Craft and Structure reporting categories. The enhanced reading test will continue to use complex literary and informational texts as a basis for items, but passage length will vary: Two passages on each test form will be approximately 750 standard words, and one will be approximately 650 standard words.

Science

The enhanced science test has stronger alignment with national science standards, with at least one passage per form now addressing an engineering and design topic. The enhanced science test also has more items that require integrating scientific background knowledge with the passage.

Optional Science Test

ACT remains committed to supporting science and STEM education. We will continue to offer a science assessment for all learners who wish to take it and will continue to work with our stakeholders to ensure that our science assessment aligns with their needs and expectations.

Beginning with ACT tests taken online on national test dates in April 2025, the ACT science test will be optional. Beginning in the fall of 2025, this change will be applied to all national ACT test administrations. Students who take the science section will receive a STEM score (if they also attempted the math test). This mirrors the approach that is used for the optional writing test: Only students who take the writing section receive an ELA score (if they also attempted the English and reading tests). The change will be applied to state and district testing in the spring of 2026; in this case, however, the decision of whether science is optional will be up to the state or district and not the individual students.

Who Should Take the Science Test?

Science and scientific reasoning remain crucial, and ACT is still the only college admissions test with a science test. While not all students will be STEM majors, nearly all colleges require science courses as part of their general education requirements. The ACT science and STEM scores can help students assess their readiness for STEM programs and identify strengths and weaknesses that might affect success in first-year courses.

While the science test can provide important information for all students, it is most important for those planning to major in a science or STEM field. Students who are retaking the ACT and who believe that their initial ACT science score did not reflect their true ability should also consider retaking the science test.

We encourage states and districts that provide the ACT test during the school day to still administer the science test. Doing so ensures that all students have the chance to demonstrate readiness in science. Assessing science is also important for state and federal accountability systems and provides insights for the evaluation and improvement of science instruction. The ACT is the only college admissions test that has met expectations for federal peer review for science. This means that states can use one assessment—the ACT—for ESSA accountability rather than a college admissions test and a separate custom-built science test.

For national test administrations, students who decide to take the ACT test with science or writing (or both) will remain in the testing room while the other examinees are dismissed after the reading test. There will be a short break, after which they will return to their seats and take either the science test (for students taking science or science and writing) or the writing test (for students taking writing but not science). Both science and writing will take 40 minutes and will be given in the same room. Following the first 40-minute optional session, anyone who opted to take only science (or writing) will then be done and will be dismissed. Examinees taking both science and writing will have a short break before beginning the writing test.

For school-day test administrations that include science and writing, states and districts have the option to administer the science and writing tests on a separate day from the three core sections of English, mathematics, and reading.

New Composite Score Definition

Beginning with ACT tests taken online on national test dates in April 2025, reported Composite scores will be based on English, math, and reading scores. Science scores will no longer be included in the Composite score calculation. Beginning in the fall of 2025, this change will be applied to all ACT test administrations. Because the science test is optional, this new Composite score definition provides a consistent measure of overall test performance for all examinees.

It is important to note that the Composite score definition applies only to ACT score reporting. Postsecondary institutions and other users of ACT scores can still define their own ACT total scores for different purposes. For example, ACT sum scores, ACT total scores that weight the section tests differently, and ACT total scores that include science are all examples of ACT total scores that may be used.

ACT Superscores

The ACT Superscore is calculated like a composite score but is based on students' best ACT section test scores across multiple test administrations. Starting in April 2025, students who take the national online ACT test will have their ACT Superscore calculated using the new method incorporating English, math, and reading. For all other testing, this change will happen in September 2025. ACT reports will still show students' highest section test scores, but the Superscore will be based on the new Composite score definition. To calculate the Superscore, ACT will use a student's best performance from each test section, regardless of which version of the test that section score came from. We are confident that combining scores from both the legacy and enhanced versions of the ACT will result in valid and predictive Superscores.

Navigating This Report

This report describes the process we undertook to develop and validate the enhanced ACT. We begin by describing the changes we made to the test content and the processes by which we implemented these design decisions. Next, we describe how we shared the overall scope of the enhancements, including the initial blueprints, with external expert panels, resulting in feedback on and iterative updates to the blueprints and proposed changes. With the revised blueprints in place, we then focused on collecting data from multiple sources and stakeholders to support decisions about the use of new and altered item types. We describe evidence collected from multiple cognitive laboratory studies conducted with students; these studies support both the cognitive processes associated with item and passage types continuing from the legacy ACT and the validity of the item- and passage-level enhancements in particular. Finally, we present statistical evidence that demonstrates continued reliability and validity of the enhanced ACT. The goal of this document is to provide transparency to stakeholders about the decisions we made throughout the design process, as well as to document supporting evidence collected for each change. Below, we describe the contents of each chapter and how each contributes to the overall validity argument for the enhanced ACT.

Chapter 1: Identifying Potential Areas for Enhancements of the ACT and the Initial Blueprints

Chapter 1 introduces the potential enhancements we considered and the rationale behind each. First, we describe the process we used to examine potential areas for enhancement. These areas included testing time and speededness, student choice and flexibility, modernization of the test, and accessibility and equity. Our goal was to introduce these enhancements while maintaining the core uses and tenets that are foundational to the ACT. To ensure we succeeded at this, we prioritized maintaining four things: continuity of scores and interpretations related to college readiness, admission, and scholarship qualifications; continued validity and reliability of scores; continued alignment to state standards; and comparability of scores. To steer this work, we relied on principles of assessment design science, which include identifying problems and intended outcomes, designing assessment requirements and constraints, and integrating tacit knowledge with explicit knowledge to develop artifacts.

This chapter describes each of the decisions made about each potential enhancement. It also includes a detailed overview of how the initial proposed changes would impact each test section and concludes with initial blueprints that reflect the desired enhancements and their implementation in each test section.

Chapter 2: Panels, Initial Studies, and Revisions to Blueprints

Chapter 2 describes the processes of collecting feedback on the initial enhanced blueprints and making subsequent revisions based on the feedback. For each subject area, we brought together a group of expert panelists that had both domain-specific expertise and knowledge in college and career readiness and alignment. The panelists evaluated the proposed changes to the ACT from the perspectives of state and district stakeholders, postsecondary stakeholders, and parent and student stakeholders. For each subject area, they considered the proposed changes at the passage, item, blueprint, and matrix sampling levels. All proposed changes were also reviewed from the perspectives of universal design for assessment and equity by design. Overall, the panelists generally supported the proposed changes for each section. Minor revisions were made to blueprints based on panelist feedback. These revised blueprints are presented at the end of Chapter 2 and were used for the June 2024 linking study (Li et al., 2025) detailed in Chapter 4.

Chapter 3: Validity Evidence Based on Test Content and Response Processes

Chapter 3 documents content-based validity evidence for the enhanced ACT. Specifically, we focused on validity evidence based on test content and evidence based on response processes that support the ACT's core claims of being predictive of college and career readiness and indicative of mastery of skills in common high school standards. Evidence in these areas supports the alignment of test items to the targeted domains being assessed and ensures that the content is representative of those intended domains. To demonstrate this, we describe how the test content for each section aligns to state content standards, including those based on the Common Core State Standards for ELA and math and those based on the NRC Framework for K–12 Science Education and the Next Generation Science Standards for science. We also present evidence of alignment to ESSA peer review requirements and alignment to findings from the National Curriculum Survey. Each of these sources of alignment evidence supports the claim that the ACT measures common high school standards and skills that are predictive of college and career readiness.

This chapter also summarizes several research studies that provide evidence based on response processes, including cognitive laboratory studies, eye tracking studies, and survey research completed with students after they completed test items. These studies provide insights into the cognitive processes test takers use when answering questions and interacting with test materials. The findings from these studies support the inclusion of various types of passages and items in the enhanced ACT, including visual quantitative information (VQI) passages and argumentative essays; they also support reducing the number of answer choices on the math test from five to four. During these studies, students consistently demonstrated the expected response processes and patterns in each newly introduced item component. The studies also provide evidence that students use the targeted skills on item and passage types from the legacy ACT that will continue to be used on the enhanced ACT. Overall, the chapter provides a comprehensive overview of the various sources of content-based validity evidence that support the intended uses of the enhanced ACT.

Chapter 4: Additional Psychometric and Validity Evidence

Chapter 4 summarizes the psychometric and statistical validity evidence collected to date for the enhanced ACT. We first present results from the June 2024 linking study (Li et al., 2025) , which provides evidence of construct equivalence and an overview of the enhanced ACT's psychometric properties. Next, we present results from the October 2024 Mode Comparability study (in preparation), which evaluated test and item comparability for tests administered online and those administered on paper. This study also includes the results of confirmatory factor analyses and tests for differential item functioning. Taken together, the results from these studies support the enhancements from a psychometric standpoint and provide evidence that the enhanced ACT measures the same constructs and reports section test scores on the same scales as the legacy ACT, with appropriate levels of reliability needed to support validity claims.

The next portion of the chapter focuses on validity evidence based on relationships with other variables. Much of the evidence reviewed in Chapter 4 is expanded upon in our white paper, Initial Evidence Supporting Interpretations of Scores from the Enhanced ACT Test (Allen & Cruce, 2025). We present concurrent validity evidence, which includes relationships between scores on the enhanced ACT, scores on the legacy ACT, and high school coursework and grades, as well as interrelationships between enhanced ACT scores across subject areas. We also present results from a simulation study that examined how well a shortened version of the ACT predicted first-year GPA and college degree attainment.

Last, we present evidence of Composite score comparability between the legacy ACT and the enhanced ACT. Specifically, we examine differences between Composite score interpretations for the legacy ACT and enhanced ACT, with and without the inclusion of science in the Composite score. Overall, the evidence presented in this chapter supports the interpretation of scores from the enhanced ACT as measures of high school academic achievement and college readiness, and the evidence also indicates that scores from the enhanced ACT can be used interchangeably with scores from the legacy ACT for most purposes. We conclude that scores from the enhanced ACT can be used for informing college admissions decisions, awarding college scholarships, placing students into programs and courses, identifying students in need of academic support, and measuring academic achievement at the school, district, and state levels for accountability systems.

Chapter 1: Identifying Potential Areas for Revision of the ACT and Drafting the Initial Blueprints

1.1 Introduction

During the 2022–2023 school year, ACT started exploring potential changes to the ACT® test as part of the periodic review process. Using a principled assessment design (PAD) approach is critical when designing a new test or revising an existing one. While various approaches to PAD exist, including evidence-centered design (Mislevy et al., 2003; Mislevy & Riconscente, 2006), interpretation and use argument (Kane, 1992, 2013), theory of action (Bennett, 2010), and assessment design science (Langenfeld & Thomas, 2016 Langenfeld et al., 2019), a common goal is to link the assessment's claims about the knowledge, skills, and abilities that test takers possess to the test, individual items, and design decisions in a defensible and logical way. We made this goal the center of our evaluation efforts and followed the assessment design science framework to guide our process.

The assessment design science approach, proposed by Thomas and Langenfeld (2017a, 2017b, 2018) based on earlier work (Johannesson & Perjons, 2014; van Aken & Romme, 2012), suggests repeating the following five steps as often as necessary.

1. Identify the problem and the intended outcomes.
2. Define the assessment requirements and constraints.
3. Integrate tacit knowledge (team-based experiential learning) with explicit knowledge (research-based literature) to develop artifacts.
4. Gather data in the artifacts and evaluate.
5. Revise the artifacts and improve the solution.

This document describes how we implemented this five-step approach to develop the enhanced ACT test. Although the process appears to be linear, it is actually cyclical. Each time you gather data and evaluate, you may revise the artifacts, which will require additional data collection to determine whether the change met the desired goal. Chapter 1 focuses primarily on Steps 1 through 3, addressing the initial gathering of the data we used to create draft blueprints for the enhanced ACT. Chapters 2–4 look at several iterations of Steps 4 and 5. Chapter 2 addresses expert panels that were brought together to evaluate proposed revisions and artifacts. Chapter 3 addresses data related to content and cognitive processes gathered through cognitive labs and interviews. Chapter 4 documents the data collection process and presents the psychometric validity evidence. All of these efforts contributed to the final enhanced ACT blueprint, technical specifications, and validity claims.

1.2 The Need for Periodic Review and Revision

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

The legacy ACT last underwent major revision in 1989. Since then, ACT has made changes such as revising the reporting categories, adjusting the percentage of items in each category, reducing the number of passages while maintaining the number of items, and introducing new passage types, but these changes have been relatively minor. Some of the changes were made to better align the test to state standards for peer review; others were made to address perceived issues related to speededness; still others were based on feedback from the ACT® National Curriculum Survey® of K–12 and postsecondary educators. To the student, the ACT likely appeared largely unchanged, with 75 English items, 60 math items, 40 reading items, and 40 science items presented with the same time allowed. These changes maintained consistency for other stakeholders as well, as scores continued to be comparable through the same 1–36 scale score for each test section and the same 1–36 Composite scale score, which maintained standard errors of measurement of approximately 2 points and 1 point, respectively.

Given all the uses of the ACT, there were constraints on how the legacy ACT could be changed.

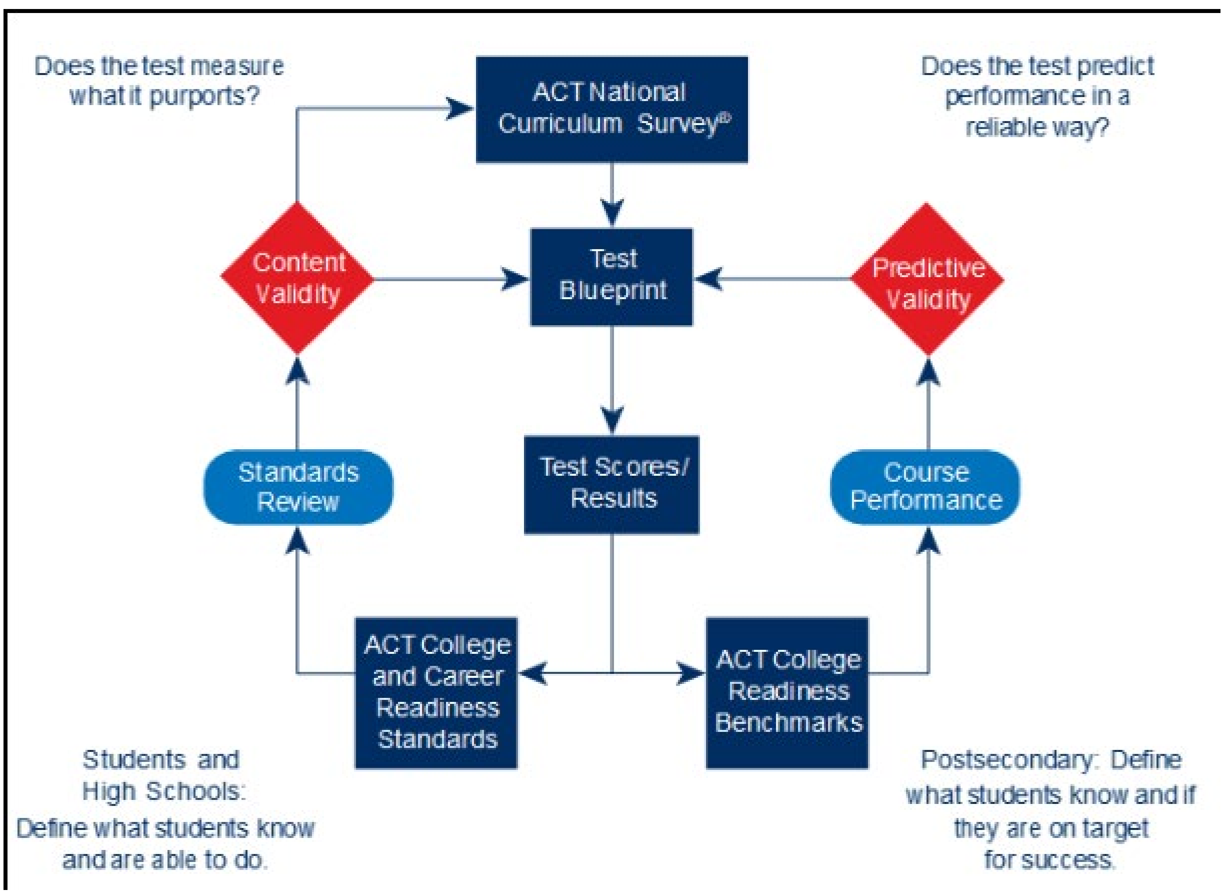
The ACT is used as an accountability measure as part of the Every Student Succeeds Act (ESSA) and must pass a rigorous peer review process. Since the ACT has received the "meets" or "substantially meets" rating for multiple states, it is important that an updated ACT test maintain both strong alignment to state standards and content coverage and predictive power for colleges and universities. Additionally, research from district, state, and postsecondary stakeholders indicated that there was a need for comparability of longitudinal data, which limited the possible changes in scores and reporting of the legacy ACT.

1.2.1 The ACT National Curriculum Survey

Part of any periodic review has traditionally been the ACT National Curriculum Survey (NCS). The NCS plays an essential role in delivering on ACT's longstanding commitment to ensuring that our assessments remain valid and relevant. Conducted by ACT every several years, the NCS is a unique nationwide survey of educational practices and college and career readiness expectations. ACT surveys thousands of K–12 teachers, administrators, and college instructors to determine which skills and knowledge in English and writing, mathematics, reading, and science are currently taught at each grade level and which skills and knowledge are currently considered essential for college and career readiness. The survey also includes a sample of workforce supervisors and employees to understand how the priority placed on skills in the workplace compares with the priority placed on those skills by educators. ACT conducts this survey to ensure that its assessments measure the knowledge and skills that workforce supervisors and instructors of credit-bearing first-year college courses identify as important for success in college and the workplace. ACT makes the results of each survey public, believing that our data can help education and workforce stakeholders make more informed decisions about the skills needed to be successful in postsecondary education and the workplace.

ACT uses the results of the NCS to guide the development of ACT assessment solutions, including the ACT test, PreACT® suite of assessments, and ACT® WorkKeys® Assessments. As such, the survey is a crucial step in the process ACT uses to build and regularly update a valid suite of assessments that is empirically aligned to college and career readiness standards. The survey results directly inform the test blueprint for the ACT test, which specifies the balance of test questions assessing different knowledge and skills within the domain of the test (refer to diagram below). Results from the NCS were used first to create and now to continually validate ACT's College and Career Readiness Standards as well as the College Readiness Benchmarks.

Figure 1.1. Dual Validity Loop



As Figure 1.1 indicates, the development of ACT's assessments begins with a dual validity loop—one for content validity and one for predictive validity. Included in the content validity loop is the ACT National Curriculum Survey. The predictive validity loop uses data from student college and career outcomes to identify which knowledge and skills are the strongest predictors of performance. Together, the content and predictive validity loops are used to establish the test blueprints. This process ensures that our assessments always measure the knowledge and skills that are being taught in schools around the country and that indicate college and career readiness.

Using NCS data is but one of numerous actions ACT takes to ensure that our assessments provide up-to-date information that enables students to be fully ready to embark on rewarding college and career journeys. Other actions and considerations will be addressed in the sections to follow as each primary consideration for updating the ACT is explored.

Section 1.1 will address the primary considerations we included in this most recent review and revision of the ACT, such as reducing overall testing time, addressing perceived speededness, maintaining the reliability and meaning of the score for both college and career readiness, and ensuring continued alignment to state standards for the peer review process.

1.3 Primary Considerations and Proposed Changes

1.3.1 Reducing Overall Testing Time

Across the country, parents, teachers, administrators, and politicians are concerned that students spend too much time testing. In addition, across the various markets where the ACT is administered, there are variations in how much time examinees spend on the ACT on test day. Students have stated that the total testing time of approximately three hours is stressful, as they rarely sit for three-hour examinations in high school.

As a drop-in large-scale college readiness assessment, the ACT can assess more material in less time than traditional end-of-course exams. Historically, the total testing time for national ACT test takers, including the mandatory 5th test pretest, was 195 minutes without writing and 235 minutes with writing. One of our driving goals as we explored updates to the ACT was to reduce this total time in order to reduce the perceived stress caused by the test and the time commitment for students. This reduction in time had to be balanced against the need to make valid and reliable claims about what students know and can do relative to both common state standards (such as the Common Core State Standards and the Next Generation Science Standards) and college and career readiness.

In an effort to respond to stakeholders' requests for a shorter test and to alleviate some of the complexities across market segments, ACT considered a number of potential solutions. The first was reducing the number of items in each section. The number of items that could be removed from each section was examined based on coverage of common state standards and the minimum number of items required to maintain psychometric properties of reliability. Since ACT was also exploring the use of embedded field-testing for all test takers as part of the modernization process (refer to 1.1.10), we knew that adding field-test items might counteract some of the reduction in time gained through dropping operational items. Therefore, a second consideration in reducing overall testing time was the need to balance adding field-test items with maintaining the necessary number of operational items. Both considerations contributed to the initial decisions related to the new blueprints.

Finally, we also determined that the science section could become optional, further reducing the overall testing time for stakeholders who choose that option (1.1.8).

Table 1.1. Testing Times for ACT Legacy and Initial Proposed Time for ACT Enhanced

ACT section	ACT Legacy national (min)	ACT Legacy state (min)	ACT Enhanced (min)
English	45	45	35
Math	60	60	50
Reading	35	35	40
Science	35	35	45
5th test (pretest)	20	N/A	N/A
Total	195	175	170

1.3.2 Addressing Perceived Issues of Speededness

Feedback from examinees has indicated a preference for more time to complete the ACT. Some internal analysis of performance data has indicated that, for certain conditions, items on the ACT might exhibit unintended speededness (the extent to which a test's time limit alters a test taker's performance). Parts of the legacy ACT appear to exhibit speededness, particularly the reading and science sections. Focus group data found that students considered the test speeded and had to rush to finish. Since speed or fluency is not part of the construct for any section of the battery, speededness represents construct irrelevant variance that should be minimized.

ACT's recent reading rate study (Arthur & Thomas, 2023) found that student reading rates were comparable to those indicated in prior research, suggesting that the estimates of time needed to read ACT legacy passages (Carver, 1992; Gallo, 1972; Taylor, 1965) are still valid. Results from the study showed that median reading rates for most passages were between 225 and 250 standard words per minute. During the study, students reported needing to rush on the last timed reading passage because they had spent too much time on the questions on the earlier passage. Additionally, the same students exhibited many nonsolution behaviors and poor metacognition on math items (Thomas & Victoria, 2023), leading many of them to not finish the test. While the standard time per item on the math section was 60 seconds per item, more than half the students spent over 180 seconds on an unusual, difficult math item rather than answering and moving on. Similar behaviors were seen in the 2017 cognitive labs (Thomas et al., 2025). Data about the science section (Thomas & Victoria, 2023) also found that the majority of the students' time was spent on cognitive processes related to decoding and answering questions and not on reading the passage.

Based on these findings, speededness concerns appear to occur at the level of answering items rather than reading passages. Consequently, we decided that solutions should focus on time per item rather than time per passage, and we used the median reading rate of 225–250 standard words per minute as an approximation for the time needed to read and comprehend a passage.

Related to speededness is student self-efficacy. The ACT should allow students to feel that they are able to show what they know and can do. Efforts to minimize speededness, especially on the reading and science sections, should better allow students to demonstrate what they learned in high school and how ready they are for college and career.

1.3.3 Ensuring Continued Use and Continuity of Scores Related to College Readiness, Admission, and Scholarship Qualifications

The ACT section tests are used to evaluate claims about college readiness in specific first-year courses. Students who meet the ACT College Readiness Benchmark score on a section test have a 50% likelihood of earning a B or better and a 75% likelihood of earning a C or better in the corresponding first-year courses listed below. The combined ELA score predicts performance in courses typically taken by students majoring in ELA-related disciplines during their first year. The combined STEM score predicts performance in the STEM (science, technology, engineering, and mathematics) courses taken by students majoring in STEM disciplines during their first year (refer to Table 1.2).

Table 1.2. Reference Courses for ACT College Readiness Benchmarks

Section	Benchmark score	Corresponding first-year courses
English	18	English Composition I
Math	22	College algebra
Reading	22	Introductory courses in American history, other history, psychology, sociology, political science, economics
Science	23	Biology
Combined ELA	20	English Composition I, American history, other history, psychology, sociology, political science, economics
Combined STEM	26	Calculus, chemistry, biology, physics, engineering

When making admissions decisions, colleges can use the ACT Composite score, individual section test scores, and combined ELA or STEM scores to compare students. Since high schools have different expectations, grading scales, course offerings, etc., a standardized measure of student achievement is needed to inform admissions decisions. Additionally, since postpandemic research indicates that grade inflation has continued to increase (Sanchez, 2023), an equitable measure of skills is needed. These scores can add incremental validity over other measures to help colleges best select students who fit a particular school, subcollege (like a college of engineering), or major.

The ACT is also used as a factor in determining scholarship eligibility at the state level (programs like Louisiana's TOPS) and the postsecondary level. Many scholarships at the local or corporate levels also use ACT scores. Many of these have shown great return on investment for states (Louisiana Board of Regents, 2021), particularly for students earning an ACT score between 19 and 26, which is below the typical ACT scholarship score for many colleges and universities.

1.3.4 Ensuring Valid Score Interpretation and Claims

One key aspect of any assessment is explaining what claims can be made about what test takers know and can do. A key component of assessment validity is having evidence that supports using scores to predict outcomes such as the likelihood of success in first-year college courses, one of the primary claims and uses of ACT scores. Surveys of postsecondary institutions and K–12 stakeholders indicated a strong desire to keep the ACT on its traditional 1–36 scale with the SEM values of close to 2 for each test section and a SEM of 1 overall for the Composite score. One concern for the legacy ACT was that there was not much differentiation among students in the lower score bands, which limited what claims could be supported for low-scoring students. Consequently, ACT conducted a study to determine whether the distribution of the 1–36 scale should or could be altered. After that study, we decided to keep the legacy ACT's distribution of the 1–36 scale, preferring continuity over adopting a new scale.

1.3.5 Maintaining Score Reliability

Since reliability depends on the number of items on an assessment, a balance must be maintained between reducing the length of the test and maintaining sufficient reliability. An assessment must reliably generate scores as part of any validity argument, theory of action, or other set of claims about what test takers know and can do. Therefore, another priority for each test section was maintaining enough items to ensure that the enhanced ACT would continue to exceed a coefficient alpha of .80 for all forms. For each section, we used preliminary models to determine the minimum number of items necessary to maintain adequate reliability. As research forms and operational forms were built throughout the enhancements process, we determined the reliability of each section score and of the Composite score.

1.3.6 Ensuring Continued Alignment to Common State Standards

The alignment of each section of the ACT to common state standards in ELA, math, and science is needed to ensure that the ACT continues to be an effective source of assessment data under ESSA. The proposed reductions in test time and the number of items on each section were thus constrained by the need to maintain alignment to standards and adequately sample each domain. Each content cluster within a set of standards (e.g., the geometry cluster within the math standards) had to maintain a sufficient number of items to adequately sample the domain for each student and sample the entire set of standards over multiple forms. Since each section of the ACT already aligns to most common standards (standards based on the Common Core State Standards for ELA and math and the Next Generation Science Standards and NRC Framework for science), ACT had to determine how many passages and items would be necessary to achieve adequate coverage in each section of the test. To address alignment coverage with a potentially shortened assessment, ACT considered comparing the current content sampling methods with a formal matrix sampling method. More information on alignment studies is included in Chapter 3 regarding validity related to content.

1.3.7 Ensuring Continued Alignment to Career Readiness

The ACT has been linked psychometrically to the WorkKeys Career Readiness system. Because there is substantial, but not complete, overlap between the skills needed for both career and college readiness, the ACT can be used to predict what portable skills test takers have when they begin a job. The ACT will continue to have a concordance table to link ACT scores to the appropriate ACT® WorkKeys® National Career Readiness Certificate® (NCRC®) level (bronze, silver, gold, and platinum).

1.3.8 Providing Student Choices for the Assessment

Student choice for the assessment is primarily related to two facets: test mode (paper or online) and the science section as optional.

To research test mode choices, ACT has conducted surveys after officially administered ACT tests, during screening procedures for cognitive labs, and during interviews after usability and cognitive labs. The results show that substantial percentages of students (25%–40% depending on the sample) prefer one mode over the other, while approximately the same percentage of students have no real preference. As a result of these surveys, ACT is committed to maintaining both modes so that students will be able to select their preferred mode for national testing. Consequently, we decided that any changes to the ACT during this periodic review must be compatible with both paper and online testing modes. Therefore, technology enhanced items, constructed response items, and composite items that integrate constructed response and forced-choice options were not acceptable solutions, as these item types do not all work equally well on paper and online tests.

The other major consideration for student choice was making the science test optional. The current ACT is used for accountability for science in some states but not others. Additionally, feedback has indicated that some students consider the science test difficult and unrelated to their intended major. Therefore, we investigated the possibility of making the science test optional, as doing so would help us meet one of the goals of the redesign, shortening the overall test. An optional science test would allow students to have four different options for taking the ACT.

- Option 1: English, math, and reading
- Option 2: English, math, reading, and science
- Option 3: English, math, reading, and writing
- Option 4: English, math, reading, science, and writing

1.3.9 Composite Scores, ELA Scores, STEM Scores, and Superscoring

The enhanced ACT will still report the ACT Composite score using a Superscoring method based on the individual components of the battery. We created a method for Superscoring tests with and without science. Results from this Superscoring method are presented in Chapter 4.

For the legacy ACT, a test taker must take the English, reading, and optional writing sections to receive a combined ELA score. This will remain unchanged for the enhanced ACT. For the legacy ACT, all students receive a STEM score, which predicts performance in the courses typically taken by STEM majors rather than performance in general education courses, which is indicated by the College Readiness Benchmarks. For the enhanced ACT, students will receive a STEM score only if they take the optional science section.

1.3.10 Modernizing the Test and Pretesting

Another goal of the periodic review was looking for ways to modernize the test development and pretesting processes so that we could create pre-equated tests that would allow for faster score reporting for all stakeholders: students, schools, and states. Currently, the pretesting of new items and passages occurs only during national testing. This presents several challenges. First, because the items are in a separate test, many students do not try their hardest, which hinders the ability to pre-equate forms. Second, since students self-select to take the national ACT, the sample is not equivalent to the population for school day testing.

To modernize testing processes, we will embed field-test items in with the operational items for all enhanced ACT test administrations, including school day and national. Since students will not know which items are operational and which are field-test items, issues with non-effortful behavior should be minimized. For the three passage-based sections (English, reading, and science), field-test units will include an entire passage with a standard number of items attached. As is typical, the embedded field-test (EFT) units will not be the first or last unit of the section. The EFT math items will be inserted throughout the test as discrete items, not as the first or last few items. Field-test items will be administered to the entire testing population under the same time constraints as those for operational items. These pretest data should be more reliable than the current fifth test model and will reflect a sample that is more representative of the entire testing population.

1.3.11 Considering Universal Design and Equity as Part of the Redesign Process

As with any potential revisions to the ACT, it was important to consider equity and accessibility when we evaluated possible changes to the test. Universal design (UD) for assessment best practices were considered from the onset at the item, passage, section, and form levels, and UD experts were brought in to review all proposed changes to the test. This commitment to universal design for assessment can be seen in changes such as adding stems to all English items to ensure that students can move through the test with “simple, clear, and intuitive instructions” (Lazarus et al., 2022, p. 3); this change will be detailed in section 1.5.1.2.3.

It was also critical to examine all potential changes through an equity lens. In addition to analyzing differential item functioning (DIF) data, ACT drew on lessons learned during the

2021–2022 Windows and Mirrors roundtable discussion series, which brought together leaders, educators, and researchers working with diverse student populations to refine ACT's best practices in incorporating ELA texts that both recognize the complex mix of student experiences and create affirming reading experiences for students. Math considerations included recent work to investigate the inclusion of culturally relevant items to address equity concerns. The science test will continue to refer to "students" and "scientists" without reference to names. Science has also investigated the use of active versus passive voice in passages and items to address potential equity concerns while maintaining the construct integrity of using materials similar to those used in first-year college courses.

ACT will continue to conduct research on equity, particularly as it relates to underserved learners (e.g., African American, Latino, and first-generation college students), to ensure that updates to the ACT allow all test takers to show what they know and can do. Changes to timing to reduce speededness should also be a benefit, since ACT research has found that traditionally underserved groups exhibit larger speededness effects.

1.3.12 Comparability of Scores from Different Modes

The enhanced ACT should not exhibit mode effects between paper delivery and online delivery. The results of the mode comparability study are discussed in Chapter 4. Current techniques for equating forms across paper and online delivery should be used.

1.4 Decisions Related to the Periodic Review and Updates to the ACT

1.4.1 Elements of the ACT That Will Not Change With Updates to the Test

1.4.1.1 The Enhanced ACT Will Measure College and Career Readiness

The ACT has a strong history of empirically linking student performance on the ACT test to various measures of college and career readiness such as grades in first-year college courses, retention, progress toward graduation, and graduation rates. Since the knowledge, skills, and abilities (KSAs) associated with college and career readiness have not drastically changed, many of the same KSAs will be included on the enhanced ACT. Some new KSAs will be added, but they should also predict college and career readiness. Studies are planned to verify that the enhanced ACT will continue to show strong correlations with positive outcomes in college.

1.4.1.1.1 First-Year Grades

The ACT College Readiness Benchmarks are based on student performance in key first-year classes in each domain. The enhanced ACT will continue to predict first-year grades. New Benchmark scores will be calculated based on the scores associated with a 75% likelihood of earning a C or better and a 50% likelihood of earning a B or better. ACT regularly conducts follow-up research on this and has adjusted the Benchmark scores accordingly to match current research data. ACT will conduct that research again and determine the new Benchmark scores. The research team may also reexamine the Benchmark courses to better align them with the general education requirements that students typically fulfill with first-year courses.

1.4.1.1.2 Retention

Higher legacy ACT Composite and STEM scores are correlated with higher retention rates from the first to the second year of college (Mattern et al., 2015; Westrick, 2017). Since the enhanced ACT will still target key KSAs related to college and career readiness, the test should continue to predict higher retention rates. Additionally, the new KSAs are associated with skills that postsecondary NCS respondents consider important to success in college, as well as skills currently taught in high school that have not been tested before. Students who are strong in these KSAs will likely be successful in entry-level college classes and will more likely continue at their first college. ACT's research team will conduct research to confirm these hypotheses for the enhanced ACT.

1.4.1.1.3 Progress Toward a Degree

ACT research has found that higher grades and other outcomes related to degree progress are correlated strongly with legacy ACT scores and with meeting the College Readiness Benchmarks. The enhanced ACT should show similar correlations because it will assess similar KSAs related to college and career readiness. The research team will conduct studies to confirm that the enhanced ACT scores show correlations with these measures.

1.4.1.1.4 Degree Attainment

Prior ACT research has shown that degree attainment at both 2- and 4-year institutions is correlated with legacy ACT scores (ACT, 2010, 2013a, 2013b; Radunzel & Noble, 2012) and stronger when students meet the STEM Benchmark score (Mattern et al., 2015). Students with higher legacy ACT scores have higher probabilities of completing an associate's degree in 3 years or less and a bachelor's degree in 6 years or less. Scores from the enhanced ACT should also show correlations with these important outcomes, and the research team will conduct studies to confirm this hypothesis.

1.4.1.2 The Enhanced ACT Will Be Validated by Experts

As with other periodic reviews, external expert panels were brought together to evaluate the proposed changes to the legacy ACT and to evaluate the KSAs that will be used to assess college and career readiness. Each test section held a separate panel with experts chosen for their knowledge, expertise, and experience relative to the constructs and assessment of English, mathematics, reading, or science. These panels also evaluated the alignment of specifications and blueprints to common state standards for high school ELA, math, and science. They reviewed matrix sampling proposals to evaluate how each assessed the full breadth and depth of skills from both the K–12 and college readiness perspectives compared to current sampling methods. Additionally, ACT's Technical Advisory Council (TAC) evaluated draft and final versions of the enhanced ACT blueprints and specifications to confirm that the assessment will continue to predict college and career readiness. Based on recommendations from the expert panels and TAC, minor revisions to the enhanced ACT have been made and may still occur.

1.4.1.3 The Enhanced ACT Will Be Validated by Outcome Research

ACT will conduct studies on college outcomes (first-year grades, retention, progress toward graduation, etc.) to provide additional evidence that the enhanced ACT predicts college and career success. To ensure representativeness of the samples, we will specifically target Hispanic serving institutions and historically Black colleges and universities.

1.4.1.4 The Enhanced ACT Will Maintain the 1–36 Scale for Composite and Section Scores

External stakeholders in both postsecondary education and at the state and district levels want continuity between the legacy and enhanced score scales. Consequently, the scale will remain 1–36. The research team will investigate and produce a new scale score distribution using the 1–36 scale. The results of a linking study connecting the new score distribution to the old score distribution are presented in Chapter 4.

1.4.1.5 The Enhanced ACT Will Be Vertically Aligned With the PreACT Suite of Assessments

Blueprints and scaling for the PreACT assessments will be vertically aligned with those for the ACT. Since many KSAs can be learned early and applied to successively more complex passages, mathematical scenarios, graphics, and scientific scenarios, the KSAs on the PreACT tests and those on the ACT will necessarily overlap. For the PreACT tests administered in Grades 8, 9, and 10, ACT will conduct studies to determine the scores that students need to earn to be on track or nearly on track to achieve college readiness.

1.5 Initial Proposed Solutions and Blueprints by Test Section

ACT examined each section of the test to determine how that section might be adapted to best meet the competing goals of increasing time per item, creating a shorter test overall, fulfilling alignment needs, and maintaining reliable scores that predict college readiness. The following sections provide details on the review process for each test section. For each section, we had to weigh the two main uses of the ACT: predicting college and career readiness and measuring student mastery of state standards for accountability purposes under ESSA.

Many decisions were based on common data such as the Reading Rate Study (Arthur & Thomas, 2023), which found that students typically read at approximately 225–250 standard words per minute until they felt rushed because of the short time remaining. This was used to approximately account for the time needed to process a passage and then spend the remaining time answering questions. Another factor common to all the test sections was the need for all items to be deliverable in a near identical manner for both paper and online testing, which would require all items and passages to adhere to universal design for assessment principles such as WCAG compliance. Other decisions were based on the unique nature of assessing KSAs in the ELA, math, and science domains.

1.5.1 Proposed Solutions for the ELA Sections: English

1.5.1.1 English Passages

The ACT English test uses passages (essays) to put the student in the position of a writer making specific choices about meaning and clarity. The test will continue to feature essays written in a variety of styles for different rhetorical purposes and reflecting a wide range of subjects and cultural contexts. However, we are making some changes to passage type and length to better address rhetorical skills related to building and supporting arguments, to ensure exposure to different writing types on each form, and to reduce the overall section test time.

Writing type. Data from the two most recent ACT National Curriculum Surveys (ACT 2016, 2020) indicate that the ability to identify and evaluate claims is important for college readiness across all subject areas. At the same time, because the ACT writing test is an optional component, students have had limited opportunities to demonstrate their mastery of writing skills related to argumentation and persuasion. Alignment studies conducted for ESSA peer review have also indicated that the range of items aligned to common state writing standards can be relatively narrow. To address these concerns, ACT will introduce an argumentative essay on each enhanced ACT English form to target important skills such as developing a thesis or claim and identifying evidence that supports or challenges a claim.

Legacy ACT English tests have historically relied on informational essays, with narrative or persuasive essays featured on some forms but not required as part of an official blueprint. On the enhanced ACT, each operational English test will feature 2–3 informational essays, 1–2 argumentative essays, and 1 narrative essay. The balance of essay types on each form is intended to reflect how most state standards approach writing types at the 11th- and 12th-grade levels, which is also in accordance with the NAEP writing framework, with informational and argumentative writing representing the bulk of student writing at this level. Because of this change, each enhanced ACT test form will not only better highlight important argumentative skills but also allow students to demonstrate key writing skills and strategies for a variety of purposes.

While each essay will support item types from all reporting categories, the specific content of each essay will allow for items that test content appropriate to different audiences and goals. For example, an argumentative essay might require students to make choices about developing arguments and selecting support for claims while maintaining a formal style and tone. On the other hand, a narrative essay might require students to make writing choices that develop a narrative with a more relaxed style or reflective tone. Essays might also blend text types.

As with the legacy ACT English test, essays will draw from a wide array of topics chosen both for their appropriateness in assessing writing and language skills and for their likelihood of engaging students' interests and experiences. Informational essays will present topics across the humanities, social sciences, and natural sciences. Narrative essays will primarily feature first-person accounts of novel activities and experiences. Argumentative essays will take a stance on a specific issue but will not ask students to assume one side of a charged political issue. The essays will be narrow in focus and written to make advanced topics accessible to all test takers. For example, an argumentative essay might claim that esports (organized multiplayer video game competitions) should be recognized as legitimate sports on par with more traditional sports like soccer and baseball, citing the cognitive demand and discipline required of esports players and the global impact of esports in general. Specialized terminology will be clearly defined and used only when relevant to the context.

Essay length. Three of the operational essays will remain approximately 340 standard words long; the other two will be approximately 185 standard words long. Keeping some essays at the current length while introducing shorter essays provides an opportunity to shorten the overall test while still addressing the three major writing types noted above. Though each essay will support items in every reporting category and all item types will appear across the different essay lengths, the shorter essays will reduce the overall reading load while allowing for items assessing the conventions of Standard English, knowledge of language, and some topic development and organization topics that do not require longer context. The longer essays will continue to allow for complex topic development, organization, and style items that require students to evaluate sentences or paragraphs in the context of the whole essay. Essay length will not be bound by writing type or genre (i.e., shorter essays will not be limited to informational, narrative, or argumentative essay types; any of these types could appear in short or long format).

1.5.1.2 English Items

Several changes are being made at the item level, all of which are intended to strengthen the English section by highlighting skills most important to college and career readiness and addressing fairness and accessibility concerns. Changes include adjustments in reporting category coverage and the introduction of items targeting argumentative skills. In addition, all item formats will be updated to include stems.

1.5.1.2.1 Reporting Categories

The enhanced ACT English test will continue to feature items in three reporting categories—Production of Writing, Knowledge of Language, and Conventions of Standard English (refer to Table 1.3). The percentage of items within each reporting category is being adjusted to reflect a greater emphasis on writing knowledge, skills, and abilities related to topic development, organization, and language. The English test will continue to feature important skills in the Conventions of Standard English reporting category, but the focus of these items will be on making choices that impact meaning or clarity.

Table 1.3. Enhanced ACT English Reporting Categories

Reporting category	Skill area	Description/examples
Production of Writing	Topic Development—Purpose and Focus: Involves the ability to make content and stylistic choices that provide support for a text's rhetorical purpose.	Determine if a text's purpose is supported by organizational structure and content. Revise text to enhance the focus and cohesion.

Reporting category	Skill area	Description/examples
	Organization, Unity, and Cohesion: Involves the ability to support a text's purpose by progressing from point to point logically and smoothly.	Order sentences and paragraphs and use transitions to enhance overall purpose, unity, and logical cohesion. Frame texts effectively with transitions, introductions, and conclusions.
Knowledge of Language	Expressing Ideas Clearly: Involves the ability to be precise and concise by using vocabulary skillfully and by avoiding wordiness and redundancy.	Use general academic and domain-specific language precisely and eliminate redundancy and wordiness when the meaning of the sentence or paragraph must be considered.
	Style: Involves the ability to use language purposefully and to maintain stylistic consistency appropriate for the communication task.	Maintain a consistent style and tone and use words, phrases, and sentences purposefully, considering their effect on the whole text.
Conventions of Standard English	Sentence Structure and Formation: Involves the ability to ensure the grammatical soundness of a variety of sentences.	Recognize and correct subtle structural errors in sophisticated sentence structures and complex contexts, including when the meaning of multiple sentences or paragraphs must be considered.
	Usage Conventions: Involves the knowledge of and ability to apply the rules of Standard English usage.	Recognize and correct usage errors in structurally sophisticated sentences, including when relevant elements are separated by intervening text.
	Punctuation Conventions: Involves the knowledge of and ability to apply the rules of Standard English punctuation.	Recognize and correct punctuation errors in sophisticated sentence structures and complex contexts, including using punctuation to reduce the ambiguity of sentences and paragraphs.

1.5.1.2.2 Argumentative Items

Items specific to identifying claims and developing arguments, including items about selecting the strongest support for specific claims and evaluating counterclaims, will most often fall under the Production of Writing reporting category. Topic development items may ask students to select the strongest support for a specific claim or to identify how specific details support a claim. Organization items may ask students to provide a conclusion that reinforces the essay's main claim or to sequence information to present an argument logically. These items will primarily be featured in argumentative essays, though they may also lend themselves to informational or narrative essays that present specific claims in service of a larger expository essay or narrative.

1.5.1.2.3 Stems

To adhere to universal design for assessment principles, every English item will now feature a stem that provides instruction for that item. Grammar items and others that were traditionally stemless (such as those testing precision in language, style and tone, or transition words), will be introduced with language that specifies the task. An example of an item testing knowledge of language is below.

Legacy ACT Item

A. NO CHANGE
B. transforming
C. dislocating
D. contorting

Enhanced ACT Item

Which choice is clearest and most precise in context?

A. **No Change**
B. transforming
C. dislocating
D. contorting

1.5.1.2.4 Item Types That Are Being Discontinued

Several English item types will be discontinued in order to adhere to universal design for assessment principles and address potential issues of bias:

- Usage items testing idiomatic language
- Sentence structure items with "NOT" in the stem
- Organization items testing paragraph ordering. While paragraph-level organization is important, items asking students to reorder the paragraphs in an essay were often confusing and did not feel organic to a typical writing process. Because these items are difficult to craft effectively, they were rare on the ACT; now they will be discontinued completely. Paragraphing skills will still be assessed through items asking for the best

introductions and conclusions to paragraphs, items asking where sentences should be placed within a paragraph, and items asking whether and where paragraphs should be separated.

1.5.1.3 English Forms and Blueprints

In modifying English forms and blueprints, ACT weighed the desire to reduce the overall number of items and the test length against the need to maintain sufficient coverage of common ELA state standards and sufficient reliability at the form level. We determined that having five operational essays with fewer items per passage would provide sufficient coverage of the standards, maintain the appropriate psychometric reliability, and allow for diversity in gender, ethnicity, socioeconomic status, and region, while the mix of longer and shorter essays would ensure that ACT English essays continue to explore diverse topics with appropriate depth and nuance.

We further determined that reducing the length of two essays would help to decrease the total testing time and address the perception of speededness, both of which were goals of the enhanced ACT. The resulting initial blueprint revisions included new shorter essays with 5 items each. The number of items attached to the longer essays was reduced from 15 to 10, resulting in an overall reduction of operational items from 75 to 40.

An additional embedded field test (EFT) slot with either one long passage with 10 items or two short passages with 5 items each brings the total number of items to 50.

1.5.1.3.1 Initial Enhanced ACT English Content Blueprint

Table 1.4. English Content Blueprint

Part I: English content		ACT Legacy	ACT Enhanced
Total testing time		45 min	35 min
Items (total number)		75	40 (+10 EFT)
Reporting categories (# items)	Production of Writing	22–24	15–17
	Knowledge of Language	11–13	7–9
	Conventions of Standard English	39–41	15–17
Passages (total number)		5	5 (+1–2 EFT)
Essay type (# passages)	Informational	4–5	2–3
	Argumentative	0	1–2
	Narrative	0–1	1
Genre (HUM, SSC, NSC, etc.)		0–2 of each	0–2 of each
Length (# passages)	340 standard words	5	3
	185 standard words	0	2
Items per passage	340-word passages	14–16	10
	185-word passages	N/A	5
Standard word count (passage total, approx.)		1,750	1,390

Note. HUM (humanities), SSC (social sciences), and NSC (natural sciences) are examples of codes used to ensure broad coverage of content.

1.5.1.4 English Matrix Sampling Proposal

The English matrix sampling proposal focused on covering skills across all three essay types. Over any three forms, this proposal would offer at least six informational essays, three argumentative essays, and three narrative essays. Though every individual form will allow students to demonstrate writing skills related to topic development, organization, language, and style for a variety of writing types, the matrix sampling proposal would ensure strong depth and breadth of coverage in the Production of Writing and Knowledge of Language reporting categories for each writing type. The proposal would also ensure depth and breadth of coverage for Conventions of Standard English items, which do not typically vary by writing type but have been reduced per form.

1.5.2 Proposed Solutions for the ELA Section: Reading

1.5.2.1 Reading Passages

The reading test uses published texts to measure a student's ability to read closely, reason about texts using evidence, and integrate information from multiple sources. The test will continue to use published literary and informational texts as a basis for items that focus on the mutually supportive skills that readers apply when studying written materials across a range of subject areas.

Excerpting passages from published sources also allows ACT to meet passage goals in two other important areas: quality and authenticity. As excerpts from real sources, ACT reading passages provide authentic reading experiences on a wide variety of topics and perspectives. Informational texts relate the domain-specific knowledge and expertise of their authors and subjects with nuance and specificity. Literary texts reflect a range of styles, voices, and points of view. Furthermore, excerpting externally published work allows ACT to feature diverse, contemporary voices and experiences in a direct and unmediated way. Diverse representation is intrinsic to the fairness of the ACT reading test. In selecting passages, the ELA team will continue to use the lessons learned during the recent roundtable discussion series Windows and Mirrors, which brought together thought leaders, educators, and researchers working with diverse student populations with a goal of further refining ACT's best practices in incorporating texts that both recognize the complex mix of student experiences and create culturally affirming reading experiences for students.

Length. A wide body of research continues to support the finding that college- and career-ready students must be able to comprehend complex texts (Adams, 2009; National Governors Association Center for Best Practices & Council of Chief State School Officers, 2010; Wei et al., 2016). While there is not a 1:1 correlation between text length and text complexity, most qualitative text complexity measures contain factors inextricably bound up in length, including levels of meaning, purpose, and structure. Reading and literacy standards for almost every state in the nation call for engagement with complex, authentic, and high-quality texts, and many common state standards at Grades 11–12 call for textual analyses such as analyzing how ideas develop over the course of a text; explaining how individuals, ideas, or events interact over the course of a text; and analyzing how an author's choices about structuring specific parts of a text contribute to meaning as a whole. Such analyses can best be achieved through engagement with multiparagraph passages.

Passage length remains a key feature that distinguishes the ACT reading test from other college entry assessments used for ESSA. At the same time, a common concern from test takers was the amount of text they had to read and analyze in the amount of testing time given. As such, updates to the reading test needed to address perceived speededness even as ACT decided to retain passage length and move to an EFT model that requires a field-test passage and item set in every test section. We determined that some passages could be shortened and still assess important skills; therefore, the reading test will continue to use 750-standard-word passages while introducing some 650-standard-word passages. Passages of both lengths will enable reading standards for literature and informational texts to be assessed with appropriate rigor, allowing students to demonstrate essential skills such as analyzing complex structures, evaluating the development of important ideas over the course of a text, and synthesizing multiple levels of meaning.

Format. Passages will continue to feature three different formats: single 650-word or 750-word passages, paired passage sets featuring two shorter passages, and passages integrating visual or quantitative data (e.g., a graph or table) alongside the passage.

Genre. The enhanced ACT will continue to include literary narratives and informational texts from the humanities, natural sciences, and social sciences. Literary narratives will be drawn from fiction and memoirs that have strong literary elements.

1.5.2.2 Reading Items

There are no major changes to the items in the reading section. Items will draw from three reporting categories and will continue to be coded to the apex skill. For example, if a student must use a close reading skill (which would fall under the Key Ideas and Details reporting category) in order to evaluate the purpose of a phrase, sentence, or paragraph (a Craft and Structure skill), that item will continue to be coded primarily to the Craft and Structure reporting category.

1.5.2.2.1 Reporting Categories for the Reading Section

The reading reporting category names and underlying skill areas will remain the same.

Table 1.5. Reporting Categories and Skill Areas for the Reading Test

Reporting category	Skill area	Description/examples
Key Ideas and Details	Close Reading: Involves the ability to attend carefully to what a text says and draw well-supported conclusions from a text.	Analyze challenging, complex, and highly complex texts to determine what the text says explicitly as well as draw conclusions based on textual support.

Reporting category	Skill area	Description/examples
	Relationships: Involves the ability to identify and understand relationships between individuals, events, themes, and ideas in a text.	Identify and infer sequences, comparative relationships, and cause-effect relationships developed across a text.
	Central Ideas, Themes, and Summaries: Involves the ability to synthesize information in a text in order to identify central ideas or themes, differentiate key ideas from less important ideas, and summarize text concisely.	Determine a central idea or theme of a challenging, complex, or highly complex text and summarize ideas and information developed across a text.
Craft and Structure	Word Meanings and Word Choice: Involves the ability to determine the meaning of words and phrases (including academic and domain-specific words, multiple-meaning words, and figurative language) based on the context of a text.	Determine the meanings (including figurative, connotative, and technical meanings) of words and phrases as they are used in more challenging, complex, and highly complex texts.
	Text Structure: Involves the ability to analyze the rhetorical aspects of a text to understand how an author's choices create effects on the reader.	Analyze rhetorical devices and the structure of more challenging, complex, and highly complex texts.
	Purpose and Point of View: Involves the ability to understand and analyze a text's rhetorical situation, including the author's intent, perspective, and use of rhetorical techniques.	Analyze stated and implied purposes in texts; analyze point of view and narrative techniques in narrative texts; analyze rhetorical techniques as well as authorial bias.
Integration of Knowledge and Ideas	Arguments: Involves the ability to understand and analyze arguments in a text, including claims, counterclaims, and supporting evidence.	Analyze the use of persuasive elements and development of an argument in more challenging, complex, and highly complex texts, assessing whether the evidence provided is relevant, sound, and sufficient.
	Synthesis of Multiple Texts: Involves the ability to make connections between, and integrate knowledge across, two or more texts.	Analyze how different literary, thematic, and structural elements inform both shared and distinct ideas when comparing more challenging, complex, and highly complex texts; synthesize information across texts to build new knowledge and insights.

Reporting category	Skill area	Description/examples
	Visual and Quantitative Information: Involves the ability to understand and analyze visual information, including tables, charts, graphs, and figures alongside text.	Analyze visual information to draw conclusions and determine how this information relates to more challenging, complex, and highly complex texts.

1.5.2.3 Reading Forms and Blueprint

As noted, the current reporting categories—Key Ideas and Details, Craft and Structure, and Integration of Knowledge and Ideas—will be retained, and the enhanced ACT reading test will continue to feature a variety of item types in all three reporting categories. The percentage of items in each reporting category is being adjusted to reflect a decrease in Key Ideas and Details (KID) items, particularly DOK 1 items that require finding less important details or sequencing multiple events that may not be essential to understanding the main points of the passage.

To determine the feasibility of reducing the total number of items on the reading test while still assessing the skills most indicative of college and career readiness and maintaining alignment to common state standards, ACT analyzed market research surveys and interviews involving K–12 and postsecondary educators and administrators, feedback from external reviewers in the high school and postsecondary spaces, recent ESSA peer reviews, and data from the most recent ACT National Curriculum Surveys. We determined that reducing the number of items in the KID reporting category would help meet both goals.

Data indicate that many skills in the KID reporting category are integral to college and career readiness, among them determining central ideas, identifying important details, and drawing conclusions and making inferences (ACT 2012, 2016, 2020). At the same time, ESSA peer review comments illustrated concerns with too many DOK 1 items that represent skills below the desired grade-level complexity and an imbalance of item types within the KID reporting category. In reducing the overall percentage of KID items, ACT will continue to test important close reading skills while reducing the time-consuming items mentioned above. The reading test will continue to use all item types within the KID category but will focus on items that are most essential to understanding the main points of each passage. Test and development blueprints will specify a range of these item types to ensure that the full breadth of KID skills is reflected in item development and on test forms. Moreover, students will continue to use close reading skills in answering items from all reporting categories.

1.5.2.3.1 Reading Proposed Content Blueprint

Table 1.6. Initial Enhanced ACT Reading Content Blueprint

Part I: Reading content		ACT Legacy	ACT Enhanced
Total testing time		35 min	40 min
Items (total number)		40	27 (+9 EFT)
Reporting categories (# items)	Key Ideas and Details	21–24	12–14
	Craft and Structure	10–12	7–9

Part I: Reading content		ACT Legacy	ACT Enhanced
	Integration of Knowledge and Ideas	6–9	5–7
Passages (total number)		4	3 (+1 EFT)
Genre (# passages)	Literary narrative	1	1
	Informational	3	2
Format (# passages)	Single passage	2–3	2
	Paired passage	1	0–1
	VQI passage	0–1	0–1
Length (# passages)	750 standard words	4	2
	650 standard words	0	1
Items per passage		9–11	9
Total word count scored passages (scored + EFT passage)		3,000 (N/A)	2,150 (2,800–2,900)

1.5.2.4 Reading Matrix Sampling Proposal

The reading sampling proposal focused on the format and genre of the reading passages to ensure that each form would have a mix of essential skills across all three reporting categories for literary and informational texts, particularly in the Integrating Knowledge and Ideas reporting category. To achieve greater depth and breadth of standards coverage, ACT used a mix of passage format (single passage, paired passage, VQI passage) and genre (literary narrative and informational) over multiple forms.

Over a set of three forms, the following passage genres and formats would all be assessed at least once, along with the skills associated with each.

Table 1.7. Targeted Skills by Genre and Passage Format

Passage genre	Passage format	Primary skills assessed
Literary narrative	Single	KID, CAS
	Paired	KID, CAS, IKI (synthesis only)
Informational	Single	KID, CAS, IKI (argument only)
	Paired	KID, CAS, IKI (synthesis, argument)
	VQI	KID, CAS, IKI (VQI, argument)

In this proposal, passages and items requiring students to synthesize information across multiple sources would be spread out so that every form had either a scored paired passage set or a scored VQI passage set. With the addition of the EFT unit, students could encounter both a paired passage and a VQI passage, but no single form would feature two paired passage units or two VQI units.

Table 1.8. Proposed Reading Matrix Sampling Protocol

Form	Passage	Genre and format
A	1	Literary narrative: single passage
	2	Informational: single passage
	3	Informational: paired passage
B	1	Literary narrative: single passage

Form	Passage	Genre and format
	2	Informational: single passage
	3	Informational: VQI passage
C	1	Literary narrative: paired passage
	2	Informational: single passage
	3	Informational: single passage

1.5.3 Proposed Solutions for the STEM Sections: Math

The goals of less time testing, more time per item, and field-testing that better represents the school day testing population are shared by ACT and external stakeholders. External stakeholders in state and district markets also desire forms that better align to common state standards. With the enhanced ACT math test, ACT accomplished these three goals simultaneously by reducing the total number of items, with the largest reduction in the Integrating Essential Skills (IES) reporting category.

1.5.3.1 Math Items

To reduce the reading load and address issues of speededness, we reduced the number of answer options in each math item from five to four, which would better allow students to demonstrate their knowledge, skills, and abilities. With four answer options, more items will completely fit on the computer screen during online testing without the need for scrolling, addressing issues of universal design for testing. Going to four answer options will also allow potentially extraneous information to be removed from some items. All of this should provide a better test experience for students. ACT has begun studies to determine how the shift from five to four answer options affects student performance as reflected in measures such as difficulty, discrimination, and DIF. All of this will be explored in Chapters 2, 3, and 4.

1.5.3.1.1 Math Reporting Categories

We determined that the current reporting categories are clear and meaningful to stakeholders, so the reporting category names and structure will remain the same. The reporting categories and skill areas align with most common state standards, so keeping the current structure should ease transitions for stakeholders and will minimize changes to ancillary materials for alignment, test prep, college and career readiness standards, and other supports. The reporting category names and structure can help students, teachers, and others understand student strengths and weaknesses and areas for study before or after taking the ACT math test. Each reporting category and skill area is described below and can be seen in Table 1.9.

Preparing for Higher Mathematics. Students apply the mathematics they have learned more recently. This reporting category is divided into five subcategories:

- **Number and Quantity.** Students demonstrate knowledge of real and complex number systems. They understand and reason with numerical quantities, including integer and rational exponents and vectors and matrices.

- **Algebra.** Students solve, graph, and model multiple types of expressions. They work with many kinds of equations, including linear, polynomial, radical, and exponential relationships. They find solutions to systems of equations and apply their knowledge to applications.

- **Functions.** Students apply their knowledge of function definition, notation, representation, and application. Items include linear, radical, piecewise, polynomial, and logarithmic functions. Students manipulate and translate functions, as well as find and apply important features of graphs.

- **Geometry.** Students define and apply knowledge of shapes and solids, such as congruence and similarity relationships or surface area and volume measurements. They understand the composition of objects and solve for missing values in triangles, circles, and other figures by using, for example, trigonometric ratios and equations of conic sections.

- **Statistics and Probability.** Students describe the center and spread of distributions, apply and analyze data collection methods, understand and model relationships in bivariate data, and calculate probabilities.

Integrating Essential Skills. Students put together knowledge and skills to solve problems of moderate to high complexity. Topics include rate, proportion, area, using expressions to represent quantities, using equations to capture relationships, rational exponents, the basics of functions, and data analysis.

Modeling. Students use mathematics to represent, through a model, an analysis of an actual situation. 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 discussed above. Thus, the Modeling reporting category is an overall measure of how well a student uses modeling skills across mathematical topics.

Table 1.9. Enhanced ACT Mathematics Reporting Categories and Knowledge/Skills Assessed

Reporting category	Skill area	Description/examples
Preparing for Higher Mathematics: Number and Quantity	Rational and Irrational Numbers	Use and apply the properties of rational and irrational numbers.
	Properties of Exponents	Use and apply the properties of exponents.
	Vectors and Matrices	Model situations, perform operations, and solve problems involving vectors and matrices.
	Complex Numbers	Perform operations and solve equations involving complex numbers.

Reporting category	Skill area	Description/examples
	Quantities and Units	Reason quantitatively and use units to solve problems.
Preparing for Higher Mathematics: Algebra	Linear Expressions, Equations, and Inequalities	Model situations, solve problems, and perform operations involving linear expressions, equations, and inequalities.
	Quadratic Expressions, Equations, and Inequalities	Model situations, solve problems, and perform operations involving quadratic expressions, equations, and inequalities.
	Rational and Radical Expressions and Equations	Model situations, solve problems, and perform operations involving rational and radical expressions and equations.
	Polynomial Expressions and Equations	Model situations, solve problems, and perform operations involving polynomial expressions and equations.
	Systems of Equations and Inequalities	Write, solve, and graph systems of equations and inequalities.
Preparing for Higher Mathematics: Functions	Properties of Functions	Evaluate and create functions. Describe their properties. Convert between different representations of functions.
	Function Composition and Inverse Functions	Compose functions, find inverse functions, and find the domain and range of a function composition.
	Sequences and Series	Model situations, perform operations, and solve problems involving sequences and series.
	Trigonometric Functions	Model situations, perform operations, and solve problems using trigonometric functions.
	Exponential and Logarithmic Functions	Model situations, perform operations, and solve problems involving exponential and logarithmic functions.
Preparing for Higher Mathematics: Geometry	Transformations	Model situations, perform operations, and solve problems involving transformations and their properties in a plane.
	Proof, Reasoning, and Constructions	Construct geometric figures and use logical arguments to prove theorems.

Reporting category	Skill area	Description/examples
	Similarity, Right Triangles, and Trigonometry	Define trigonometric ratios in terms of right triangles. Apply trigonometric ratios to general triangles.
	Conic Sections	Model situations, perform operations, and solve problems involving conic sections.
	Coordinate Geometry	Model situations, perform operations, and solve problems in the coordinate plane.
	Properties of Circles	Model situations, perform operations, and solve problems involving properties of circles.
	Geometric Measurement and Modeling	Apply geometric concepts in modeling situations.
Preparing for Higher Mathematics: Statistics and Probability	Interpret Data on a Single Variable	Summarize, represent, and interpret data on a single count or measurement variable.
	Interpret Data on Two Variables	Summarize, represent, and interpret data on two counts or two quantitative variables.
	Making Inferences from Experiments and Surveys	Interpret and evaluate random processes underlying statistical experiments.
	Rules of Probability	Use rules of probability to compute probabilities and expected values.
	Counting, Permutations, and Combinations	Use counting principles, combinations, and permutations to solve problems and compute probabilities of compound events and solve problems.
Integrating Essential Skills	Properties of Real Numbers	Interpret and apply the properties of real numbers to aid problem-solving.
	Computation and Problem-Solving with Real Numbers	Use all types of real numbers to compute and answer questions.
	Ratio, Proportion, and Percent	Use ratios, proportions, and percents in problem-solving situations.
	Writing Algebraic Expressions	Write algebraic expressions to represent situations including linear and polynomial expressions.
	Writing and Solving Simple Equations and Inequalities	Write equations in one or two variables with linear relationships and use these equations to answer questions.

Reporting category	Skill area	Description/examples
	Perimeter, Circumference, and Area	Calculate perimeter, circumference, and area of polygons and circles.
	Surface Area and Volume	Calculate surface area and volume of solids, including prisms, cylinders, cones, and spheres.
	Measurement Units and Unit Conversion	Model situations, perform operations, and solve problems involving measurement units.
	Properties of Lines, Angles, and Shapes	Use the properties of lines, angles, two-dimensional shapes, and three-dimensional shapes to describe situations and to solve problems.
	Pythagorean Theorem	Use the Pythagorean theorem to solve problems and to find distances.
	Scatterplots and Association	Construct and interpret scatterplots and use linear models.
	Data Summaries and Displays	Describe measures of center, spread, and shape for a data set. Display data in displays such as line plots, dot plots, histograms, and box plots.
	Basic Probability	Compute probabilities for simple events and for compound events where the sample space can be listed.
Modeling	Producing	Produce a model for a given real-world or mathematical context.
	Interpreting	Take information from a model and interpret the information in terms of the situation.
	Understanding	Show understanding by determining conditions under which a model works or does not work.
	Evaluating	Choose the best model for a situation or decide if a model is good enough for a given situation.
	Improving	Change a model or change assumptions of a model by iterating.

1.5.3.2 Math Passages

The enhanced ACT math test does not include passages. Older legacy ACT materials had passage sets with three items that focused on a common dataset, complex geometric figure, etc.; however, those items had already been phased out in an earlier periodic review.

1.5.3.3 Math Forms and Blueprints

For the math section, ACT needed to balance the desire to reduce the total number of items and the testing time with the need to maintain a sample large enough to assess the breadth and depth of the domain appropriate to meeting the goals of predicting college and career readiness and maintaining the "meets expectations" rating for ESSA peer review. Because peer reviewers often identify Integrating Essential Skills items as being below grade level, this reporting category was reduced substantially. However, the NCS has shown that postsecondary educators view these items as important indicators of critical thinking and problem-solving skills that are essential for success in first-year math courses, so the new blueprint maintained 20% of the items in this category (down from 40% in the legacy ACT).

1.5.3.3.1 Calculator Policy

ACT's calculator policy will not change, as there is no evidence from market research or National Curriculum Survey data to show that stakeholders want a change. ACT will continue to prohibit calculators with computer algebra system functionality, which might erode the credibility of claims about what some items measure. Both the ACT National and ACT State and District online testing platforms have a built-in calculator, alleviating fairness and equity concerns.

1.5.3.3.2 Formula Sheet

ACT's formula sheet policy will not change; the enhanced ACT will not provide a formula sheet for the math test. Formula sheets compete for a student's testing time and for screen or desk space. They can also give the illusion that students need to know and apply only the formulas on that sheet. Moreover, for many items that require a complex or unfamiliar formula, the needed formula is included in the item.

1.5.3.3.3 Advanced Topics

There will be fewer items that cover advanced topics in common state standards. These topics are frequently taught in Grade 12 and sometimes in Grade 11. Few students are exposed to many of these advanced topics. Students in high poverty schools or schools with high percentages of traditionally underserved groups are less likely to have access to these courses and more likely to have inexperienced teachers or teachers without proper qualifications for those courses (Bryant, 2015; Flores, 2007). Decreasing the number of items in this category is desirable, but some items that cover advanced topics will remain. These topics can identify students who have taken more advanced math classes, who will score in the higher score ranges, and who are better prepared for STEM majors. Moreover, keeping advanced topics in the domain demonstrates coverage of these topics during ESSA peer review. The legacy ACT forms average six advanced topic items (10%), and the enhanced ACT forms will reduce that number to three items (7%) per form. Decreasing these types of items should help ease concerns about overall form difficulty and reliability, because many students exhibit nonsolution behavior (Wise & Smith, 2011) on these items and reduce the reliability of the form.

1.5.3.3.4 Items in Context

Some math items contain a real-world context. These items require students to understand and process a mathematical concept in a word problem. Items in context will be retained. Currently, there is a range of 17 to 23 in-context items (28% to 38%). On the enhanced ACT math test, 24% to 29% of the items will be in context. These items are valuable, but they take more time to read and are generally more difficult than items covering the same topic without context.

ACT is currently researching culturally relevant items in context and examining how these items affect student performance across groups (Anguiano-Carrasco et al., 2025). The results of this initial study showed that culturally relevant math items exhibited less differential item functioning than the original legacy ACT items did. Additionally, equity-based items may elicit more solution behaviors and fewer nonsolution behaviors (Wise & Smith, 2011), which may lead to students being better able to show what they know and can do.

1.5.3.3.5 Math Content Blueprint

Table 1.10. Initial Enhanced ACT Math Content Blueprint

Part I: Math content		ACT Legacy	ACT Enhanced
Total testing time		60 min	50 min
Total items		60 scored	41 scored and 4 EFT = 45 total
Reporting categories (# items)	1. Preparing for Higher Math (PHM)	36 (60%)	33 (80%)
	1a. Number and Quantity	6 (10%)	4–5 (10%–12%)
	1b. Algebra	8 (13%)	8 (20%)
	1c. Functions	8 (13%)	8 (20%)
	1d. Geometry	8 (13%)	8 (20%)
	1e. Statistics and Probability	6 (10%)	4–5 (10%–12%)
	2. Integrating Essential Skills (IES)	24 (40%)	8 (20%)
	3. Modeling	≥12 (≥20%)	≥8 (≥20%)

1.5.3.4 Math Matrix Sampling Proposal

The math matrix sampling proposal focused on the Preparing for Higher Math (PHM) reporting category and the topics within each PHM subcategory to ensure that each form would have a mix of skills essential for college and career readiness and alignment to common state standards. To achieve greater depth and breadth of standards coverage, ACT used a three-tiered approach to sampling clusters of topics over multiple forms. Clusters in the first tier would be sampled more frequently and would be more likely to appear on every form. Clusters in the second tier would be sampled less frequently, and clusters in the third tier would be sampled rarely. Over a set of three forms, all clusters in all three tiers would be assessed at least once.

ACT developed the tiers by analyzing data from alignment studies, PLDs, state standards, NCS results, and student performance on the legacy ACT. Table 1.11 shows an example of a possible matrix sampling solution that covers the most important topics most frequently while still maintaining complete coverage across three forms. Each form comprises mostly Tier 1 topics and a sampling of Tier 2 and Tier 3 topics. Items on each form and in each reporting category will vary in DOK, difficulty, and complexity, and forms will contain a mix of items in real-world contexts and not.

Table 1.11. Proposed Math Matrix Sampling Protocol

PHM subcategory	Form A	Form B	Form C
Number & Quantity (4–5 items)	Tier 1 Cluster 1	Tier 1 Cluster 1	Tier 1 Cluster 1
	Tier 1 Cluster 1	Tier 2 Cluster 1	Tier 2 Cluster 1
	Tier 2 Cluster 1	Tier 3 Cluster 1	Tier 3 Cluster 1
	Tier 3 Cluster 2	Tier 3 Cluster 2	Tier 3 Cluster 3
	Tier 3 Cluster 3	—	Tier 3 Cluster 4
Algebra (8 items)	Tier 1 Cluster 1	Tier 1 Cluster 1	Tier 1 Cluster 1
	Tier 1 Cluster 2	Tier 1 Cluster 2	Tier 1 Cluster 2
	Tier 1 Cluster 3	Tier 1 Cluster 3	Tier 1 Cluster 3
	Tier 1 Cluster 4	Tier 1 Cluster 4	Tier 1 Cluster 4
	Tier 1 Cluster 5	Tier 1 Cluster 5	Tier 1 Cluster 5
	Tier 1 Cluster 6	Tier 1 Cluster 6	Tier 1 Cluster 6
	Tier 2 Cluster 1	Tier 2 Cluster 2	Tier 2 Cluster 3
	Tier 2 Cluster 4	Tier 2 Cluster 1	Tier 3 Cluster 1
Functions (8 items)	Tier 1 Cluster 1	Tier 1 Cluster 1	Tier 1 Cluster 1
	Tier 1 Cluster 2	Tier 1 Cluster 2	Tier 1 Cluster 2
	Tier 1 Cluster 3	Tier 1 Cluster 3	Tier 1 Cluster 3
	Tier 1 Cluster 4	Tier 1 Cluster 4	Tier 1 Cluster 5
	Tier 1 Cluster 5	Tier 1 Cluster 5	Tier 1 Cluster 6
	Tier 2 Cluster 1	Tier 2 Cluster 1	Tier 2 Cluster 1
	Tier 2 Cluster 2	Tier 2 Cluster 3	Tier 2 Cluster 2
	Tier 3 Cluster 1	Tier 3 Cluster 2	Tier 3 Cluster 3
Geometry (8 items)	Tier 1 Cluster 1	Tier 1 Cluster 1	Tier 1 Cluster 1
	Tier 1 Cluster 2	Tier 1 Cluster 3	Tier 1 Cluster 2
	Tier 1 Cluster 4	Tier 1 Cluster 4	Tier 1 Cluster 3

	Tier 1 Cluster 5	Tier 1 Cluster 5	Tier 1 Cluster 4
	Tier 2 Cluster 1	Tier 2 Cluster 2	Tier 1 Cluster 5
	Tier 2 Cluster 3	Tier 2 Cluster 4	Tier 2 Cluster 5
	Tier 3 Cluster 1	Tier 3 Cluster 2	Tier 2 Cluster 6
	Tier 3 Cluster 3	Tier 3 Cluster 3	Tier 3 Cluster 4
Statistics & Probability (4–5 items)	Tier 2 Cluster 1	Tier 2 Cluster 1	Tier 2 Cluster 1
	Tier 2 Cluster 2	Tier 2 Cluster 3	Tier 2 Cluster 2
	Tier 2 Cluster 4	Tier 2 Cluster 5	Tier 2 Cluster 6
	Tier 3 Cluster 1	Tier 2 Cluster 4	Tier 3 Cluster 1
	—	Tier 3 Cluster 2	—

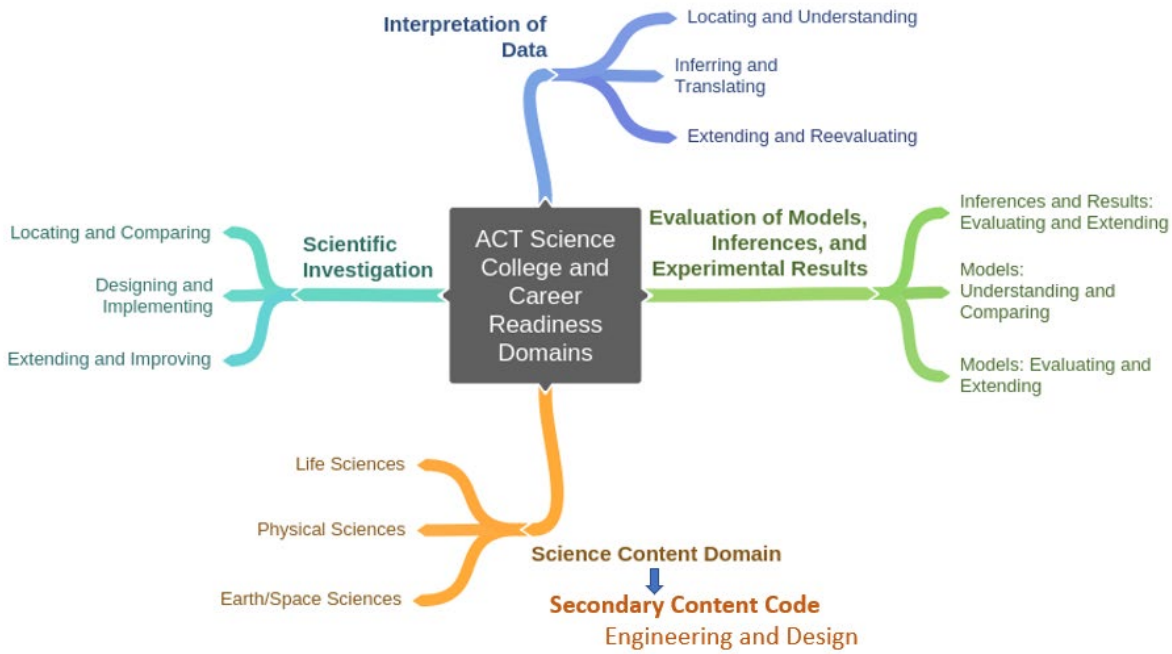
1.5.4 Proposed Solutions for the STEM Sections: Science

The changes to the science section focused on adding more time per item and addressing issues related to engineering and design thinking in common state standards that were not systematically sampled in the legacy ACT.

1.5.4.1 Science Items

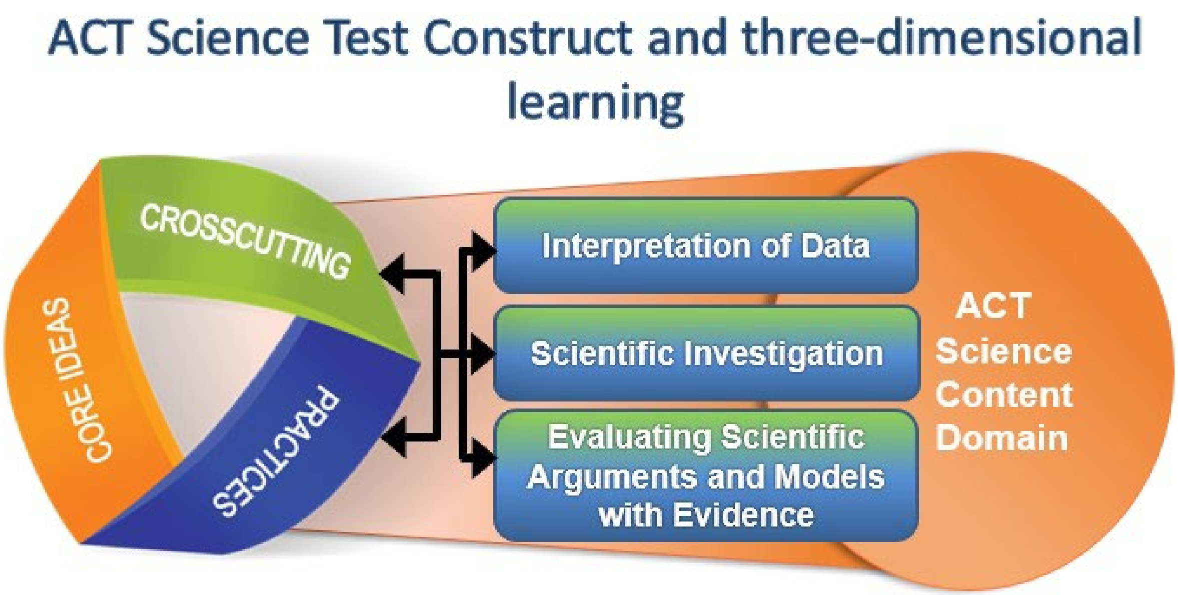
The enhanced ACT science section will continue to focus on the transferable science skills and practices represented in the legacy ACT reporting categories. Enhanced ACT science will continue to cluster key science and engineering practices and crosscutting concepts in science into three reporting categories: Interpretation of Data (IOD), Scientific Investigation (SIN), and Evaluation of Models, Inferences, and Experimental Results (EMI). The name of the EMI reporting category will be changed to Evaluating Scientific Arguments and Models with Evidence to better align with the terminology in state standards and science education. Interpretation of Data involves the ability to manipulate and analyze scientific data presented in tables, graphs, and diagrams. Scientific Investigation involves the ability to apply science knowledge, skills, and practices to understand scientific tools and procedures, comprehend the designs of scientific experiments, and compare, extend, and modify those experiments. Evaluating Scientific Arguments and Models with Evidence involves the ability to apply science knowledge, skills, and practices to evaluate the validity of scientific claims and models and formulate conclusions and predictions based on that information. Each reporting category has three subcategories in the legacy ACT science test (refer to Figure 1.2). The figure also illustrates how the foundation or root of every passage is the science content domains.

Figure 1.2. Science Reporting Categories and Subcategories for the Legacy ACT Science Test



The enhanced ACT science section aligns with the underlying strands of most state science standards that show a three-dimensional learning and assessment model of science. Science and engineering practices (what scientists do, SEP), crosscutting concepts (how scientists think, CCC), and disciplinary core ideas (what scientists study and research, DCI) are intertwined. The science content domain (DCI) is the foundation that allows the ACT test to present scientific phenomena so that students can engage in scientific sense-making with real data, experiments, and scientific explanatory models. Because of the large amount of overlap in the eight SEP and seven CCC categories in state standards, ACT has opted to cluster these domains into the reporting categories IOD, SIN, and EMI, which is similar to the frameworks of other large-scale science assessments like NAEP and PISA.

Figure 1.3. Relationship Between the Three Dimensions of Science Standards and the ACT Science Reporting Categories



The enhanced ACT science section will continue to code items to only the apex skill, the most cognitively complex skill used in answering the question. For example, a question asks students to evaluate a prediction and justify it with evidence; this requires the students to locate the appropriate data and identify the trend in the data (IOD skills), extrapolate the trend (IOD) to determine if it is consistent with the claim (EMI skill), and then justify their answer using a claim-evidence-reasoning scientific argument (EMI). The most cognitively complex skill is linking the claim to evidence in a scientific argument, so that item would be coded as EMI at the item level and the blueprint level even though other skills were involved.

This method of looking at the underlying components of the standards rather than the performance expectations has successfully met ESSA peer review and aligns with the postsecondary educator view that the transferable science skills and practices in these domains are more important than specific content knowledge (ACT,2012, 2016, 2020). It also aligns with the stated goal of helping high school graduates become scientifically literate, a goal common to state standards.

1.5.4.1.1 Reporting Categories

Table 1.12 shows the reporting categories and subcategories into which these skills and items are clustered. Two subcategories have been added for future development (SIN.ED and EMI.ED).

Table 1.12. Reporting Categories and Skill Areas for the Science Test

Reporting category	Skill area (subcategory)	Description of skills
Interpretation of Data (IOD)	—	Students manipulate and analyze scientific data presented in tables, graphs, and diagrams.
	Locating and Understanding (LU)	Students locate one or more pieces of data and understand features of graphs such as units, tables, legends, and axes.
	Inferring and Translating (IT)	Students use data from one or more graphs to identify trends, make inferences and comparisons, or translate data into other graphic formats.
	Extending and Reevaluating (ER)	Students make predictions based on trends in data.
Scientific Investigation (SIN)	—	Students understand experimental tools, procedures, and experimental designs. They compare, extend, and modify experiments.
	Locating and Comparing (LC)	Students locate, compare, and contrast information about one or more scientific investigations or experiments.
	Designing and Implementing (DI)	Students understand and evaluate aspects of experimental design such as methods, tools, variables, and controls.
	Extending and Implementing (EI)	Students make predictions about future experiments or experimental conditions and determine additional methods to improve or evaluate investigations.
	Engineering and Design (ED)	Students apply engineering and design thinking to investigations by, for example, identifying constraints, goals, or tradeoffs in an investigation or prototype.
Evaluation of Scientific Arguments and Models with Evidence (EMI)	Inferences and Results: Evaluating and Extending (IE)	Students judge the validity of scientific information and claims and formulate conclusions and predictions based on that information. Students also

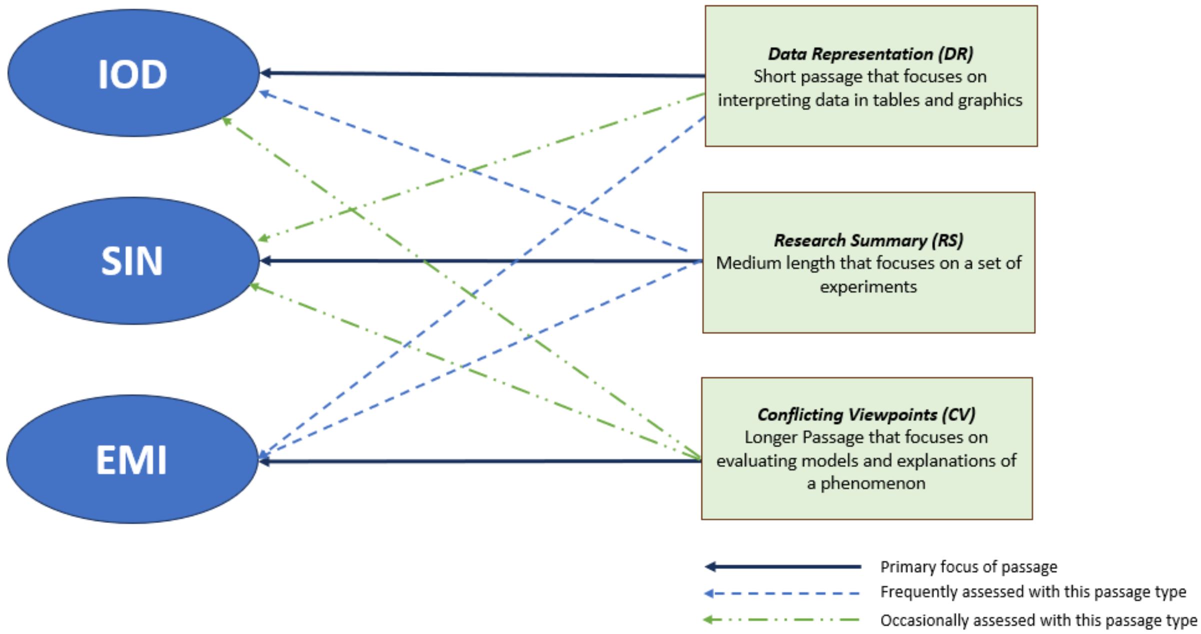
Reporting category	Skill area (subcategory)	Description of skills
		engage with scientific arguments using the claim-evidence-reasoning argument structure.
	Models: Understanding and Comparing (MU)	Students locate and compare information within a theoretical model or across competing models. (These skills are used only with Conflicting Viewpoints passages.)
	Models: Evaluating and Extending (ME)	Students evaluate and formulate predictions and hypotheses based on the examination of competing theoretical models. (These skills are used only with Conflicting Viewpoints passages.)
	Engineering and Design Thinking (ED)	Students evaluate claims about solutions, suggest new solutions or prototypes, and identify what evidence supports or refutes claims about the success of a potential solution or prototype.

1.5.4.2 Science Passages

The three science passage formats (Data Representation, Research Summaries, and Conflicting Viewpoints) are not changing for the enhanced ACT, nor are the primary targeted skills for each passage type. The mix of passage types allows ACT to sample skills in any content area using rich, authentic scientific scenarios.

Data Representation (DR) passages primarily focus on assessing Interpretation of Data (IOD) skills with some Evaluating Scientific Arguments and Models with Evidence (EMI) skills. Because these usually contain graphics with little text describing the investigations that collected the data, they rarely assess skills in the Scientific Investigation (SIN) domain. Research Summaries (RS) passages focus on how investigations were conducted, the research questions they sought to answer, and analysis of the results in the SIN domain. Because the experiments generate data, and the claims associated with the experimental hypotheses must be evaluated, RS passages are also able to commonly assess both the IOD and EMI domains. Conflicting Viewpoints (CV) passages present a central phenomenon and two or more conflicting models (causative explanations of that phenomenon). These focus largely on EMI skills, particularly in the EMI-ME and EMI-MU subcategories. Any passage type can be used to assess any of the four main science disciplines: biology (life science), physics, chemistry, and earth and space science.

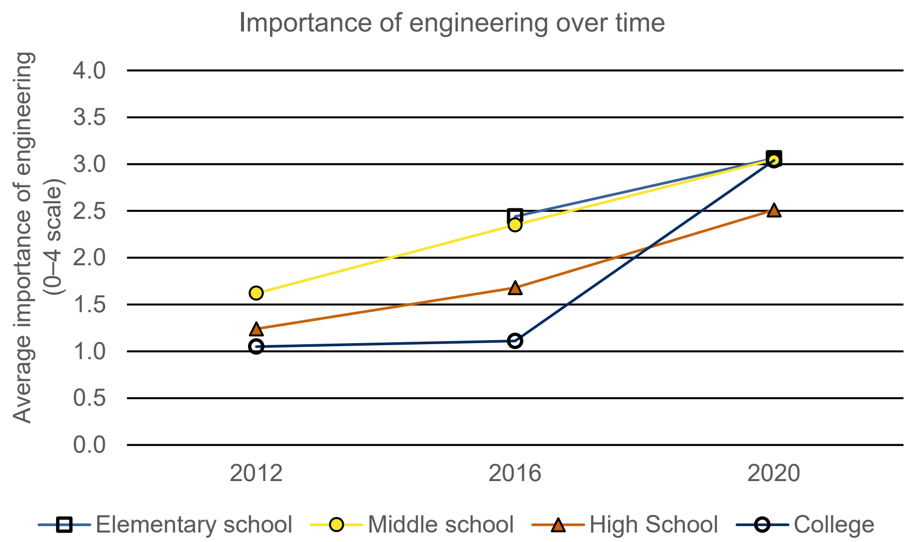
Figure 1.4. Science Reporting Category Relationships to Passage Types



We have changed how word count is determined for tables, graphics, and diagrams based on the results of ACT's 2019–20 eye-tracking study (Thomas & Victoria, 2023). Since students do not read every data point in a table, and students with strong graphic and data literacy skills focus on headings, legends, keys, and axes, the word count for graphics focuses on these and not on the data, which students generally scan when answering questions. This change will allow the passage to contain more context for why the phenomenon is being studied, which should increase engagement. Earlier cognitive labs (Thomas et al., 2025) found that students considered engaging scenarios easier even if the scientific content, graphics, and experiments were more complex. Adding context to the experiment or data will help make the Engineering and Design Thinking passages clear, as the problem being investigated can be more explicitly stated.

Passages will now have the potential to have a secondary content code for Engineering and Design Thinking (EDT; refer to Figure 1.2). Engineering and Design Thinking is explained in state standards as applying science to investigate or solve real-world problems. This will help the enhanced ACT science test achieve stronger alignment to common state standards for ESSA peer review. This secondary EDT code beyond science content (e.g., life science, physical science) will also allow ACT to track EDT materials in the pool and make EDT part of the enhanced ACT test blueprint with 1–3 EDT passages per form. (The legacy ACT contains engineering and design passages, but they are not part of either the blueprint or the sampling plan.)

Figure 1.5. Longitudinal Results of Importance of Engineering and Design Thinking From National Curriculum Survey



The 2020 NCS data showed a continuation of the increase in the perceived importance of teaching engineering and design thinking across all grade levels from lower elementary through postsecondary. However, 60% of the responding teachers reported that they rarely or never used activities specifically designed to teach these topics. After a few more years of educators teaching state science standards based on the NRC Framework for K–12 Science Education and Next Generation Science Standards, more students should have had the opportunity to learn these skills. These NCS data, combined with data from more recent alignment studies and peer reviews for ESSA, indicate that formally adding engineering as a targeted construct in the blueprint is appropriate.

1.5.4.3 Science Forms and Blueprints

A key consideration of the enhanced ACT science blueprint and form plan was how to maintain consistent coverage of all science content areas while also shortening the test. Given the wide variety of science course-taking patterns in high school, we determined that each form had to contain at least one passage from each scientific content domain. Additionally, since all states that require students to take a specific science course to graduate include a lab-based biology (life science) course, we decided that each form needed two life science passages. Simulations determined that 34 items could produce a test that would consistently meet reliability standards based on the item statistics in the pool.

1.5.4.3.1 Initial Enhanced ACT Science Content Blueprint

Table 1.13. Science Content Blueprint

Part I: Science content	ACT Legacy	ACT Enhanced
Passages (total number)	6	6 + 1 EFT (6 pretest items)

Part I: Science content		ACT Legacy	ACT Enhanced
Items (total number)		40	34 scored + 6 EFT
Reporting categories (# items)	IOD	16–20	13–17
	SIN	8–12	6–11
	EMI	10–14	8–13
Passage count by format	Data Representation	2	2
	Research Summaries	3	3
	Conflicting Viewpoints	1	1
Number of items per passage format	Data Representation	10–13	10–12
	Research Summaries	19–23	16–20
	Conflicting Viewpoints	6–8	6–7
Number of passages by content area	Earth/space science	1–2	1–2
	Life science: biology	2	2
	Physical science: physics	1–2	1–2
	Physical science: chemistry	1–2	1–2
	Engineering and design (secondary code)	Not tracked	1–3
Background knowledge items		4–8 (10%–20%)	5–8 (15%–24%)
Primary standard		Only one unit from a primary content standard	Only one unit from a primary content standard

Note: The numbers of items and passages given in the ACT Enhanced column are only for operational content; they do not include EFT content. Including the EFT passage, the maximum number of passages by content area will be three for life science and two for all other content areas.

1.5.4.4 Science Matrix Sampling Proposal

The science sampling proposal focused on the content domain, as the blueprint ensures that each form will have a mix of the transferable science and engineering practices across all three reporting categories predictive of college and career readiness and aligned to common state standards. The proposal used the content clusters in the NRC Framework and NGSS Appendix E, as shown in Table 1.14.

Table 1.14. Content Clusters Used for Matrix Sampling in Science

Science discipline	Content cluster
Physical science	PS1: Matter and its interactions
	PS2: Motion and stability: forces and interactions
	PS3: Energy
	PS4: Waves and their applications in technologies for information transfer
Life science	LS1: From molecules to organisms: structures and processes
	LS2: Ecosystems: interactions, energy, and dynamics

Science discipline	Content cluster
	LS3 Heredity: inheritance and variation of traits
	LS4: Biological evolution: unity and diversity
Earth and space science	ESS1: Earth's place in the universe
	ESS2: Earth's systems
	ESS3: Earth and human activity

Over a set of nine forms, three per year, all clusters would be assessed at least once. Because of the nature of science, some units may cover more than one cluster. For example, a unit on the effects of different colors of light on photosynthesis could target knowledge from LS1, LS2, PS3, and PS4. Some clusters would be sampled more than once because of their ubiquity across all three branches of science and prominent place in most science curricula, like energy (PS3) and matter and its interactions (PS1), which both overlap greatly with life science and Earth and space science. Most clusters would be covered more than once, but every cluster would be the basis of a passage at least once.

1.6 Overview of Initial Review and Decisions

As part of the periodic review of the ACT test, the assessment and domain sampling plans were evaluated, and the changes presented in this chapter were sent to external panels for review and potential revision. A summary of the proposed changes and rationales is shown in Table 1.15.

Table 1.15. Summary of Proposed Changes and Rationales

Change	Rationale
Reduction of total testing time, particularly for national test takers	Reduction of overall student stress in testing experience
Increase in time per item	Reduction of speededness and perceptions of speededness
Embedding field-test questions	Modernizing the test and following a best practice
Universal design and equity by design	Following a best practice and consistent with ACT's mission
Changes to passages (new topics, lengths, genres)	Alignment to common state standards
Changes to item types and weighting in blueprint	Alignment to common state standards
Reducing the number of options on math from five to four	Addresses speededness and maintains consistency across sections of the battery
Explore science optional	Reduction in overall testing time

Chapter 2: Panels, Initial Studies, and Revisions to Blueprints

In this chapter, we will discuss results from panels and initial studies conducted with the blueprints that were presented in Chapter 1. We used Steps 4 (gather data in the artifacts and evaluate) and 5 (revise the artifacts and improve the solution) from the Assessment Design Science framework (Thomas and Langenfeld; 2017a, 2017b; Langenfeld et al, 2019) to guide these studies and subsequent revisions. At the end of the chapter, we present the revised blueprints based on the results from these studies.

2.1 External Expert Panel Overview

To recruit participants for the expert panel reviews, each content area created a list of potential panel participants that would include experts from both K–12 and postsecondary education, such as high school teachers, college professors, and assessment designers. Additionally, each area recruited at least one expert from a traditionally underserved population. Overall, the panelists' expertise represented viewpoints that address key claims for the ACT® test: first, that it is predictive of college and career readiness; second, that it aligns to common state standards and is an appropriate assessment tool for accountability under the Every Student Succeeds Act (ESSA). Each of the content areas convened virtual panels of five experts plus ACT personnel during the winter of 2023–24. Panelists were paid for their participation.

Panelists were asked to evaluate the proposed changes to the legacy ACT as well as the portions of each section that were not changing. Prior to the panel, panelists were given information to review and comment on asynchronously, including information about the goals and constraints of the periodic review; the initial blueprints, specifications, and matrix sampling plans; retired forms of the legacy ACT; and samples of new items and passages. After asynchronous review, the panelists met remotely to discuss the potential changes and rationales as well as to make suggestions related to the enhanced ACT decisions. After the virtual panel, each panelist was asked to asynchronously provide additional feedback on their specific content area and the enhanced ACT project as a whole.

An additional review of the proposed changes was undertaken as part of the Universal Design for Assessment and Equity by Design initiatives, which included two or three panelists for each discipline who are experts in such areas of compliance as WCAG guidelines and universal design for assessment. These reviews were considered essential to minimize the possibility of construct-irrelevant variance, especially when delivering the test in multiple modes (paper and online). Accessibility experts met with each content team (ELA, math, and science) to give feedback on the Universal Design (UD) for testing aspects of the enhanced ACT. These experts clearly stated that adherence to the construct being assessed supersedes universal design constraints.

Both sets of panelists were asked to evaluate the proposed changes and recommend potential revisions based on the following criteria.

2.1.1 State and District Stakeholder Perspectives

Panelists were presented with information from voice-of-customer research as well as a summary of peer review results under ESSA. Some of the information dealt with the overall test (e.g., general issues regarding length and timing) while other information was subject specific (e.g., adding argumentative essays to English, adding engineering and design thinking specs to the science test). Panelists were asked to consider the alignment of each section to the relevant high school standards and the weighting of each content cluster in the blueprints.

2.1.2 Postsecondary Stakeholder Perspectives

Panelists were also asked to consider the needs of postsecondary stakeholders in recruitment, admissions, and retention of students. Questions included the following: What skills are most important to college readiness? How do postsecondary stakeholders use the current information (composite scores, section scores, cluster scores [ELA or STEM], and Superscoring)? What impacts on the concordance tables and raw-to-scale conversions would be acceptable to this group? What information is needed for placing students in course sequences?

2.1.3 Parent and Student Stakeholder Perspectives

Panelists were asked to analyze what impacts the proposed changes might have on student performance, particularly regarding efforts to reduce speededness at the item level and overall testing time. They were also charged with evaluating the effect of proposed changes on students from traditionally underserved populations (e.g., Black, Latino, Indigenous, low-income, and first-generation students) and detailing concerns parents and students might have.

2.1.4 Communication Across Stakeholder Groups

Panelists were tasked with suggesting how ACT should most effectively communicate the changes, the reasons for the changes, and the evidence supporting the decision-making process to state and district, postsecondary, and family stakeholders.

2.2 External Expert Panel Results

2.2.1 ELA Panels: English

Overall, the panelists were in favor of the new proposed blueprint, indicating that the changes seem to reflect the work students are doing in high school classrooms and are expected to do in college. Feedback from the reviews is organized at the passage, item, blueprint, and matrix sampling levels.

2.2.1.1 English Passages

Panelists were strongly in favor of adding argumentative essays to the English test, as argumentative writing is a major focus in Grades 11 and 12 and is critical to college success. They were also strongly in favor of including shorter essays alongside essays of the current length, noting that shorter essays might feel more accessible to students, ease stress and fatigue, reduce attention span demands, and increase student confidence. Panelists also noted that the sample items appearing with short essays did not appear to represent a decrease in complexity and stressed that maintaining appropriate rigor would be important. Panelists suggested that ACT select topics for short essays carefully because some sensitive or complex topics, especially those involving traditionally underserved or oppressed groups, would be better addressed in longer essays, where it would be easier to avoid abstractions or oversimplifications that could disadvantage some readers.

2.2.1.2 English Items

As well as favoring the addition of argumentative essays, panelists were strongly in favor of adding items that test argumentative skills. They noted that claims, supporting evidence, and counterclaims should all be explored in English items, and they reviewed common argumentative vocabulary for terms that would be most accessible to all students for use in enhanced ACT item construction.

The panelists strongly supported adding stems to all English items, noting that the addition may help minimize bias for students from traditionally underserved groups without adding to the reading load for the section. They reviewed sample stem language and agreed that the proposed standard stem language was clear and appropriate to the tasks.

Panelists also strongly supported removing item types that might disadvantage English learners (e.g., idiom items) and those that may present accessibility issues from the standpoint of universal design for assessment (e.g., items framed with “not” language).

2.2.1.3 English Blueprints and Specifications

The panelists were in favor of reducing the number of items on the English test and felt that the new form blueprint maintained enough items to sample the domain effectively. They were also in favor of giving students more time per item to reduce potential speededness issues. In terms of the blueprint, there were no significant concerns with the plan to move to embedded field test (EFT) items, and panelists thought the move to six or seven essays (with the inclusion of shorter essays and EFT units) might prove beneficial to students for the reasons noted above. They felt that the increase in the number of essays would be offset by shorter passages and fewer items.

The changes to the percentages of items in each reporting category were generally viewed favorably. The reduction in the number of Conventions of Standard English (CSE) items, which include sentence structure, punctuation, and usage items, was viewed as having multiple potential benefits, including reducing bias for English learners and speakers of nonstandard dialects. This reduction also allows for greater focus on Production of Writing (POW) and Knowledge of Language (KLA) skills, which the panelists agreed are more important to student success in college. At the same time, panelists did not want the CSE reporting category to be eliminated entirely. K–12 educators, especially, were concerned about reducing these items too much, noting that the amount of instructional time spent on these skills can differ between high schools and that students continue to struggle with important grammar skills. Though the panel represented a small sample of educators, it was suggested that K–12 and postsecondary educators may differ in their feelings about the importance of testing grammar; this will be further investigated in the next National Curriculum Survey.

As noted, panelists agreed that the skills in the Production of Writing (POW) and Knowledge of Language (KLA) reporting categories are most important for both K–12 and postsecondary success, but there was some concern that increasing the number of items in these categories may offset the additional time per item allowed by the new blueprint, as these items can be more time-consuming than grammar items. The panelists agreed that the planned timing study could address these concerns (refer to Chapter 4).

Panelists were also in favor of a blueprint that ensures that all forms have a mix of argumentative, informational, and narrative essays, as this reflects both classroom experience and postsecondary expectations. They suggested even more argumentative content, if possible, and a blueprint of two argumentative, two informational, and one narrative essay per form. Since the proposed blueprint (one or two argumentative essays, two or three informational essays, and one narrative essay) allows for the ideal mix suggested by panelists and flexibility is necessary as ACT builds a pool of argumentative content, blueprints were not changed.

2.2.1.4 English Matrix Sampling Proposal

The panelists thought the sampling proposal was an acceptable way to appropriately sample the domain across forms. As part of this discussion, panelists also evaluated individual skills within each reporting category, discussing whether the sampling proposal resulted in any skills being over- or underrepresented. There were no concerns that necessitated a change to the sampling proposal or blueprints, but suggestions for new item types have been taken into consideration for item development and were incorporated into the cog lab discussed in Chapter 3. The importance of these skills will continue to be explored through the next National Curriculum Survey.

2.2.1.5 Applying Universal Design Principles to Testing

The UD panelists were in favor of applying principles that do not interfere with the English constructs being assessed. They noted the need to balance clarity and word count and offered suggestions for how items might be worded or formatted to increase accessibility. Suggestions also included ways to address technical concerns beyond the scope of the content changes (e.g., maintaining passage highlighting in passage-based test sections when moving from item to item). ACT will continue to examine these suggestions as we strive for continuous improvement in all aspects of test taking.

2.2.2 ELA Panels: Reading

The panelists were largely in favor of the proposed changes to the test, noting that the necessary rigor of the reading test would be maintained even with fewer items and added time per item. All panelists noted that fewer items and more time would benefit students. There were concerns about the sampling proposal, which are detailed below.

2.2.2.1 Reading Passages

The panelists discussed passage length and complexity in depth. While they acknowledged that many students lack reading stamina and have reduced attention spans, they agreed that the ability to read complex multiparagraph texts is a critical component of college and career readiness, as postsecondary success requires engagement with longer complex texts. Panelists noted the large gap between expectations for high school and first-year college students, particularly for traditionally underserved populations who are still coping with COVID learning loss. As such, panelists felt it was valuable to maintain 750-word passages while introducing some 650-word passages. ACT was urged not to reduce reading passage length further.

It was acknowledged that maintaining passage length means having fewer scored passages on a form, which limits opportunities for diverse representation. However, the authenticity and specificity of the longer published passages was noted as a positive, and panelists were encouraged by the work done through Windows and Mirrors, an ACT initiative to find and feature passages that center traditionally underserved populations in positive ways (further discussed in Chapter 1).

Panelists were also given twelve sample reading passages and asked to assess text complexity. The panelists agreed that all sample passages were of appropriate complexity for the reading test, even while there were differences of opinion about how ACT's qualitative complexity rubric might be applied and at which text complexity level individual passages might fall. In addition to reviewing sample passages, panelists evaluated ACT's approach to assigning text complexity, which incorporates the text complexity triangle (National Governors Association for Best Practices & Council of Chief State School Officers (2010) Appendix A), and requires evaluation of three facets: qualitative, quantitative, and reader/task. The panel's feedback prompted an examination of ACT's qualitative and quantitative guidelines to ensure that there is an appropriate progression between each text complexity level, especially between the "complex" and "highly complex" levels. As a result, the entire pool of ACT reading passages eligible for enhanced ACT forms was evaluated by ACT ELA staff, and quantitative guidelines were revised to help ensure appropriate leveling. Feedback to inform the qualitative rubric will be included in the next NCS as part of the continuing improvement processes.

2.2.2.2 Reading Items

In addition to reviewing sample forms, panelists examined a variety of Visual and Quantitative Information (VQI) items and passages. All panelists agreed that reading graphics alongside text is an important skill for both college and career readiness and should be measured. They noted that while these skills may not be widely addressed in literature classes, students are expected to encounter graphics in other reading-heavy classes. Additionally, they noted that these skills are included in common ELA standards. As a whole, the sample VQI items were viewed as appropriately complex.

2.2.2.3 Reading Blueprints and Specifications

As noted, panelists were in favor of reducing the number of reading items and adding more time per item. They were also largely in favor of reducing the percentage of Key Ideas and Details

(KID) items, especially the effort to eliminate DOK 1 KID items that may be difficult for construct-irrelevant reasons. There was some concern that KID items may be an access point for emerging readers and that the reduction may reduce the number of questions these readers are able to answer; however, it was acknowledged that this concern must be balanced with the goals of assessing achievement of common ELA standards and college and career readiness.

Panelists understood the requirement to move to EFT items and approved of the change. There were some concerns that scores based on only three operational passages leave a relatively high percentage of EFT items on a form; however, panelists understood that this is a necessity given the passage-based nature of the section and the desire to not lengthen the overall test.

2.2.2.4 Reading Matrix Sampling Proposal

Panelists were concerned about the proposal to have either a scored VQI passage or a scored paired passage on a given form. The panelists perceived that paired passages are more difficult overall, requiring a higher cognitive load from readers. All agreed that a form that featured two single passages, one VQI passage, and one paired passage would likely be more challenging than a form that featured three single passages and one VQI passage. However, during discussion of the two sample forms (one with two single passages, one VQI passage, and one paired passage and the other with three single passages and one VQI passage), opinions differed as to the comparability of the forms. There was no consensus as to whether one form, if either, was more difficult, with a few panelists noting that the forms seemed comparable. One panelist noted that increased familiarity with VQI units may decrease the perceived difference between forms, which reinforces the need to increase VQI content in test prep and elsewhere. Since these two sample forms were not built according to statistical difficulty specifications, the perceptions may come from the difficulty of the item set as well as the passage type.

As a result of panel discussions, both a paired passage and a VQI passage were selected for the cognitive lab research that came in the next stage of gathering evidence. In addition, ACT's research team conducted analyses of VQI unit performance, using legacy ACT forms and the June study forms discussed in Chapter 4. Results were consistent with previous studies indicating that VQI items do not exhibit significant statistical differences from other items. These analyses will be repeated with more data to confirm findings.

Panelists offered several suggestions for how ACT might reduce the perceived difference while maintaining the sampling plan, including ensuring that graphics in VQI units are well connected to the passage and establishing form specifications that require a set number of synthesis or VQI items per form. Panelists agreed that it will also be important for ACT to offer more VQI practice materials so students can gain familiarity with these item types on the reading test.

Panelists also stated there is a need for clearer communication about the complexity of the passage; the mix of passage complexity levels on a form; and the interplay of the complexity of the passage, the cognitive demands of the skill being tested in a single item, and the item set as a whole. Consequently, ACT will need to communicate more clearly for both the reading and science tests that passages of any complexity level can support questions that also vary in complexity, whether that be quantified with something like Webb's DOK, Bloom's taxonomy, the College and Career Readiness Standards, or some other metric.

2.2.2.5 Applying Universal Design Principles to Testing

Panelists were in favor of principles that do not interfere with the reading constructs being assessed. There was a lengthy discussion about accessible vocabulary, and these comments will help inform item development and passage selection for the enhanced ACT. Internal processes for reviewing UD for VQI graphics were acceptable.

2.2.3 STEM Panels: Math

Generally, the panelists approved of the proposed changes to the enhanced ACT math section.

2.2.3.1 Math Items

The panelists felt that reducing the number of response options from five to four was a positive. This could reduce potential bias for English learners by decreasing the reading load in some items. This would also reduce the amount of time needed to solve some items, particularly when students work backward from the answers. Panelists did not interpret this reduction as simplifying the test or making the test easier since the three wrong answer options could include many common misconceptions and errors for any given item. It was noted that this could also streamline items for which a fourth plausible foil is difficult to generate.

2.2.3.2 Math Passages

Panelists did not consider math passages because these were previously removed from the test.

2.2.3.3 Math Blueprints and Specifications

The panelists were in favor of fewer items with more time per item. They did not believe that this diminished the rigor of the test. They thought that the changes could reduce test anxiety and fatigue while potentially increasing student confidence. One panelist was concerned that 41 scored items was too few; however, the other panelists thought that 41 operational items were sufficient. The panelists' evaluation that the test did not lose rigor ("was not dumbed down") was an important piece of support against a potential perceived weakness of the shorter test.

The panelists were in favor of reducing the number of Integrating Essential Skills (IES) items on a form. They thought this may reduce bias for English learners and others as IES items typically get more DIF flags than items in other reporting categories. Postsecondary instructors expressed concern that too many IES items were being removed from the test. There were differences of opinion between high school and postsecondary educators about the importance of IES skills. This mirrors findings from recent National Curriculum Surveys (2016, 2020) in which the postsecondary instructors supported the use of IES for targeting critical thinking and problem-solving skills that were rated as extremely important for success in first-year math courses. This could be further researched in future iterations of the NCS.

Panelists agreed that the slight decrease in the percentage of items in context was consistent with the overall goals of the enhanced ACT math section. Panelists acknowledged the difficulty of balancing the additional length and reading load of these items with the complexity of tasks needed for college readiness.

The panelists saw the reduction in the percentage of advanced items as a strong move from an equity perspective, as this may reduce bias for lower achievers in high school or those with less opportunity to access advanced math classes. This echoed the research by Flores (2007) and Bryant (2015) mentioned in Chapter 1.

Panelists understood the need for embedded field testing. The plan for the number of EFT items was acceptable.

The panelists agreed that reporting category names and structures are understandable to parents and align well with common math standards and that the score report should continue to provide usable information for students by reporting category.

The panelists stated that ACT should continue to ban calculators with computer algebra systems (CAS) as that would change what the test is measuring. They supported the use of calculators in general, as students can use a tool that they are familiar with on all math items.

The use of formula sheets was seen as a more complex issue. Some of the panelists supported providing a formula sheet while the majority did not. The argument for including a formula sheet is that in the real world, students can look up formulas. However, that contrasted with the realities of timed high stakes assessments. A formula sheet can take up desk or screen space. It can be more time-consuming and complex to look for the formula on the sheet than to recall the formula. It would be difficult to decide what to put on the sheet, as that might imply that these are the only things that will be tested. The decision was made to continue with the same method as the legacy ACT. There will be no formula sheet, although items with unusual formulas (e.g., the law of sines) will be scaffolded with that information.

The panelists supported the initial enhanced ACT math blueprint. They suggested increasing the number of items in the Statistics and Probability reporting category to reflect college course-taking trends and the increased demand for skills in data analysis and data science. There was no consensus on which reporting category should have an item removed. The updated enhanced ACT math blueprint will increase the number of Statistics and Probability items from four or five to five or six. To compensate, the Algebra, Functions, and Geometry reporting categories may have seven or eight items each (23 total) as opposed to the eight items each in the initial blueprint.

2.2.3.4 Math Matrix Sampling Proposal

The panelists were generally supportive of the proposal. There was discussion about whether the placement of topics in the three tiers discussed in Chapter 1 was appropriate. The panelists stated that the two sample forms were both good assessments separately and that the overall depth and breadth of topic coverage was better when both forms were considered together. The topic tiers continue to be reviewed and revised. The next National Curriculum Survey will provide more feedback related to topics that are difficult to place in tiers.

2.2.3.5 Applying Universal Design Principles to Testing

Content panelists were in favor of principles that do not interfere with the math constructs being assessed. Panelists shared comments about scaffolding and the appropriateness of certain contexts; these comments were later shared with the math content team. The panelists recommended that scrolling on math items, particularly those with graphics, be minimized for online testing.

2.2.4 STEM Panels: Science

The science panel was generally in favor of the initial enhanced ACT science blueprint. Panelists were strongly concerned that making the science test optional would devalue science education and hinder the goal in state standards that all students be scientifically literate. Several panelists echoed the common idea that only what is tested will be taught and felt that this change may have unintended consequences for science education. In a post-COVID world with misinformation abundantly available on the internet, it is important that all students be able to engage in scientific sense-making. The concern that only what is tested will be taught echoes findings about time allowed for science instruction in lower elementary grades when science is not tested (ACT, 2016, 2020).

2.2.4.1 Science Items

Panelists were strongly in favor of adding the new Engineering and Design skills under the Scientific Investigation (SIN) and Evaluating Scientific Arguments and Models with Evidence (EMI) reporting categories. They suggested that these skills, which relate to applying science to real-world problems, connect strongly to college and career readiness and common state standards and should therefore be a key part of communications to stakeholders.

Panelists were strongly in favor of clustering the Science and Engineering Practices and Cross-cutting concepts in science into the three reporting categories. They also stated that items were able to test the targeted skills.

Panelists were strongly in favor of the increased percentage of three-dimensional items that required background science knowledge. They also stated that some items in a passage should focus on the big ideas and the “So what?” aspect of scientific sense-making. This led ACT science staff to examine and update some of the internal guidelines for overlap among items in a set. This will also lead to targeted revisions of existing passages to add context to passages and to add more three-dimensional items to these units for use in enhanced ACT forms.

Panelists would prefer the use of some constructed-response or composite items (items that mix forced choice and constructed response) because multiple-choice tests cannot assess all the science standards, particularly those that involve the student generating a scientific argument or model. They suggested exploring the scalability of these items using AI for even paper-based assessments. These ideas were passed on to the AI research group.

2.2.4.2 Science Passages

Panelists supported the interpretation of Engineering and Design Thinking (EDT) based on the NRC framework. They did posit that some stakeholders, particularly postsecondary instructors who are less familiar with state standards, will not consider this the same as traditional engineering. They suggested targeted communication to explain to all stakeholders that EDT is applying scientific knowledge and methods to solving real-world problems and not just traditional mechanical, civil, or chemical engineering.

Panelists stated that the amount of text used is appropriate to introduce the phenomena and activate scientific sense-making. They also stated that giving the scientific scenario a real-world context might increase engagement and further suggested including passages that might address problems in a variety of locations (e.g., inner city, suburban, rural) that students might connect with. Care must be taken to balance the desire to give context and spur interest with the risk of adding too much to the reading load.

Panelists said that the variety of passages used on the enhanced ACT science test allows for a variety of scenarios and skills to be tested.

2.2.4.3 Science Blueprints and Specifications

Panelists were strongly in favor of increasing the time per item as they thought that many students would struggle to finish the 40 items on the legacy ACT science test in 35 minutes. Furthermore, increasing the time per item would better allow students to engage in authentic scientific sense-making. However, adding a seventh passage (the EFT passage) could counteract the increase in time for the science test. Panelists were presented with both a 45- and a 40-minute option for the enhanced ACT science test. They were concerned that adding only 5 minutes to the science test would not accomplish the desired reduction in speededness. They strongly supported the 45-minute plan; however, strong evidence from the timing study could allay panelists' concerns that a 5-minute addition of time would be offset by adding an additional passage.

Panelists reported that the proportion of items in each reporting category was appropriate and would allow students across ability levels to demonstrate what they know and can do. The mix of content areas and passage types also allows students to demonstrate their knowledge and skills in scientific sense-making across the breadth and depth of the science domain.

2.2.4.4 Science Matrix Sampling Proposal

The panelists supported the sampling proposal. Many of the passage and item sets contained more than one content cluster, which was viewed as a strength because science is often interdisciplinary. Some panelists thought that certain clusters (specifically PS4 waves) could be sampled less frequently as they may not be taught as often. This is another area in which the next National Curriculum Survey could better inform future plans.

2.2.4.5 Applying Universal Design Principles to Testing

Panelists were in favor of principles that do not interfere with the science constructs being assessed. There was some concern among the panelists about the amount of scrolling that would be necessary for online testing.

2.2.5 Impacts of Panels on Cognitive Labs and Other Studies

The comments of the panelists guided the selection of items and passages for the cognitive lab study. Specifically, some culturally relevant math items (Anguiano-Carrasco et al., 2025) were added to address concerns about the reading load of items with culturally relevant context. The new engineering and design thinking items were included to determine whether they elicited evidence of those specific new constructs. Additionally, parallel math items were used with four or five foils to determine whether they elicited different evidence of student behaviors. Both a paired passage and a VQI passage were included to gauge relative student performance, perception of difficulty, and interest level. These decisions and others will be addressed in full in Chapter 3 of this document.

A list of topics and questions for the next iteration of the National Curriculum Survey also arose from these panels. Insights from a large sample of educators across K–12, postsecondary education, and workforce will better inform decisions and communications about the enhanced ACT.

2.3 Panel Impact on Final Content Blueprints

The panel recordings were made available to content specialists in all content areas. The content specialists continue to use the feedback from the panel to guide the enhanced ACT test development process from topic selection for passages through item development, form building, and sampling across forms.

2.3.1 Overall Results

The panels generally supported the changes that were proposed for the new enhanced ACT and the matrix sampling protocols. Some minor changes were made to blueprints, and the post-panel blueprints that were used for the initial field test and timing study are shown below.

2.3.2 Revised Blueprints Post Panel

The following represent the content blueprints used to create the two forms for the June 2024 enhanced ACT study. Results from this study are described in Chapter 4. They have minor differences from those shown in Chapter 1, reflecting the changes made in response to panelists' comments (e.g., increasing the number of probability and statistics items in math).

2.3.2.1 English

Table 2.1. English Content Blueprint

Part I: English content		Enhanced ACT
Total testing time		35 minutes
Items (total number)		40 (+10 EFT)
Reporting categories (operational items)	Production of Writing	15–17 (38%–43%)
	Knowledge of Language	7–9 (18%–23%)
	Conventions of Standard English	15–17 (38%–43%)
Passages (operational)		5 (+1–2 EFT)
Writing type (# passages)	Informational	2–3
	Argumentative	1–2
	Narrative	1
Genre (HUM, SSC, NSC, etc.)		0–2
Length	340 standard words (approx.)	3
	185 standard words (approx.)	2
Items per passage	340-word passages	10
	185-word passages	5
Standard word count (operational passage total, approx.)		1,390

2.3.2.2 Math

Table 2.2. Math Content Blueprint

Part I: Math content		Enhanced ACT
Total testing time		50 minutes
Total items (ordered in increasing difficulty)		41 scored and 4 EFT = 45 total
Reporting categories (# items)	1. Preparing for Higher Math (PHM)	33 (80%)
	1a. Number and Quantity	4–5 (10%–12%)
	1b. Algebra	7–8 (17%–20%)
	1c. Functions	7–8 (17%–20%)
	1d. Geometry	7–8 (17%–20%)
	1e. Statistics and Probability	5–6 (12%–15%)
	2. Integrating Essential Skills (IES)	8 (20%)
	3. Modeling	≥8 (≥20%)

2.3.2.3 Reading

Table 2.3. Reading Content Blueprint

Part I: Reading content		Enhanced ACT
Total testing time		40 minutes
Items (total number)		27 (+ 9 EFT)
Reporting categories (operational items)	Key Ideas and Details	12–14 (44%–52%)
	Craft and Structure	7–9 (26%–33%)
	Integration of Knowledge and Ideas	5–7 (19%–26%)
Passages (operational)		3 (+1 EFT)

Part I: Reading content		Enhanced ACT
Genre (# passages)	Literary narrative	1
	Informational (humanities, social science, natural science)	2
Format (# passages)	Single passage	2
	Paired passage	0–1
	VQI passage	0–1
Length (# passages)	750 standard words (approx.)	2
	650 standard words (approx.)	1
Items per passage		9
Standard word count (operational passage total, approx.)		2,150

2.3.2.4 Science

Table 2.4. Science Content Blueprint

Part I: Science content		Enhanced ACT
Passages (total number)		6 + 1 EFT (6 pretest items)
Items (total number)		34 scored + 6 EFT (1 passage)
Reporting categories (# items)	IOD	13–17
	SIN	6–11
	EMI	8–13
Passage count by format	Data Representation	2
	Research Summary	3
	Conflicting Viewpoint	1
Number of items per passage format	Data Representation	10–12
	Research Summary	16–20
	Conflicting Viewpoint	6–7
Number passages by content area	Earth/space science	1–2
	Life science: biology	2
	Physical science: physics	1–2
	Physical science: chemistry	1–2
	Engineering and design (secondary code)	1–3
Background knowledge items		5–8 (15%–24%)
Total word count		NA
Primary standard		Only one unit from a primary content standard

Note: The numbers of items and passages given in the ACT Enhanced column are only for operational content; they do not include EFT content. Including the EFT passage, the maximum number of passages by content area will be three for Life Science and two for all other content areas.

2.3.3 Sampling Proposal Decisions Post Panel

Given the general agreement from panelists that individual forms in all four content areas were able to adequately sample the domain and that multiple forms were able to provide a greater sample of the breadth and depth of skills aligned to state standards and college and career readiness, it was decided that formal matrix sampling proposals were not needed for the enhanced ACT. Forms and groups of forms will continue to be built with content sampling plans that ensure appropriate coverage of standards at the individual test form level. Content sampling using multiple forms will continue as the enhanced ACT becomes operational, and a formal matrix sampling plan may be implemented in the future, if needed.

2.4 Summary of Panels and Future Plans

Across all four domains, the external expert panels supported the proposed changes. All panels stated that the new items, passages, and blueprints would achieve the stated objectives of measuring college and career readiness and assessing student achievement of common high school standards. They stated that the sample forms in all domains were reasonable samples of the corresponding domain and that looking at multiple forms improved the depth and breadth of the coverage. Some of the minor changes that were incorporated were increasing the number of statistics items on math by one, revising text complexity guidelines for reading, and changing item overlap procedures in science.

The panelists provided many suggestions for what additional information should be collected to strengthen the argument for the changes made to the enhanced ACT; many of these suggestions were incorporated into cognitive labs or other research activities, such as the timing study. Other comments from the panelists have been catalogued so that they can be used in the upcoming National Curriculum Survey, particularly for areas in which the postsecondary and K–12 panelists disagreed. Additional information from thousands of educators in each group will be used to better inform any future revisions to the enhanced ACT.

Chapter 3: Validity Evidence Based on Test Content and Response Processes

According to the *Standards for Educational and Psychological Testing* published by the American Educational Research Association (AERA), American Psychological Association (APA), and National Council on Measurement in Education (NCME) in 2014, the five primary sources of validity evidence are content, response processes, internal structure, relations to other variables, and consequences of testing. Together, these sources of validity evidence support the intended uses and interpretations of test scores.

The ACT has two main claims that serve as the foundation of its validity argument. First, ACT scores are predictive of college and career readiness. Second, ACT scores are an indicator of mastery of common high school standards. Validity evidence supporting both of these intended uses comes from alignment of the skills and content to what is taught at both the high school and postsecondary levels. Further support for their intended use comes from evidence of the cognitive processes that test takers (generally high school students) use when they answer questions correctly or when interacting with the material in ways that prepare for answering questions (reading passages, item stems, item options, etc.). Additionally, the content should be deemed appropriate to the targeted skill level of the test takers and claims.

Validity is considered a unitary and ongoing process. Some pieces of validity evidence represent early steps in gathering information and data to make initial design decisions and create artifacts related to new item templates, passage guidelines, and content specifications. Other sources of evidence were collected in the later steps of the iterative design process. We conducted research, e.g., cognitive labs, to determine if the design changes elicited evidence of the targeted skills, both new and old. We also looked for evidence of a reduction in construct-irrelevant variance. This chapter provides an overview of all validity evidence based on test content and response processes that support the intended use of the enhanced ACT scores.

3.1 Evidence Based on Test Content

Sources of evidence based on test content are crucial for establishing the validity of a test. These sources ensure that the test items are representative of the domain being assessed and the content aligns with the intended constructs being measured. The ACT content blueprints are created to ensure that each form samples from the same domain of skills at appropriate levels of difficulty. This ensures that each test allows students to demonstrate what they know and can do across the spectrum of abilities for high school juniors and seniors.

In this section, we describe multiple sources of test content evidence collected related to the enhanced ACT® test: evidence of alignment and content validity, and artifacts from the item review process, including bias and sensitivity reviews.

3.1.1 Alignment to Common Standards and Peer Review Requirements

To support claims required to achieve “substantially meets” or “fully meets” requirements for Every Student Succeeds Act (ESSA) accountability and peer review, the ACT must be aligned to the standards that it is used to assess. Although each state has its own set of standards, the majority of the content of those standards within a given domain is shared. The ELA and math standards are commonly based on work from the Council of Chief State School Officers (CCSSO), which used data from the ACT and SAT to identify upper anchors for college and career readiness in 12th graders. Over 90% of states use standards that are based on the NRC Framework for K–12 Science Education (NRC Framework) or the related Next Generation Science Standards (NGSS).

The legacy ACT has fully met the rigorous technical requirements for assessments used to measure achievement under the ESSA, as evaluated by the U.S. Department of Education's (USED) assessment peer review process. ACT's track record of success is demonstrated by the states that currently have approval to use the legacy ACT as their accountability measure. As of 2024, eight states use the ACT under ESSA as their high school achievement indicator. Nearly all these states use math and ELA/literacy standards that are adapted from the Common Core State Standards (CCSS) and science standards adapted from the NGSS or the related NRC Framework. The legacy ACT is the only college entrance exam to have received a "fully meets requirements" designation in USED's peer review process, which evaluates comprehensive evidence of the soundness and validity of the assessments that a state uses for accountability.

3.1.1.1 ELA Alignment to Common Standards

The ACT focuses on the general content areas of college and high school curricula and on the knowledge, skills, and abilities (KSAs) represented in common state standards. Although not designed to measure any single state's set of educational standards, the ELA sections of the enhanced ACT offer several features that account for their strong alignment with rigorous academic standards across different states.

The enhanced ACT measures mutually supportive reading, English language, and writing knowledge and skills with rich passage-based tasks that report several scores directly linked with most states' clusters of ELA standards. To accomplish this, the reading section uses authentic literary and informational texts drawn from contemporary and historical U.S. and world literature sources. Every reading test includes a mixture of informational passages on engaging and diverse topics from the humanities, social sciences, and natural sciences, and includes passages that require students to integrate information presented from multiple sources. The enhanced ACT English test features essays in all three writing types: narrative, informational, and argumentative/persuasive. Furthermore, when administered with the optional writing component, the enhanced ACT includes a direct writing task that requires students to engage critically with other perspectives and develop their own thesis on a contemporary issue. The analytic scoring rubric links directly to expectations in most common writing standards.

As part of the review process for the enhanced ACT, the ELA team reviewed alignments conducted between 2015–2019 related to ESSA peer review and performed a comprehensive review of common state standards. Because alignment studies are subjective and different methodologies can result in different outcomes, the team attempted to address several common concerns regardless of the alignment model used.

3.1.1.1.1 Argumentative Writing (English Test)

Data from the two most recent ACT National Curriculum Surveys (NCS) (ACT, 2016; 2020) indicate that the knowledge and skills requiring students to identify and evaluate claims and evidence are deeply important to college readiness across all subject areas and are commonly taught across all disciplines in high school. Since the ACT writing test is an optional component, there have been fewer opportunities for students to demonstrate their mastery of writing skills related to argumentation and persuasion. Consequently, alignment studies conducted for ESSA peer review for the legacy ACT indicated that the alignment to common state writing standards for argumentative writing was relatively weak. To address these concerns, the decision was made to introduce an argumentative essay on each ACT English form to target important skills such as developing a thesis or claim and identifying evidence that supports or challenges a claim.

Moving forward, every enhanced ACT English test will feature at least one essay in each category: informational, argumentative, and narrative. The balance of essay types on each form is intended to reflect how most state standards approach writing types at the Grade 11–12 level, which is also in accordance with the NAEP writing framework, with informational and argumentative writing representing the bulk of student writing at this level. In making this change, ACT will not only be better able to highlight important argumentative skills but will also allow students to demonstrate key writing skills and strategies for a variety of purposes on each test form. These changes should lead to stronger alignment with common writing standards.

3.1.1.1.2 Integration of Knowledge and Ideas (Reading Test)

Data from the most recent NCSs indicate a continued need for students to read a variety of text types and genres to be college ready, including passages with graphic and quantitative elements. In fact, “a large majority (82%, on average) of ELA and SS instructors agreed or strongly agreed that ‘reading tests that measure readiness for college-level reading should include passages with graphic and quantitative elements (diagrams, graphics, charts, tables)’” (ACT, 2020). Moreover, the majority of state standards call for engagement with these texts. Since 2018, several small changes have been made to the legacy ACT reading test to strengthen Integration of Knowledge (IKI) representation and address concerns raised during independent alignment studies, including increasing the number of IKI items per form, revising development task models for “claims and evidence” items, and incorporating passages with graphic and quantitative elements. It was important that the enhanced ACT not take steps backward in any of these areas with a reduced test length. Though the overall number of items will be reduced, IKI will increase from 15–23% of the test to 19–26%. The domain sampling plan will ensure that there is depth and breadth of IKI coverage over multiple forms.

3.1.1.2 Math Alignment to Common Standards

While not designed to measure any single state’s standards, as a curriculum-based test designed to measure what students learn in high school, the ACT is well aligned to rigorous math standards across different states. Internal alignments from state standards to skills assessed on the legacy ACT math test consistently show over 95% of state standards are in the domain of the ACT math test. As part of the review process for the enhanced ACT, the math team reviewed ESSA peer review alignments conducted between 2015–2019 and performed a comprehensive review of common state standards.

The review found that gaps were generally observed in areas that would be difficult to assess in a large-scale summative assessment (e.g., standards asking students to analyze data reports or complete long-term projects) or caused by using only forced-choice items. Given the constraint of only using multiple-choice items on the enhanced ACT, these concerns could not be addressed. However, concerns about the high percentage of items based on math topics introduced before high school (IES reporting category) were addressed by reducing the percentage of the IES category from 40% to 20% and increasing the Preparing for Higher Math (PHM) reporting category from 60% to 80%. With these changes, the enhanced ACT math test should maintain at least the same alignment strength.

3.1.1.3 Science Alignment to Common Standards

Over 90% of states use science standards that are based on the NRC Framework or the related NGSS, which put scientific sense-making using three-dimensional learning and assessment at the center of the standards. Science and Engineering Practices (SEP) are described as what scientists do (e.g., design and conduct investigations, analyze data, and engage in scientific arguments with evidence). The Crosscutting Concepts (CCC) in science are described as how scientists think (e.g., patterns, cause and effect, scale, and quantity). The Disciplinary Core Ideas (DCI) are what scientists are engaged in studying (e.g., big ideas in life science, chemistry, physics, and the Earth and space sciences). The NGSS lay out exemplars of what that three-dimensional learning might look like in a class with one SEP, one CCC, and one DCI combined to form a Performance Expectation (PE). ACT has used conclusions from the NRC report *Developing Assessments for the NGSS* (National Research Council, 2014) to influence our assessment design. This report suggests that the PEs listed in the NGSS represent only one potential way to mix SEP, CCC, and DCI. It also suggests that using the PE as the basis for assessment blueprints would likely produce “teaching to the test,” as many topics could only be assessed in one potential scenario (e.g., momentum (Newton’s third law) could only be assessed through an engineering solution to collisions).

Using the 2014 NRC report as a guide, the ACT science test involves scenarios with multiple tasks (Conclusion 2-1) that require sense-making of the scenario. We use Conclusion 2-4 as justification for not using the PE as one-to-one mapping of the test specifications. We use Conclusions 2-3 and 5-3 to acknowledge that it is impossible to cover all of the NGSS in one assessment, and we will use a content sampling plan over several forms to cover the breadth and depth of the standards. Recommendation 7-2 suggests using a different logical grouping structure of the standards, the ACT Science Reporting Categories cluster the standards into a logical grouping related to data interpretation, scientific investigation, and scientific argument clusters. Since this logical structure has been accepted under peer review, it appears to meet this goal. Finally, the Achievement Level Descriptors (ALD) or Performance Level Descriptors (PLD) provide a way to locate students along a progression of successively more sophisticated application of SEP and CCC (Conclusion 2-2) when used for accountability purposes.

Because there is complicated overlap among the CCCs (refer to Figure 3.1), the SEP (refer to Figure 3.2), and across these two dimensions (refer to Figure 3.3) that would be extremely difficult to disentangle, the legacy ACT clustered these skills into three reporting categories: Interpretation of Data (IOD), Scientific Investigation (SIN), and Evaluating Models, Inference, and Experimental Results (EMI). In the enhanced ACT, the EMI reporting category will be renamed to Evaluating Scientific Arguments and Models with Evidence so that confusion caused by the previous name is reduced and the language of scientific argument common to state standards and postsecondary instruction is used instead. This name change is intended to more clearly link the skills taught in high school and used in college courses to the reporting category.

Figure 3.1. Overlap Among the Crosscutting Concepts in Science Standards

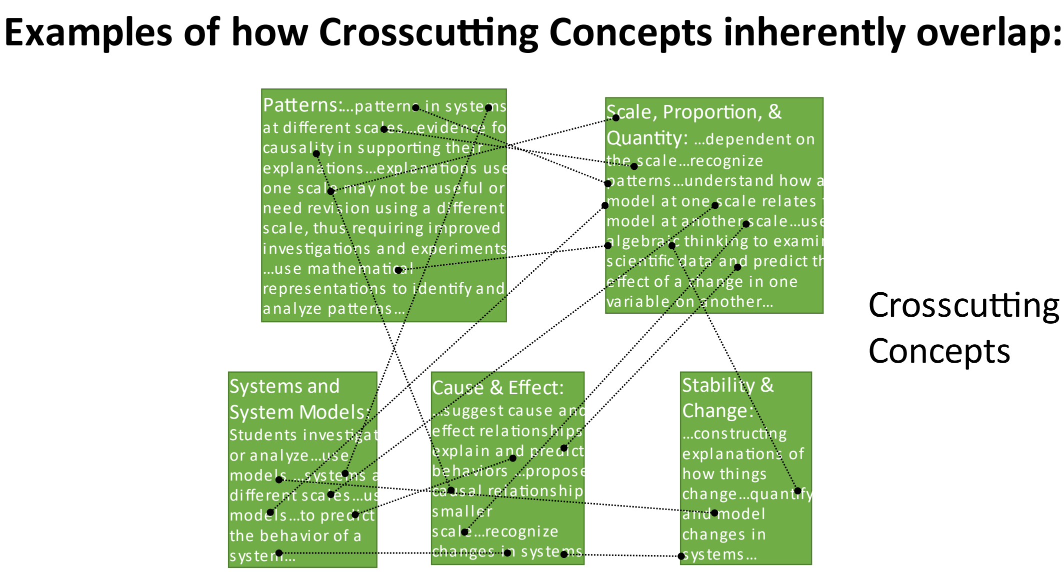


Figure 3.2. Overlap Among Science and Engineering Practices in Science Standards

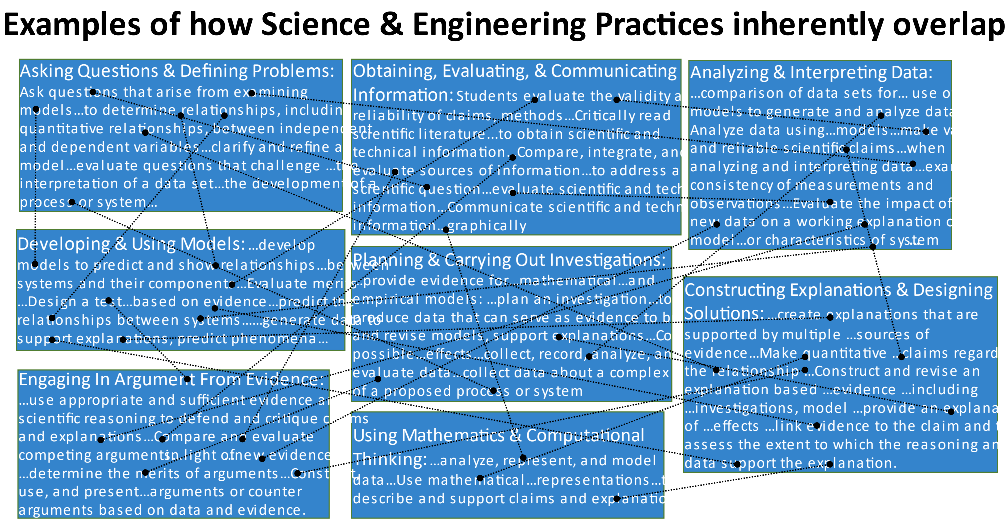
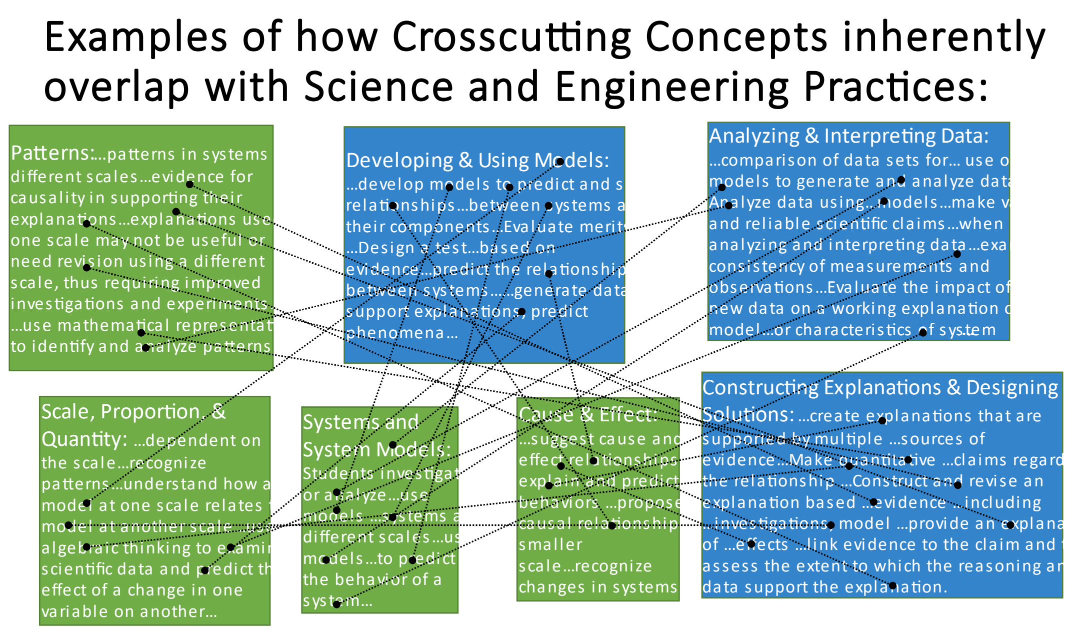
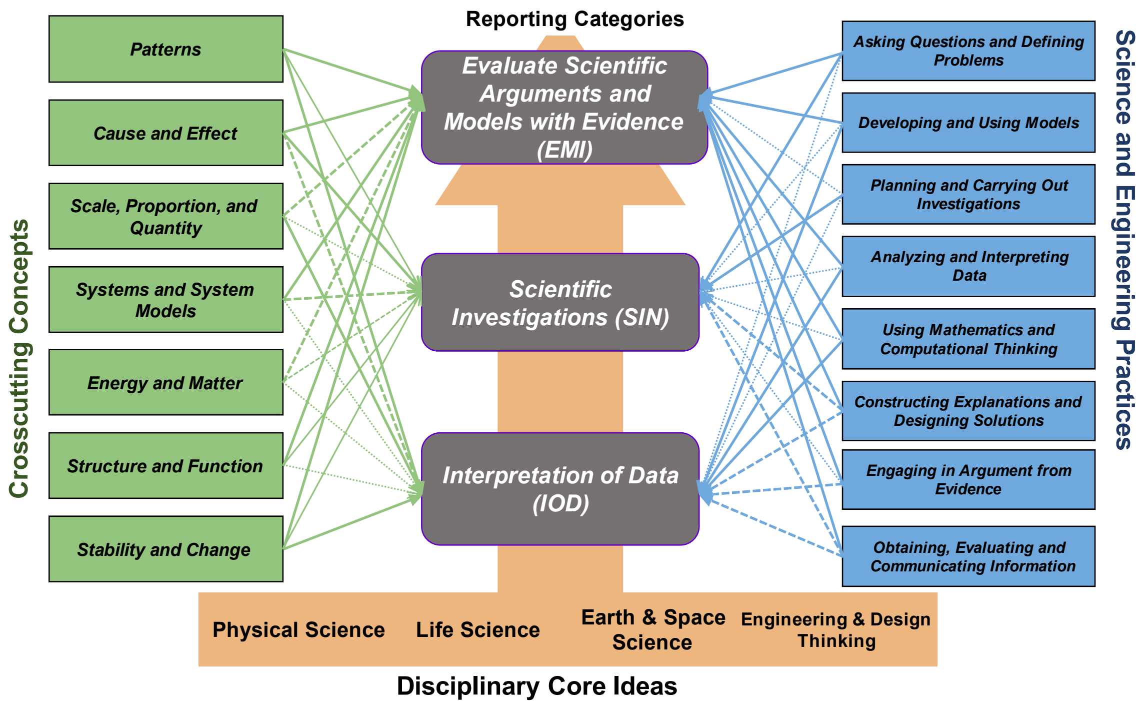


Figure 3.3. Overlap Among SEP and CCC in Science Standards



Because not all SEP or CCC contribute equally to each reporting category, Figure 3.4 shows the strength of alignment for each CCC and SEP to the ACT Science Reporting Categories. The solid lines indicate strong alignment between the skills in the SEP or CCC of science standards based on the NRC Framework or NGSS. The long-dashed lines indicate moderate alignment between the associated knowledge and skills and that reporting category. The fine dots represent some alignment between the appendices and the reporting category. For example, Engaging in Argument from Evidence (SEP 7) should strongly align with ACT's reporting category cluster of Evaluating Scientific Arguments and Models with Evidence (EMI) while having moderate alignment with IOD (the evidence for the argument) and some alignment with SIN (the source of the evidence or a way to test the argument).

Figure 3.4. Strength of Alignment Between Reporting Categories and SEP and CC



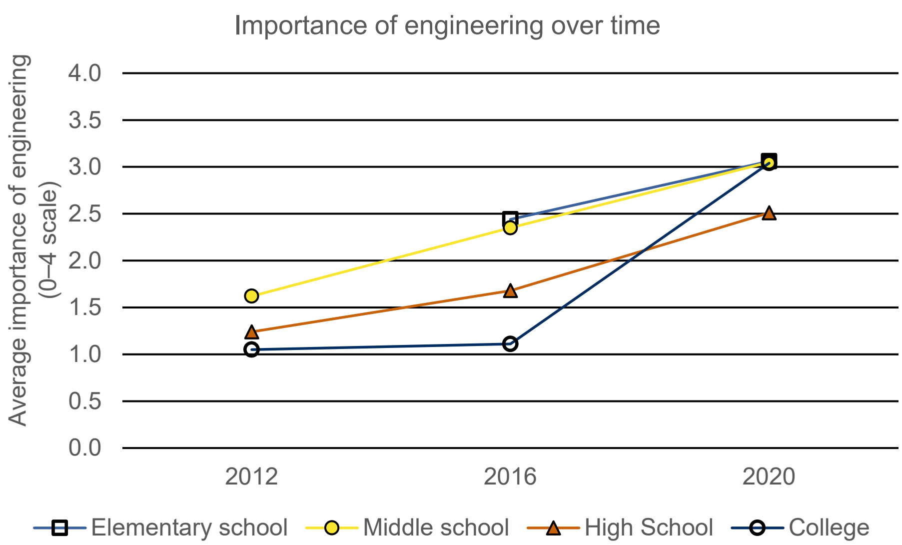
Even when examining the Performance Expectations of common state standards based on the NGSS, the enhanced ACT science test has very strong alignment as all performance expectations can be assessed to some extent. Because of constraints as a large-scale timed multiple-choice assessment, the ACT science test is unable to directly assess skills that involve students creating scientific arguments or engaging in long-term research and experimentation. However, students are required to evaluate those arguments, scientific models, experimental methods, and investigations, as well as the research based on authentic scientific scenarios, data, and graphics. Although selecting which option could best strengthen the claims or experimental procedure is not the same skill as creating the options, it is an appropriate measure of those skills given the constraints of the enhanced ACT.

As part of the review process for the enhanced ACT, the science team reviewed alignments conducted between 2015–2019 related to ESSA peer review and performed a comprehensive review of common state standards. Because alignment studies are subjective and different methodologies can result in different outcomes, the team attempted to address several common concerns regardless of alignment model used.

3.1.1.3.1 Engineering and Design Thinking

Previous alignments found the legacy ACT to have insufficient content on Engineering and Design Thinking. In NGSS and similar standards, engineering and design thinking focuses on applying scientific practices, methods, and knowledge to solving real-world problems. That process includes identifying problems that might be solved using science, identifying solution criteria (goals) and constraints, creating potential solutions, and testing those solutions. On the enhanced ACT science test, these skills will be addressed at both the passage and item level. Moving forward, there will be at least one passage per form that has an engineering and design focus. Many legacy ACT science forms had passages that had an engineering and design secondary focus; now, by design, they will appear on every form.

Figure 3.5. Importance of Engineering Across Grade Bands on the National Curriculum Survey (2012–2020)



The 2020 NCS data showed a continuation of the increase in the perceived importance of teaching Engineering and Design Thinking across all grade levels from lower elementary through postsecondary (approximately 3.05 for all but high school for the 2020 report). However, 60% of responding science educators reported that activities specifically designed to teach these topics were rarely or never taught. With a few more years of teaching based on the NRC Framework and NGSS, more students should have had access to opportunity to learn these skills. This will be addressed in 3.3.4.5 with information from cognitive labs.

Additional sub-reporting categories for Engineering and Design Thinking have been added to both the SIN and EMI reporting categories. Because these are below the reporting category level, there are not specific guidelines in the blueprint for their inclusion on individual forms; however, they will be part of the domain sampling plan to ensure that these skills are tested sufficiently.

3.1.1.3.2 Three-Dimensional Science Items

Based on concerns from this review and the external expert panel, the percentage of items that require the integration of prior content knowledge for three-dimensional scientific sense-making will be increased from 10%–15% of the test to 15%–24%, which should further strengthen alignment to standards.

3.2 Expert Judgements of the Relationship Between Parts of the Test and the Construct

The *Standards of Educational and Psychological Testing* (2014) states that one source of validity evidence is "the judgements of the relationship between the parts of the test and the construct" (p.14). The legacy ACT has used many processes to document this type of validity evidence. The enhanced ACT will continue to use best practices for gathering this evidence, such as the item and form review process listed subsequently, while also engaging in external review of new items, passages, and blueprints. Information from both the legacy and enhanced ACT will be discussed in this section.

3.2.1 National Curriculum Survey

The ACT has traditionally conducted the NCS every 3–5 years to gather information from K–12 teachers, postsecondary instructors, employers, and employees about what KSAs are needed for student success. This feedback is critical to aligning what is taught in the K–12 classroom with what is needed to succeed in college or career. ACT uses longitudinal data as well as snapshot data to inform data-driven decisions about what to continue to test, what should be added to the test, what should be removed from the test, and what should be monitored for later decisions. For example, the 2020 NCS found that both high school and postsecondary instructors thought that science should be assessed using scenarios with authentic scientific data and experiments, as well as with multiple-item arguments that require students to engage in scientific sense-making. Those findings, similar to the longitudinal data, support the decision to maintain only passage-based scenarios with no discrete science content knowledge questions that might be used on an end-of-course exam. Other specific findings have been addressed with the relevant test sections in this chapter as well Chapters One and Two.

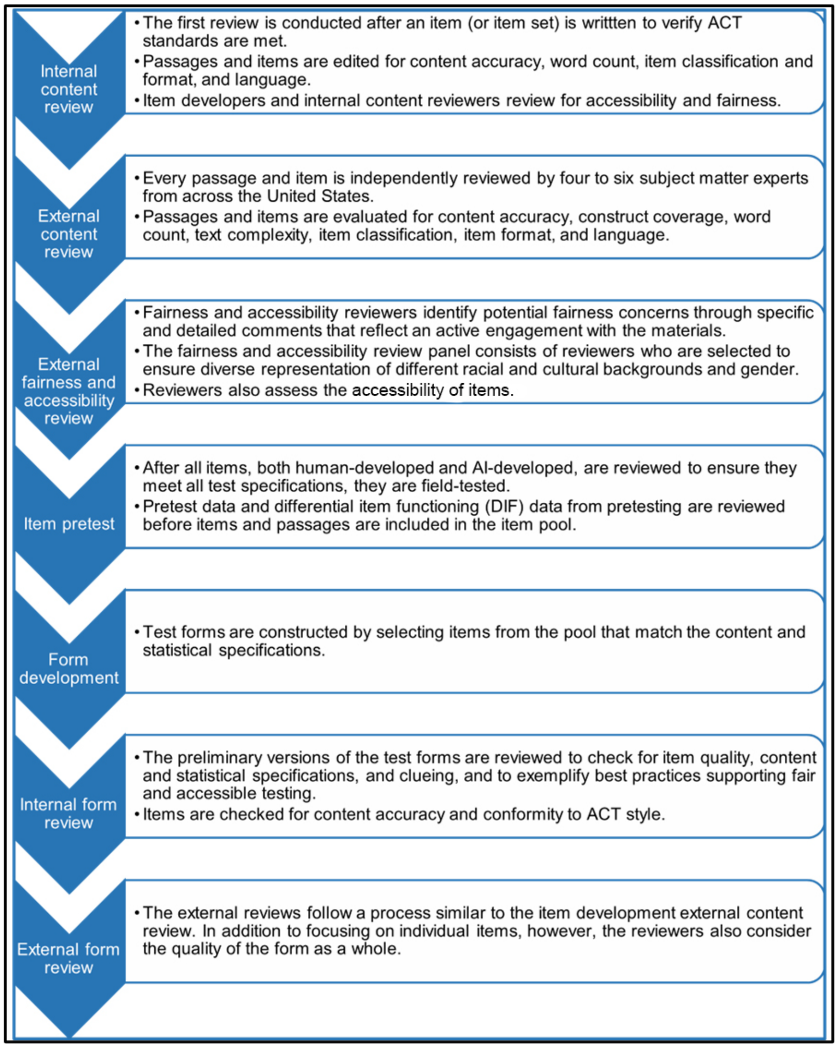
3.2.2 External Expert Panels Related to the Enhanced ACT

ACT conducted external expert panels to evaluate proposed changes to the enhanced ACT. These findings are addressed in detail in Chapter 2.

3.2.3 Item Review Processes from the Legacy ACT That Will Be Continued for the Enhanced ACT

All items undergo multiple rigorous content reviews, including internal peer reviews, internal senior-level reviews, and external reviews. Reviewers verify that the items elicit evidence of the targeted knowledge and skills, are developmentally appropriate, and are error-free. They verify that each item is not likely to produce construct-irrelevant variance. External experts also participate in fairness reviews to help ensure that items and tasks are fair and unbiased. The review process detailed in this section was followed for all items appearing in the legacy ACT forms and will continue to be the process for the enhanced ACT.

Figure 3.6. High-Level Overview of Item and Form Review Process for Enhanced ACT



3.2.3.1 Internal Reviews

After an item (or set of items) is written, it is reviewed multiple times by peer and senior content specialists, who use a standardized rubric to verify that items and passages meet ACT's standards. Test content 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 to minimize the possibility of construct-irrelevant variance.

3.2.3.2 External Reviews

Every passage and 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 either the high school or postsecondary level. During the external content review, passages and items are evaluated for content accuracy, construct coverage, word count, text complexity, item classification, item format, and language. ACT contracts external reviewers with knowledge and experience in relevant content areas to participate in reviews. Reviewers are selected to ensure diverse representation of different racial, cultural, and geographic backgrounds, as well as gender.

3.2.3.3 External Accessibility and Fairness Reviews

Fairness reviews play an essential role in the development of the ACT to minimize construct-irrelevant variance. To help ensure that content is fair, unbiased, and accessible, we conduct external fairness reviews for all items and writing prompts before pretesting and for entire test forms before they become operational. All comments are reviewed by content staff and appropriate changes are made, including the removal of items or passages deemed unacceptable.

3.2.3.4 Documentation of Item Issues During Pretesting and Operational Use

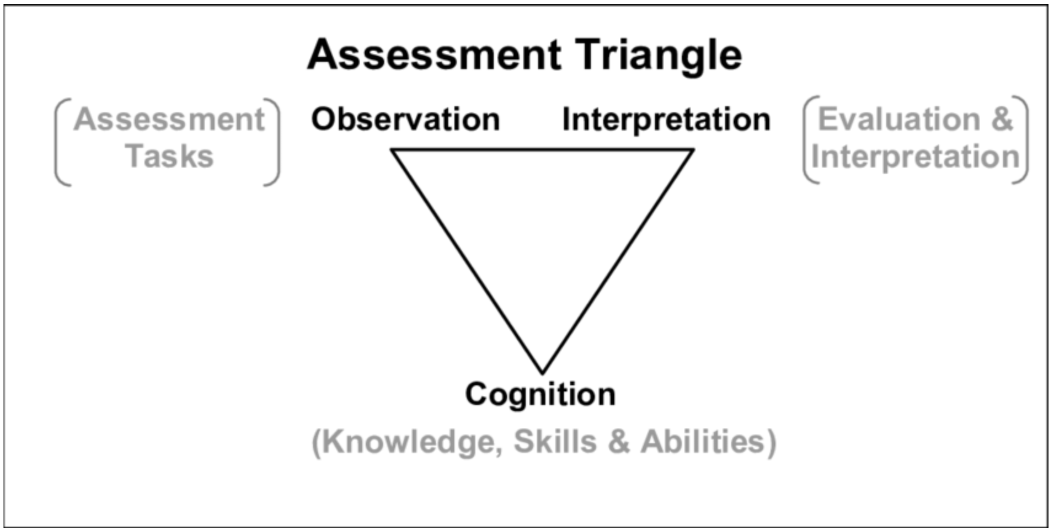
ACT pretests every item before it can appear on an operational form to verify that the item functions properly—that is, the item is not too easy or too difficult, the item contributes to precise measurement of the intended construct, and there are no problems with the answer choices. Subject-matter experts review pretest units to identify possible problems related to statistics or content. All items are checked for Differential Item Functioning (DIF) to ensure that no group is significantly disadvantaged by construct-irrelevance in an item. Items may be eliminated from consideration for operational use at this point based on either content or statistical reasons. In some cases, items are revised and re-pretested.

After each operational administration, item analysis results are reviewed for any anomalies such as substantial changes in item difficulty and discrimination indices between pretesting and operational administrations. DIF is checked for all items to determine that the items performed as pretested.

3.3 Evidence Based on Response Processes

Since 2001, when *Knowing What Students Know* (Glaser et al., 2001) was published with the assessment triangle model (Figure 3.7), it has frequently been posed that the cognition vertex of the triangle is often a black box of unknown activities occurring within the mind of a test taker. This section will summarize several ACT research studies over the last few years that have used cognitive labs (primarily think-alouds and eye-tracking) as well as student interviews and surveys related to these cognitive processes. Each study serves as a source of validity based on response processes.

Figure 3.7. Assessment Triangle Model



Additionally, earlier work on the ACT® WorkKeys® Assessment system provided further evidence of the skills used in literacy and graphic literacy (Langenfeld et al., 2020; Thomas & Langenfeld, 2017a; Kliewer et al., 2018) that overlap with the ACT reading and science tests, providing both tacit and explicit knowledge used to generate the initial artifacts for the enhanced ACT (Thomas & Langenfeld, 2017a; 2017b).

3.3.1 2017 Study Related to Proposed Adaptive Test Model

This study was undertaken as part of previous periodic reviews of the legacy ACT to examine the possibility of using an adaptive test model. The study involved students interacting with discrete English items, novel math items, and shortened passage sets for both reading and science sections. The findings on discrete items and shorter passages are not relevant to the enhanced ACT redesign and will not be included here. This section will focus on specific knowledge gained to inform decisions for the enhanced ACT.

This study included 20 participants from Grades 11 and 12. Fourteen students answered English and reading items and were interviewed about the differences in completing the assessment in different modes (paper, laptop with external monitor, Chromebook). A subset of six students (three male and three female) engaged in think-aloud protocols for English and reading items.

Students were presented with five items in the Production of Writing reporting category: three organization items and two topic development items. These items are designed to simulate editorial decisions that students make while revising essay-length academic texts. The multiple-choice items present students with potential changes such as the decision to reorganize the text, delete portions of the text, or add information to produce a specified rhetorical effect. All items in this category are passage-based and include a stem and four response options; in some cases, response options refer to an underlined portion of text, and one of the response options can specify “NO CHANGE,” indicating that the best option is to leave the text as is.

Student think-aloud evidence supported that each of these items required students to use the targeted knowledge and skills. Responding to the items that assessed organization, students recognized that the items focused on transition words. For all organization items, students carefully studied the passage context and considered the different response options before selecting an answer. They did not seem to be confused by the items, but in each case, they performed close reading of the item text and the passage context. One of the transition items was more difficult (A4); only one of the three students answered correctly. For this item, it is possible that one of the students selected the incorrect response accidentally, because the rationale provided referred to the correct option. The other two students did not seem to be confused by the item. These items elicited more than one cognitive step in a fairly routine way appropriate to the targeted knowledge and skills.

For both topic development items, students carefully evaluated the passage context and drew inferences to arrive at an answer (e.g., “The bike was rented, so he probably was a tourist”). Because these items ask about potential changes to the meaning (“If the writer were to delete . . .”), students demonstrated evidence of speculative thinking. In all cases, the students provided a rationale for their answer. In some cases, the rationale indicated that they were considering a rhetorical perspective, such as which information would be the most useful to a potential reader. The item is designed to test this type of knowledge about text development. The items appeared to elicit multistep cognitive processes in both routine and non-routine ways.

Students also completed one item in the Knowledge of Language reporting category. This item targeted the skill of ensuring logical word choice. All three students who responded to the item took time to consider all response options and deliberated over the best response, returning to the passage multiple times to examine the context. Students acknowledged that multiple options would work, but one was stronger—an important characteristic of item design. The student who answered the item incorrectly provided a weaker rationale for the selection. These findings support claims that these items are able to measure the nuances of the targeted skills, even when using multiple-choice formats. There is little doubt, however, that given the considerable time the three students spent evaluating the item, they were required to carefully evaluate the passage and apply discourse knowledge that they would use in the composition of their own essays.

It is interesting to note that this evidence parallels the English panelists’ observation that including a higher percentage of POW and KLA items might counteract the increased time per item because these items require more time to process than most Conventions of Standard English items. Further results from timing studies could clarify these findings.

Many of the other findings are similar to the 2020 studies’ findings and will be addressed there with the larger, more representative sample.

3.3.2 Cognitive Labs 2019–2020: Overview

These cognitive labs gathered evidence of student cognitive processes of knowledge and skills based on items that are appropriate for the enhanced ACT across the reading, math, and science sections. Some items could have been used in the legacy ACT, but the primary goal was to gather evidence on new aspects of the construct, such as new engineering items in science, or new ways to authentically assess the construct, such as novel math items or grid ins. Results and methods are discussed in detail following this section.

3.3.2.1 Participants and General Methodology

During the first semester of the 2019–2020 school year, another cognitive lab study was conducted that focused on reading, particularly Visual and Quantitative Information (VQI) passages. The study, which also included some math and science items, included 28 students in Grades 11 (71%) or 12 (29%) from two different midwestern states. The sample was 54% male and 46% female. The ethnicity of the participants was as follows: 75% White or Caucasian, 10.7% Hispanic, 7.1% African American, 3.6% Asian or Pacific Islander, 3.6% Native American (self-reported as two races), and 17.9% who self-identified as two or more races. Students who self-identified specific ethnicities as well as two or more races are included in all ethnicities that they identified. Students were paid a $40 electronic Amazon gift card for their participation in the two-hour session.

The protocol for this study included eye-tracking for paper-based forms. We utilized the goggle method approach, which creates a video file that shows the eye focal position overlaid on the visual field by using a camera located just above the bridge of the nose on the eye-tracking goggles. Students were also recorded with two cameras (one on the student’s face and one on their paper forms). After calibrating the goggles, there was no additional change from normal student behavior in answering test questions to accommodate the eye-tracking part of the protocol.

All reading passages were based on legacy ACT specifications for word count and included 10 items across the reporting categories, as would be used on operational legacy ACT forms. Stage One used retired legacy ACT passage sets so that conversion to a scale score could be used. Stage Two used non-operational research passages designed to provide validity evidence for using VQI passages in future legacy ACT forms.

3.3.2.1.1 Stage One

Students were given 17.5 minutes to silently complete two released reading units from the same form, which corresponds to the time for half of the reading section. A timer was placed in front of the students so that they could monitor the time remaining. Additionally, the participants were reminded when they had five minutes remaining, similar to actual test conditions. This dynamic was designed to examine changes in student testing behaviors and cognitive processes when they felt that time remaining was short. The scores for this half of the test were used to estimate the students’ scores on the reading section so that students could be placed into LOW, MID, and HIGH scoring categories for further analysis. Students also completed a short survey at the end of these two units.

3.3.2.1.2 Stage Two

Students were presented parallel versions of three reading passages in which one version was a VQI passage, with one or more graphics (table, process diagram, or bar graph). Each student answered questions on three passages in total after being randomly assigned to one of the three options categorized in Table 3.1.

Table 3.1. Stage Two Reading Passage Group Assignments

Random group assignment	Passage 1 (silent)	Passage 2 (think-aloud)	Passage 3 (silent)
A	Les Paul (VQI)	Aquatic Universe (VQI)	Strad (non-VQI)
B	Strad (VQI)	Les Paul (VQI)	Aquatic Universe (non-VQI)
C	Aquatic Universe (VQI)	Strad (VQI)	Les Paul (non-VQI)

This counterbalanced approach was designed to examine differences in cognitive demands and cognitive processes for incorporating the VQI graphics with a similar traditional passage. The information in the graphics was summarized in the text for the non-VQI version of the passage. The think-aloud was performed second so that the first VQI passage behaviors and cognitive processes would not potentially be affected by engaging in the think-aloud on the first passage. These passages were not timed. Students were instructed to answer these questions as if they were actually receiving a score so that processes like reading speed could be compared between these passages and the timed passages. After completing items at this stage, students were given a short survey and a guided interview about the reading section passages and items which are discussed subsequently.

3.3.2.1.3 Stage Three: Math and Science

Students were given ten math items to complete in ten minutes across multiple reporting categories. They were instructed to answer the questions as they would on test day. On the facing page of each math item, students were given a Likert scale to rank the difficulty (easy, medium, challenging, very challenging) and confidence in correct answer (very confident, pretty sure, not sure, confident that I was wrong). Time calls were given at five minutes and one minute remaining, and students could monitor the same countdown timer that was used in Stage One. Some of the math items were created as fill-in-the-blank with numerical answers, like a grid for paper testing or numeric entry for online testing. A TI-84 calculator was available if students did not bring their own calculator.

Students were given twenty minutes to complete questions related to three science passages. Each passage was presented in either the long form (with 6–7 items, using legacy ACT specifications for word counts, number of graphics, and graphic types) or a short form (with 3–4 items, shorter text, and only one graphic needed to answer the specific items selected). The tropical storm passage only added a second graphic as part of the stem of the item that required the use of both graphics. Similar to Stage Two, the reading passages were assigned using random group assignment and counterbalanced design (refer to Table 3.2).

Table 3.2. Stage Three Reading Passage Group Assignments

Random group assignment	Passage 1 (long)	Passage 2 (short)	Passage 3 (short)
A	Orange juice	Tropical storm	Carbon phase
B	Tropical storm	Carbon phase	Orange juice
C	Carbon phase	Orange juice	Tropical storm

After Stage Three, students completed a survey focusing on math and science items and participated in a guided interview that focused on how graphics were used in each section of the test. Students were also asked about the similarities and differences in the use of graphics among the three subjects. Discussion of the interviews and surveys appear subsequently.

3.3.3 Cognitive Labs 2019–2020: Reading Findings

Videos were created that overlaid the eye-tracking focal point on the student's view of the test materials (passage and questions on facing pages) and then synced with videos of the student's face that included audio from both the proctor and the student. These videos were then coded for a variety of information by two ACT staff members who were given training and completed a test sample video to verify that the coding process was consistent with the official coding of the sample item. If there was more than 10% disagreement in a time stamp or other code, then a senior researcher (Principal Content Specialist) would rewatch and recode the video to give the final values used for analyses.

Overall, students exhibited many of the same cognitive processes as judged by their eye-tracking, which will be discussed subsequently in further detail. The first round of analysis was completed in 2020 regarding validating the inclusion of VQI passages and items to better align with state standards as part of an ESSA review. The reading rate analysis was completed using the videos in 2022.

One student noted that she normally gets time and a half for these assessments. We had previously avoided asking about accommodations for research studies but determined that it would be important to specifically track the accommodations of participating students in the future given the anonymization of the data past the initial recruitment stage. This can be seen in the review of the 2023–2024 study.

3.3.3.1 Initial VQI Analysis 2020

Students generally found the inclusion of graphics to be surprising for the ACT reading section. Comments like "I haven't seen graphics in passages when I took the ACT before" or "I didn't see any of these when I practiced for the ACT" were common. Most students thought that the graphics were similar to graphics they had seen in class. The Les Paul diagram took more time for students to process than the more familiar bar graph and table in the other passages. This was consistent with the WorkKeys 2.0 cog labs and eye-tracking that found familiarity with the type of graphic was related to the overall complexity and amount of time needed to process the passage and set of items (Thomas & Langenfeld, 2017a; 2017b). Overall, students stated that the graphics were manageable and helped them to understand the passages. Some students admitted to not even noticing the graphics when they first read the passage, which was confirmed with eye-tracking analysis, and only noticed them when a question asked about information that could not be found in the passage text. This occurred in several cases, even when the question was scaffolded with language like “based on the figures” or “based on the passage and the figures.” Again, failure to apply scaffolding language is a common maladaptive behavior also exhibited in the WorkKeys research (Thomas & Langenfeld, 2018), the 2017 science cog lab items, and the 2024 science cog labs, described subsequently.

Students generally gauged the difficulty of the passage based on the familiarity with language being used. The use of jargon (scientific, musical, etc.) unfamiliar to the student had a greater impact on the perception of difficulty than other factors. Students also stated that the passages they found interesting were easier to comprehend. The interplay of familiarity, interest, and perception of difficulty may provide further validation of the Windows and Mirrors initiatives to refine ACT’s best practices in incorporating texts that recognize the complex mix of experiences students navigate and to create culturally affirming reading experiences for students.

Generally, students used the targeted skills for the VQI passages. There was strong think-aloud and eye-tracking evidence that students who answered correctly used the appropriate parts of the passage and graphics. Comments about specific graphics were used to make modifications to VQI passages and graphics for the legacy ACT.

3.3.3.2 Reading Rate Analysis of the Five Reading Passages

The use of eye-tracking equipment in both timed and untimed sections of the reading test posed a unique opportunity to determine how students read passages when they believe they are under time constraints and when they do not believe they are under time constraints. ACT could examine differences in behavior, if any, when the timer was low (less than five minutes or less than one minute) and when the student was untimed or believed they had a relatively substantial amount of time remaining. This would allow for the comparison of more current reading rates of students with the reading rates used to determine the timing for the legacy ACT which were based on data from the 1970s and earlier.

The combined videos of the student’s face and the eye-tracking path were coded by two trained coders. The initial time stamp for the beginning and end of each paragraph, column, overall passage, and item stem were recorded to the millisecond. Additionally, the time stamp for the beginning and end of each solution behavior was recorded to the same precision. This would allow for several analyses. Each student’s reading rate could be calculated for any of these possible combinations. Changes from passage to passage could be observed. Differences between the first and second passage would give support to how the perception of speededness or rushing leads to changes in cognitive processes. We also examined how much time students need to make sense of the graphics.

As is customary in eye-tracking research with small samples, the median value will be used for the discussion as an outlier can have a large impact on the mean. The results in Table 3.3 indicate that participants read the first timed passage at a mean rate of 355.82 standard words per minute, followed by the second passage at 1372.40 standard words per minute, indicating that participants read Passage 2 much faster than they read Passage 1. The last three passages were untimed, but the mean reading rate for these passages was 451.60, 335.10, and 239.40 standard words per minute (WPM), respectively. The median for Passage 1 was 241.94 WPM and Passage 2 was considerably higher at 370.50 WPM. The large differences between mean and median support the use of median for conclusions related to the enhanced ACT. Combined with the two students who did not read Passage 2 at all, there is evidence here to support that students engage in different cognitive processes when they perceive they are rushed and use techniques to overcome the speededness aspect of the reading assessment.

Table 3.3. Reading Rate in Standard Words Per Minute by Passage

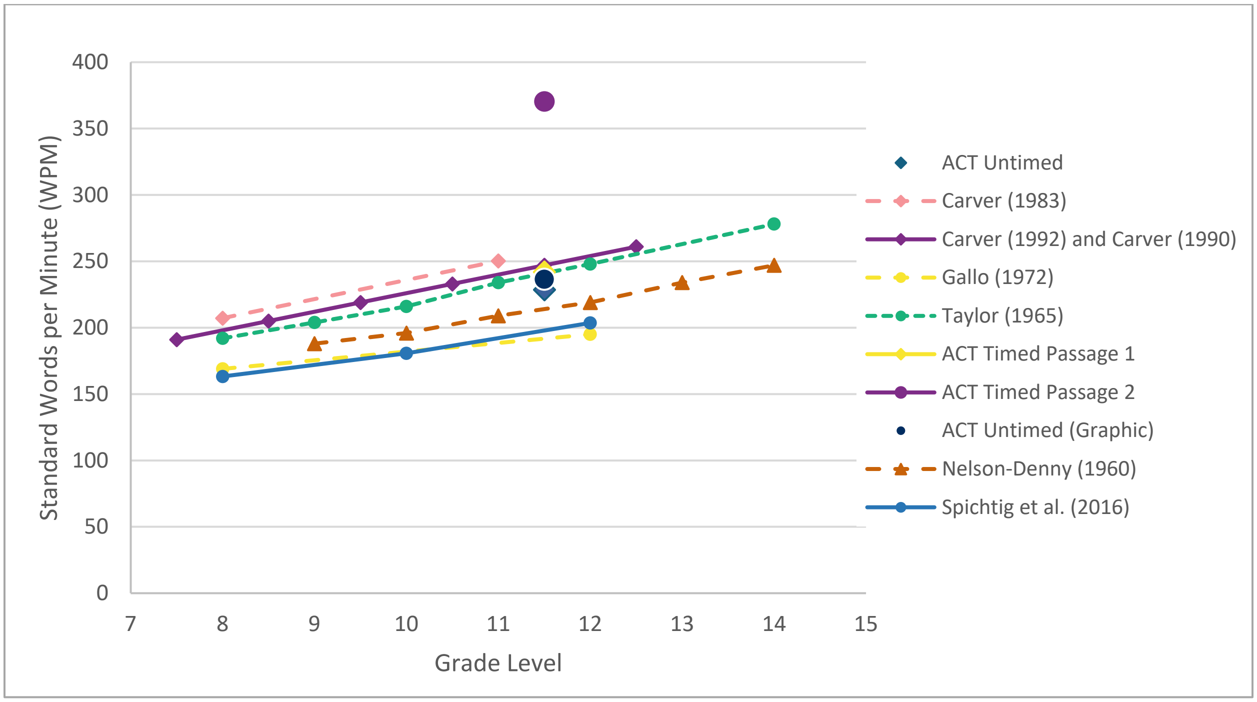
Passage	Timed	Standard words	Mean WPM	Median WPM
Passage 1: Pride	Y	741.70	355.82	241.94
Passage 2: Our Place in the Universe	Y	755.00	1,372.40	370.50
Les Paul	N	663.00	451.60	236.30
Our Aquatic Universe	N	767.00	335.10	272.80
Strad Religion	N	682.00	239.40	176.90

***Note*:** Two students did not read Passage 2.

In interviews and surveys, several participants said that they read Passage 1 and completed the associated items at their normal speed but then realized that they spent too much time answering questions and sped up when reading Passage 2 and its items. For the remaining three passages, their reading rates dropped to a speed similar to Passage 1, even though the passages were untimed, evidence that the reading rates for Passage 1 and the three untimed passages were representative of the participants’ typical reading speed. The overall average of the median passage reading rate was 254.29 WPM. Figure 3.8 shows the median rates for the timed (separately) and untimed passages (combined by type), along with the data of previously cited research.

Figure 3.8 illustrates the average reading rates for Passage 1, Passage 2, and the average reading rate for the untimed VQI and standard reading passages, along with estimates from Carver (1983, 1990, 1992), Gallo (1972), Nelson et al. (1960), and Spichtig et al. (2016).

Figure 3.8. Reading Rate in Standard Words Per Minute (WPM) for Various Studies



It should be noted that Taylor (1965) and Spichtig et al. (2016) excluded participants that did not answer 70% of reading comprehension items correctly. ACT also did not truncate data based on reading comprehension, similar to NAEP (Gallo, 1972).

Figure 3.8 indicates that the average reading rates for ACT timed Passage 1, the untimed VQI passages, and the untimed reading passages are comparable to previous studies, excluding Spichtig, falling near the middle of the range for students in the same grade. This figure also shows how extreme the average reading rate was for Passage 2. Again, participants reported the need to rush when reading Passage 2 based on the time used to answer items associated with Passage 1.

To evaluate how much reading time is needed to make sense of graphics, the reading rates for the versions of a passage with and without graphics were compared. Table 3.4 shows that the reading rates for the passage format with graphics were similar to the parallel versions without graphics, with the mean difference approximately zero.

Table 3.4. Median Reading Time in Minutes for Passages With and Without Graphics

Passage	Standard Passage			Passage with Graphics				
	N standard words	Median WPM	Median time	*N* standard words	Median WPM	Median time	Diff	% Diff
Les Paul	663	236.26	2.81	703.2	259.82	2.71	−0.10	−3.6%
Aquatic	767	272.82	2.81	768.7	199.97	3.84	1.03	36.7%
Strad Religion	682	176.86	3.86	706.2	249.51	2.83	−1.03	−26.6%
Mean	—	228.65	3.16	—	236.43	3.13	−0.03	2.2%

Differences in the reading rates for different types of graphics may involve a combination of comfort with the topic and type of graphic since the passages had similar text complexity ratings. For example, the Les Paul was a simple process diagram, but the Aquatic passage had a dense table with unfamiliar scientific terms as headers. Students stated that questions referring to graphics were easier than others. Similar comments were made about questions that referred to specific lines or paragraphs. Thus, it appears that passages with and without graphics had approximately the same equivalent number of standard words, although there is a lot of variability based on the type of graphic and passage topic.

3.3.3.2.1 Summary of Reading Rate Study Impact

Students were reading at approximately the same rates (225–250 WPM) as were used for determining the timing of the legacy ACT until they engaged in maladaptive speededness behaviors, such as skimming Passage 2 in less than 5 seconds or not reading the passage at all. The timing analysis indicated that the students spent much more time per item than the legacy ACT allows, both in timed and untimed conditions when they did not engage in speededness behaviors. This supported the decision that the timing for passage-based sections on the enhanced ACT would be based on time per item rather than time per passage. Given the reading rate and passage lengths of approximately 750 words for standard passages and 650 words for the new shorter passages, a typical student would be expected to spend between 180 seconds (3 minutes) and 200 seconds (3.33 minutes) reading a long passage, leaving between 325 seconds (5.42 minutes) and 345 seconds (5.75 minutes) to answer the 10 items associated with the passage. Students often took longer than this amount of time to answer questions, so adding an additional 8.3 seconds of solution time per item (plus the additional time saved on the shorter passage) should address the perceived speededness issues on the legacy ACT reading section.

It is important to note that students often engage in poor metacognition when answering test questions, spending large amounts of time on an item that they do not know how to answer (e.g., cannot decode the question, lack the targeted KSA, misunderstood that part of the passage), which greatly contributes to the appearance of speededness. If students do not know what they know and can do, adding additional time may not solve the appearance of speededness issues as students may continue to spend large amounts of time early in the test on items they are not likely to answer without guessing.

3.3.3.3 Cognitive Labs 2019–2020: STEM Findings

Scores from the timed reading test were used to identify students in one of the three groups: HIGH (upper quartile); MED (inner quartile); LOW (lower 25th percentile). This was used as a proxy for math and science as there are correlations between the sections scores of 0.68 for math and 0.80 for science (refer to the ACT technical manual). Differences among the groups as well as overall patterns were analyzed using goggle-based data collected using the Tobii-Pro Glasses 2 hardware and Tobii Pro Software for data collection.

3.3.3.3.1 Math Findings

Gaze patterns for the HIGH performing group showed greater consistency than either the MID or LOW group. This corresponds to previous findings for the WorkKeys Graphic Literacy Assessment (Langenfeld et al., 2020) as well as with findings of other researchers. Gotwals and Songer (2009) referred to the group at intermediate steps in a learning progression as the "messy middle," as these students use inconsistent reasoning and understanding, making it difficult to classify their errors. They will use practices and content knowledge in some contexts, but not in others. The more novel the problem, the more likely they are to revert to a lower level of thinking in the progression. This inconsistency mirrors earlier findings in the SOLO taxonomy of Biggs and Collis (1982). Gotwals and Songer also found that scaffolds in items help students to perform closer to the upper limit of their apparent capacity, and that tasks with fewer scaffolds lead to more errors. Unconventional items and those with fewer scaffolds will be addressed in these findings as well.

The HIGH performing group used time more efficiently and followed expected patterns. The amount of time per question was lowest for the easy items, increased slightly for the items of moderate difficulty, and was much higher for difficult items, as anticipated by the construct.

Unusual problems took more time than traditional problems. Unusual problems involved using math in ways that are not often taught or tested as presented in this study. The only exception to this was the first item on finding the median of a list. The HIGH group took more time than the other two groups. However, this was indicative of attention to detail and checking work rather than indicating a sub-optimal solution path.

Figure 3.9. Average Time Working on Question by Group

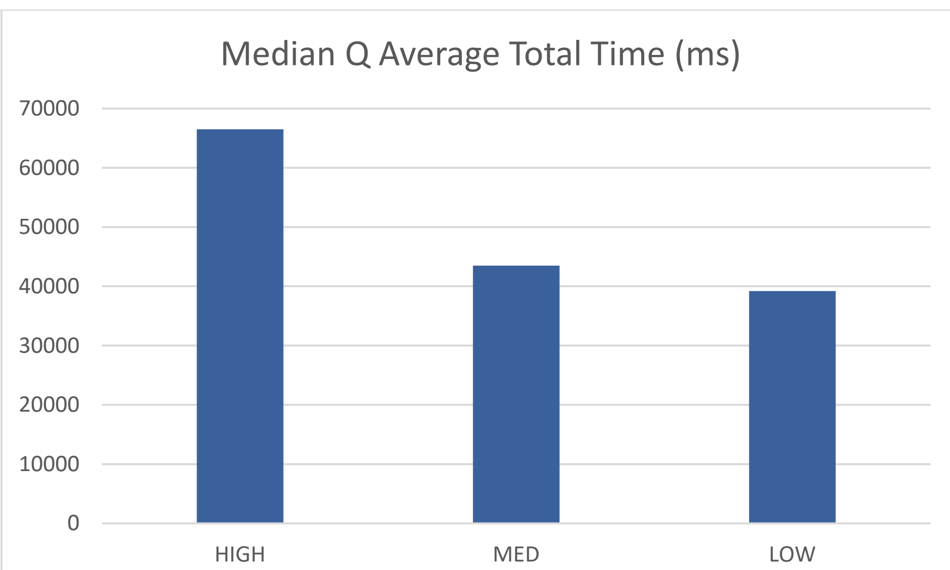
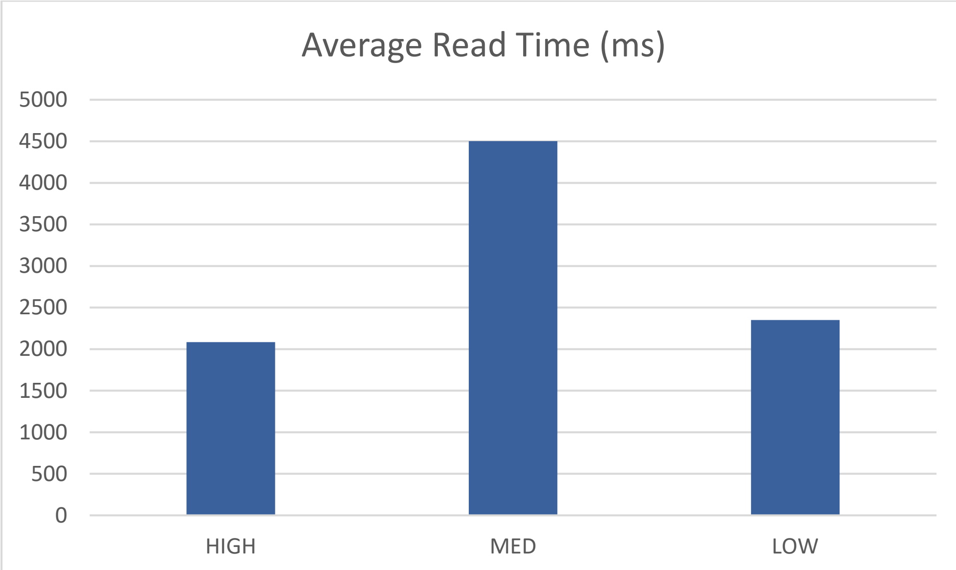


Figure 3.10. Average Time Reading Question Stem by Group



Figures 3.9 and 3.10 show that HIGH performers required less time to decode the question and spent more time on solution behaviors, such as reordering the numbers in the set to find the median. The MED group spent twice as much time reading and re-reading the stem to the question. This may indicate attempts to identify the correct measure of central tendency. In the follow-up interviews, many students identified this median question as the easiest for them, even when they answered it incorrectly. The students were confident that they had found the correct central tendency because the mean, median, non-reordered median (maladaptive behavior), and mode were four of the five answer choices. This provides evidence that plausible foils can lead to a false sense of confidence in students. Since no student selected the mode or the fifth option, which did not use a common error or misconception related to central tendency, having only four options rather than five would not have changed student solution behaviors.

The HIGH group spent more time on solution behaviors than on reading the question stem or scanning the question and figures for relevant information. They were able to efficiently use their time to work on solving problems rather than trying to decode questions that did not follow a traditional path.

Eye-tracking revealed an additional step of having to reorder the numbers in the set. Although many would consider reordering a routine step in these problems, the gaze patterns showed that many of the incorrect students in the MED group used an elimination strategy (finger covering or crossing out) to find the middle term of the original list of 11 items. 75% of the HIGH, 58% of the MED, and 25% of the LOW groups answered this item correctly.

Unlike the other groups, particularly the LOW group, students in the HIGH group were more willing to skip a question and move on. Most did not spend a significantly larger amount of time on any one question. This metacognitive awareness was echoed in their answers to the survey question about how to decide when to skip a math problem (e.g., “when I start to get confused”;

“when I have no clue what to do”; “when I spend more than a minute on a question”; “skip a question if you can’t answer it and don’t stress too much”).

Only five of the participants chose to skip one or more problems and all were from the MED or HIGH group. They spent an average of 66.5 seconds on the item before they skipped the item. Of the 21 participants who answered that they would skip a question, if necessary, 14 of them did not skip any items. All 14 of these individuals spent at least 120 seconds on the Venn diagram or unusual series item, despite their survey responses. This maladaptive use of time was also seen on a graphing item for a mode study in 2018 when all five students in the study spent more than four minutes on a graphing item before guessing their answer without using their graphing calculators.

In reality, the MED and LOW scorers gave similar responses about when to guess; however, those two groups did not apply their reasoning to actual items. None of the four LOW-scoring individuals skipped or guessed and moved on. Consequently, none of the LOW scorers finished all the items within the time limit. This metacognitive awareness of knowing what the student can answer within the timeframe and when to move on is a valuable skill for students to use if they want to show what they know and can do on a standardized assessment. This lack of awareness of “knowing what you don’t know” contributes to poor performance in the LOW group, as these students may never attempt to complete tasks they do know how to do because they have wasted time on tasks that they do not know how to do, spending significant time re-reading question stems and scanning diagrams for a place to start the problem. Even among the HIGH students, there was difficulty in applying the rules that they claimed to use when deciding to guess and move on to another question. This maladaptive behavior was consistent with behaviors exhibited in Part One of the study, which caused students to rush to finish Passage 2.

Extended time spent on one of the three following maladaptive behaviors indicated that a student did not know how to approach a problem:

- repeatedly reread the question stem
- continuously scanned the page
- long fixations on both relevant and irrelevant portions of the page

These behaviors were much more prevalent in LOW and MED scorers than they were for HIGH scorers. Particularly, the LOW and MED groups engaged in these three maladaptive behaviors with greater frequency, for longer durations, and with greater repetition of these non-solution behaviors. These behavioral differences, as observed by gaze-pattern data, were less obvious on routine tasks, such as the median question. Questions that required complex problem-solving steps showed significant differences between the LOW/MED group and the HIGH group. This confirms data from the last two NCSs that “weak problem-solving skills” was ranked as one of the top impediments to success by educators at both the high school and college level.

Figure 3.11. Average Total Time on Item by Group for Math Items

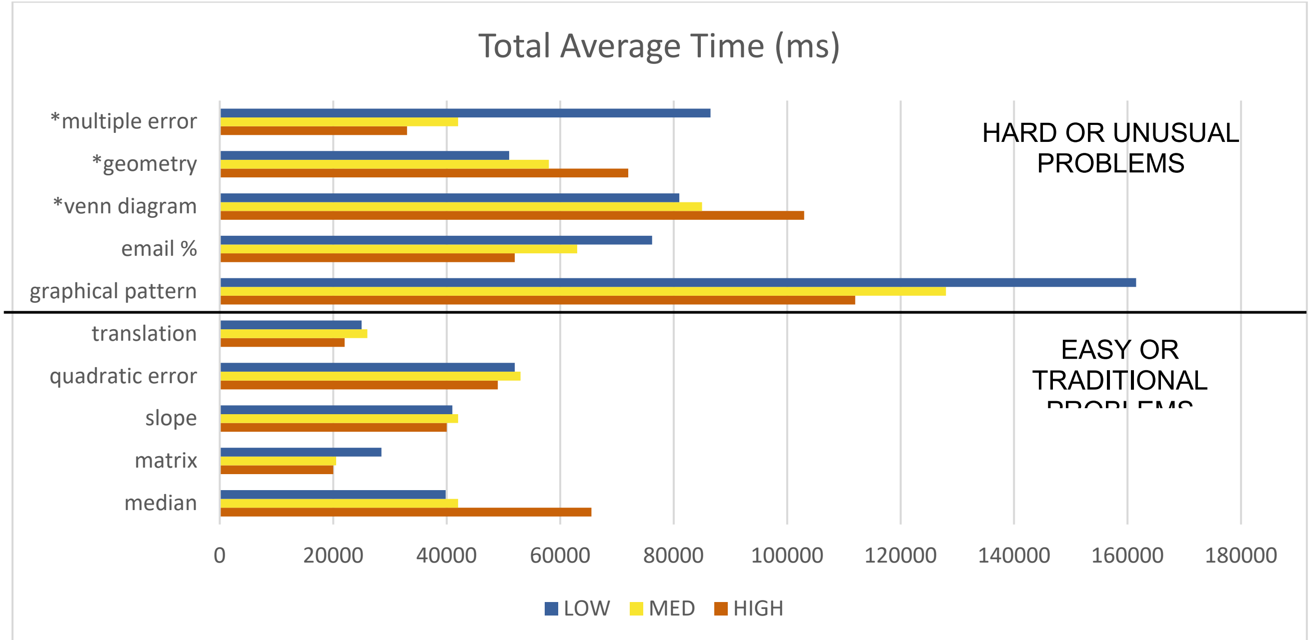


Figure 3.12. Average Question Stem Reading Time by Group for Math Items

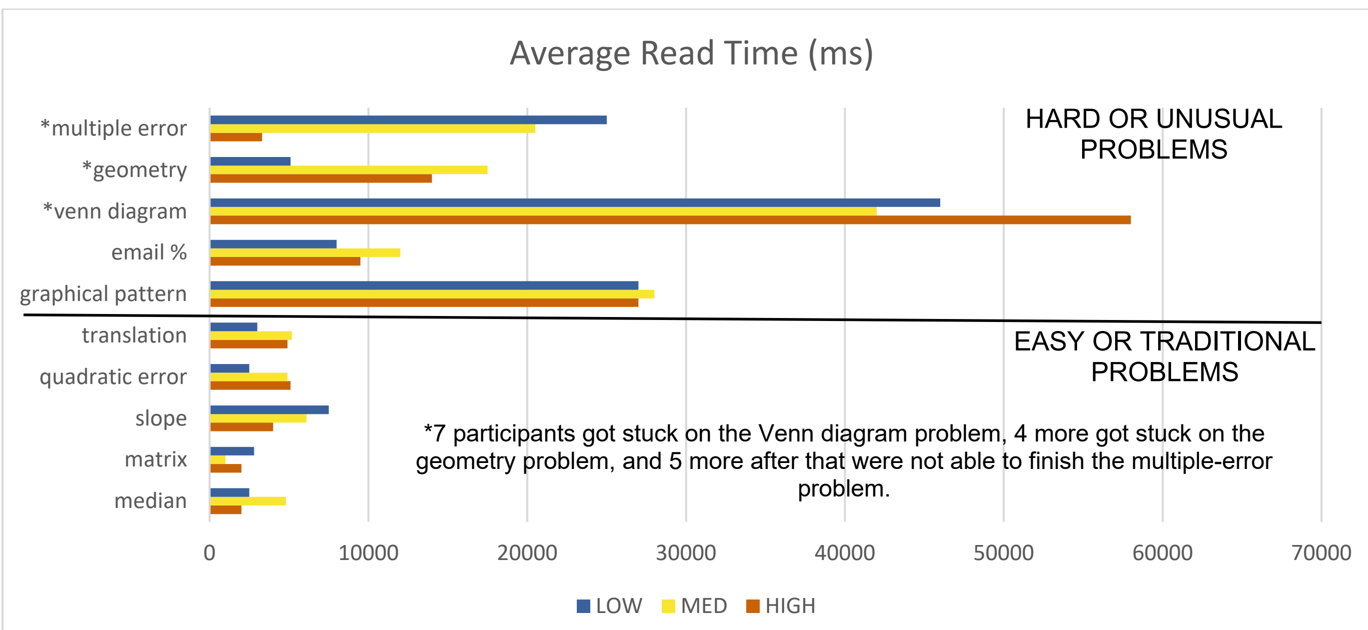
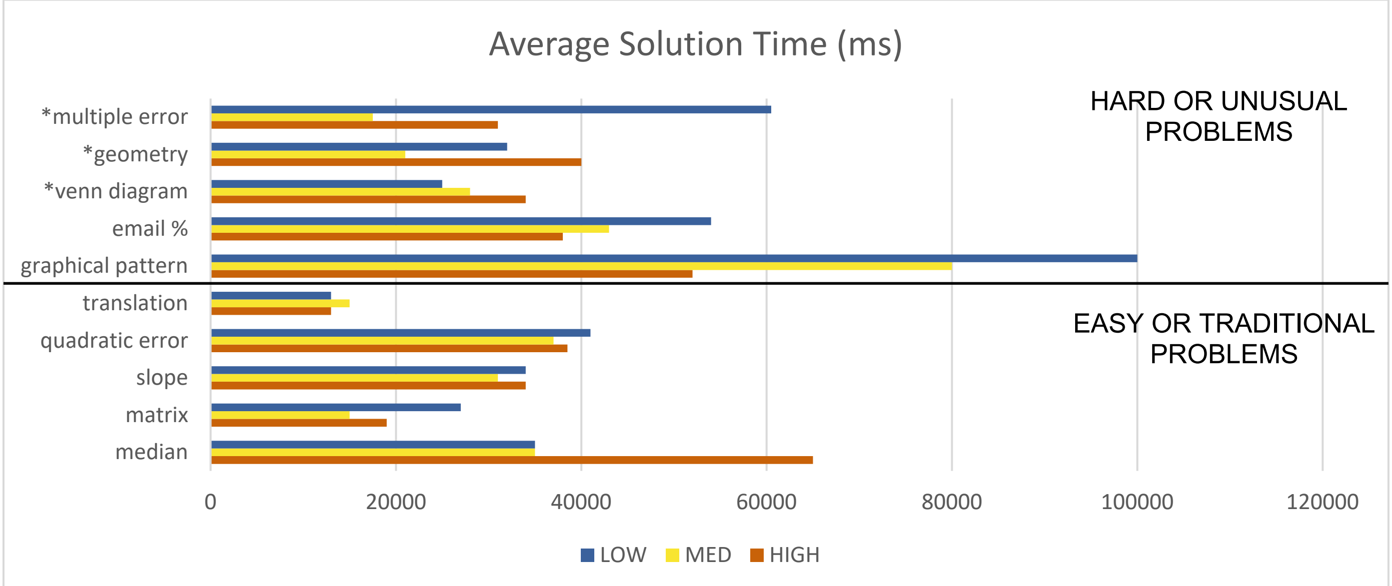


Figure 3.13. Average Solution Time Behaviors by Group for Math Items

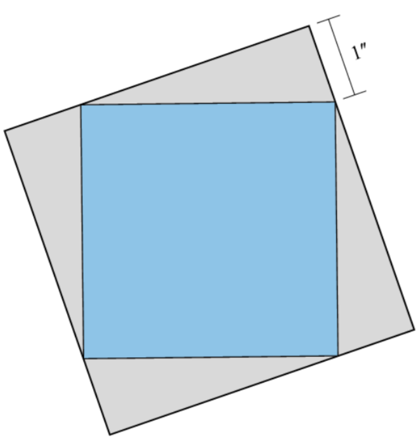


Unfortunately, so many participants used exorbitant amounts of time on the unusual graphical pattern problem and the Venn diagram problem that the data for the multiple-error (logical fallacy) problem is not useful. The data for the LOW group shows that the average time spent on the item was over 160 seconds. With only 10 minutes to answer the 10 questions, it is not surprising that these individuals ran out of time before completing the final three problems.

One other finding of note was that students who were on an accelerated math track (including calculus in senior year or before) exhibited poor performance and maladaptive behaviors when presented with unusual applications of common math content such as the Pythagorean theorem (Figure 3.14) or Venn diagrams.

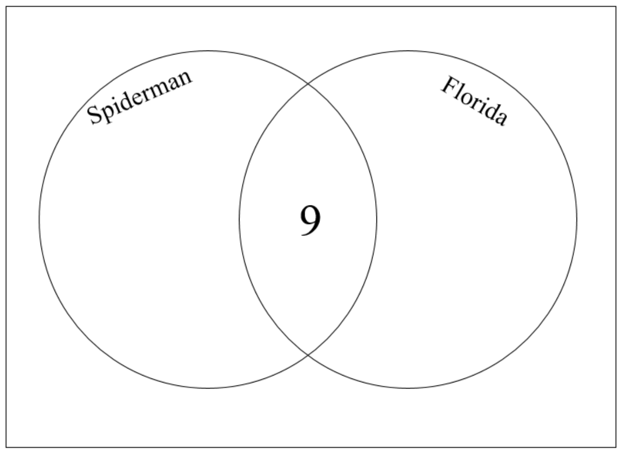
Figure 3.14. Pythagorean Theorem Problem

The smaller square below has an area of 10 square inches. The sides of the larger square go through the vertices of the smaller square. What is the area of the larger square, in square inches?



Only a few HIGH scorers approached this problem with a systematic approach labeling sides and using area and Pythagorean formulas. Most students' gaze patterns involved long time periods rereading the question stem or scanning the figure. Students did not appear to follow line segments to break the diagram into easy-to-use shapes—triangles and squares. The overwhelming majority of students spent a large amount of time with little meaningful solution behavior and then guessed a value.

Figure 3.15. Venn Diagram Problem Stimulus, Question Stem, and Answer Options



Each day in class, a different student gets to ask the class a question. One day, the students learned that 14 of their classmates have seen all the Spiderman movies, while 15 have not. The next day, the students found out that 17 of their classmates have been to Florida, while 12 have not. Today, you asked your classmates if they have been to Florida and have watched all the Spiderman movies. 9 of your classmates answered "Yes," while 20 answered "No." The following diagram captures some of this information.

How many of your classmates have NOT been to Florida and have NOT seen all the Spiderman movies?

(Note: All students were present in class each day.)

A. 0

B. 3

C. 5

D. 6

E. 7

If this problem was surprisingly difficult given that the use of Venn diagrams is common beginning in elementary school across all content areas, and the item was scaffolded with a graphic organizer. Additionally, there are two common solution paths to consider.

In the solution based on critical thinking and pattern analysis, the problem should have been solved by using the statement that 14 students had seen all the Spiderman movies to realize that the part of the Spiderman circle in the diagram that did not include Florida should include the 5 remaining students. The statement about Florida (17 had been there) should allow students to label the Florida circle, not including Spiderman, as 8. The labeled parts of the circle would then account for 22 of the 29 students, leaving 7 students in the outside region with “No” for both Spiderman and Florida. No student approached the problem as suggested above.

The second option to solve the problem is to use a traditional math formula:

A + B + Neither – Both = Total

A few of the HIGH performing students used this formula.

On average, students who completed the item spent 42–47 seconds rereading the item, depending on group classification, which was evidence of maladaptive behavior. This is further evidence of not applying metacognition to decide when to guess and move on to another problem. Many students labeled 14 and 17 on the Spiderman and Florida circles respectively, ignoring the overlap section in the middle that was correctly labeled in the question stem. Again, this problem involved some critical thinking as to what many would consider a “routine” Venn diagram problem with a provided diagram. This item was labeled as DOK3, and the students’ difficulty as well as the multistep nature of the problem support that claim.

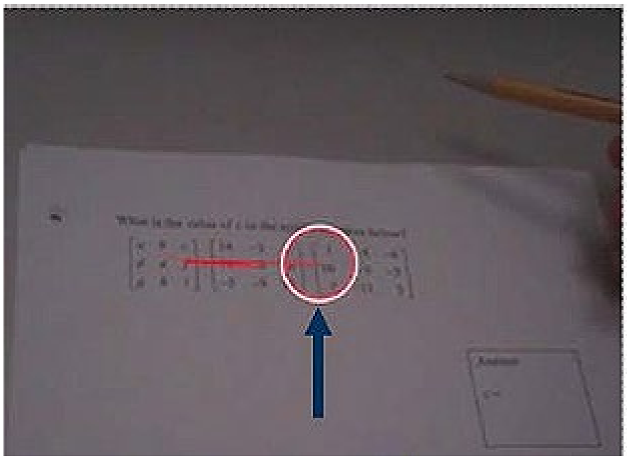
For 7 of the 16 MED/LOW performing students, this problem was the last one that they attempted as they ran out of time.

Figure 3.16. The Matrix Problem Stimulus

What is the value of c in the matrix equation below?

$$\begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix} = \begin{bmatrix} 14 & -3 & 2 \\ 11 & 0 & 8 \\ -5 & -9 & 15 \end{bmatrix} - \begin{bmatrix} 7 & 4 & -4 \\ 16 & -5 & -3 \\ 5 & -12 & 5 \end{bmatrix}$$

Figure 3.17. Eye-Tracking Data for the Matrix Problem Showing a Saccade Line From the Negative to a Fixation Circle on the Equals Sign



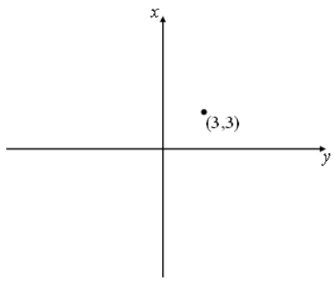
Note: Circle indicates fixation. Red line is a saccade.

This problem illustrates the value of eye-tracking. Students who had a fixation or saccade on the negative sign between the two matrices were much more likely to answer the item correctly. This was especially true of students in the LOW and MED category. All the LOW/MED participants who answered correctly (69%) had either a fixation (73%) or a saccade (27%) over the negative sign (refer to Figure 3.17). Again, the gaze data (Figure 3.17) supports claims that students who answer the item correctly understand the basic operations of addition and subtraction of matrix components. The circle indicates an extended time on that location (fixation, more than 300 milliseconds), while the red line represents a rapid eye movement from one area of focus to another (saccade).

As shown in Figure 3.17, a saccade indicates a gaze jump from the negative sign to the equal sign. Students in the HIGH performing group may have used peripheral vision to ascertain the mathematical operation (subtraction) between the two matrices. This chunking of information would be typical of people with high skill in a task like this. This is an example that shows that HIGH scorers, i.e., experts, have automatized or made routine this procedure (Chi et al., 1981), while students in the LOW and MED group must still work through this in a step-by-step manner.

Figure 3.18. The Translation Problem as a Multiple-Choice Item

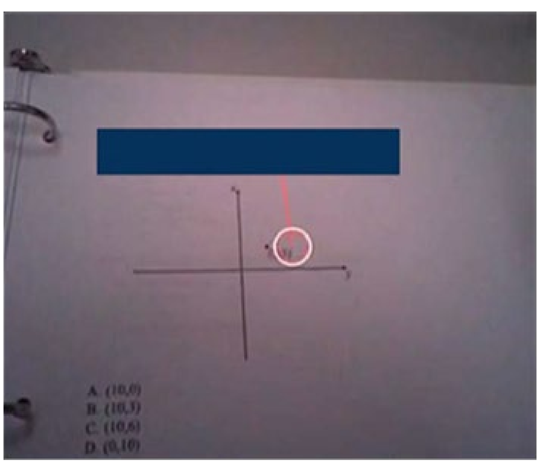
The point shown in the standard (x,y) coordinate plane below will be translated 7 coordinates up and 3 coordinates right. What are the resulting coordinates?



A. (10,0)
B. (10,3)
C. (10,6)
D. (0,10)
E. (6,10)

The problem shown in Figure 3.18 had two forms: one was the multiple-choice form referenced in Figure 3.18, and the other was a fill-in-the-blank form for the ordered pair. This problem was more difficult than expected based on the skill domain—a simple translation of both the *x* and *y* coordinates. Gaze patterns confirmed the common error for both the fill-in-the-blank and multiple-choice version of the item. The first translation is 7 coordinates up, which should result in an increase of the *y*-coordinate by 7. Ninety-one percent of the students who answered this correctly had a gaze event (fixation or saccade) on the *y*-coordinate in the graph after reading that part of the sentence, as shown in Figure 3.19. The common error of applying the first translation to the *x*-coordinate instead of the y-coordinate was seen in students who had a gaze event along the *x*-axis.

Figure 3.19. Eye-Tracking Capture of Adaptive Saccade Showing Correct Translation of *y*-Coordinate



Some students labeled the coordinates as they worked, which decreased the working memory load, which could be further validated with pupillometry data analysis. Fifty-six percent of the HIGH scorers wrote the coordinates, while only 33% of the MED scorers did. In interviews, several MED participants said that they wrote less on the math test to save time. This approach appears to be a maladaptive strategy, as illustrated by items like the one in Figure 3.19, as the time saved is offset by answering a question incorrectly that the student otherwise has the knowledge and skills to complete. Again, the idea of metacognitive awareness as to how to best utilize time appears to be a problem for the MED group in the "messy middle" (Gotwals et al, 2012). The problem was answered correctly by 47.1% of the multiple-choice participants, but only by 37.5% of the fill-in-the-blank participants. Since the most common incorrect answer (10,6) was one of the wrong answer choices, the scaffold of having the correct answer available provided only a slight improvement.

Table 3.5. Eye-Tracking Data for Translation Problem by Group

Categories	HIGH correct	HIGH incorrect	MED correct	MED incorrect	LOW correct	LOW incorrect
Total # of participants	5	4	5	7	2	2
Average total time (s)	28.3	17.0	24.8	27.0	20.5	29.9
Average read time (s)	6.3	3.4	5.3	5.6	3.9	1.9
Average scan time (s)	6.5	3.3	7.8	1.6	3.9	14.8
Average solution time (s)	15.5	10.3	11.7	19.8	12.6	13.2
% of participants who wrote their coordinates	80%	25%	20%	43%	0%	100%

This is one of the few items that LOW participants spent little time reading. Notice the extremely short time (1.9 seconds) for LOW students who answered incorrectly, which likely illustrates a lack of attention to detail as they frequently applied the translation to the incorrect coordinate. Also, students in the HIGH category who spent much less time reading were much more likely to make the mistake of reversing the coordinates.

Together, these exemplars provide evidence that test takers are using the skills that were targeted and that there are observable differences in the cognitive processes of students who answer correctly and incorrectly as well as between students who are in the different general scoring categories (HIGH, MID, LOW) that are used for many discrimination indices statistically.

Again, issues of metacognition and time management could be observed here. Many students used far more time than they should have on the Venn diagram problem. This finding was similar to observations from an earlier mode study on a graph-matching question. Several students who were using eye-tracking spent more than 4 minutes on this problem without using the graphing function on their calculators. Afterward, as part of the follow-up interview, one researcher asked the student to look back at that item and try to graph the answer choices on his graphing calculator. "I never thought of that," was the student's response. In less than 20 seconds, the student found the answer using the graphing function on his calculator; however, he had spent more than 5 minutes on that item without making any substantial progress during the original part of the study.

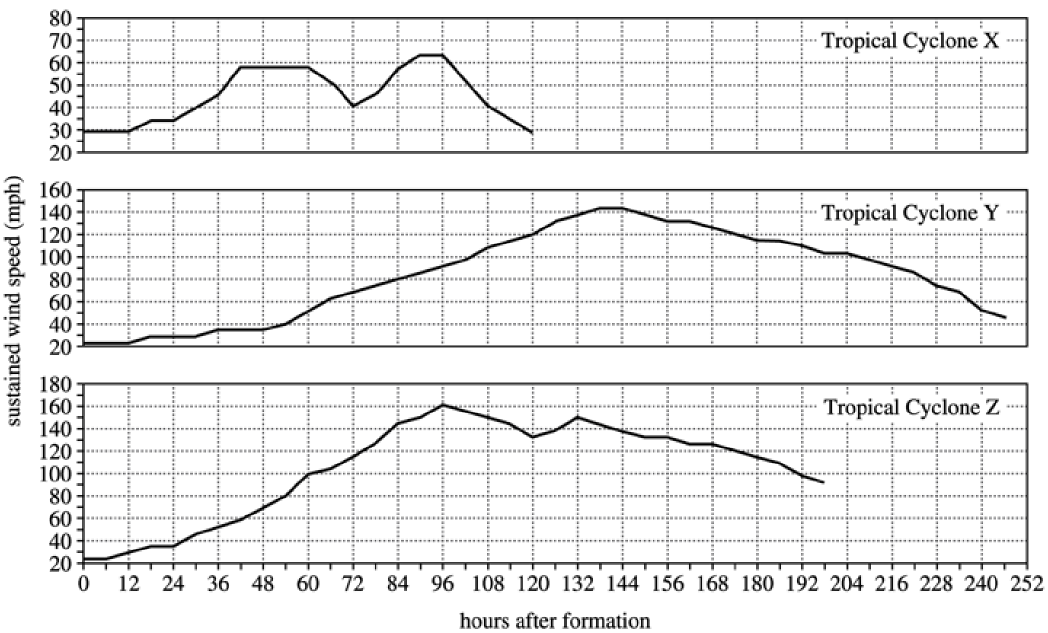
Since time management and metacognition have been listed by postsecondary instructors as being important skills needed to be successful when entering first-year college classes, the appropriate use of time on a standardized test may serve as a proxy for assessing those skills.

3.3.3.3.2 Science Findings

The importance of examining axes labels, column headers, legends, and keys was identified as a key differentiating skill. As was found in the validation studies for graphic literacy (Thomas and Langenfeld, 2018), test takers who spend time examining keys, legends, and labels are much more likely to answer items correctly. When presented passages with two or more graphics, many students engaged in non-solution behavior by searching for the necessary data in the wrong graphic, or in an incorrect component of the graphic (for example, looking at the wrong line in a multiple-line graph that was identified in the key or legend).

Writing, circling, and drawing on the graphic were associated with higher rates of correct answers, particularly for large or complicated graphics, as shown in Figure 3.20. This should be investigated further as these physical and cognitive processes do not transfer equally onto computer-based testing platforms.

Figure 3.20. Multipart Line Graph From Tropical Storm Passage



Many students struggled with finding data relevant to the correct tropical cyclone. Labeling relevant points made using the correct portion of the graphic more common.

Even when questions specifically state "Based on Table 1 . . ." students spent time looking at the wrong graphics. In short versions, graphics that were only required for a single item were placed within the question stem, which seemed to reduce the percentage of time wasted looking at the wrong graphics. Many students spent a long time reading and rereading questions, similar to maladaptive behaviors on the math section. Retrograde saccades were common when initially decoding questions, which is indicative of students attempting to understand specific words in the item and the overall meaning of the question. Items that required the use of two graphics were more difficult for students to answer correctly, even when the item was scaffolded or the second graphic was added to the stem of the relevant item.

Familiarity with the science content of the passage made it easier for students to answer questions even if the graphic was not identical to what they had seen in class. Several students commented that the carbon phase diagram was similar to phase diagrams that they had used in chemistry. This was a good example of transfer from the traditional three-phase diagram seen in textbooks and the four-phase diagram present with the two allotropes of solid carbon (graphite and diamond). Student familiarity with the two forms of solid carbon may have also helped.

Students liked the mix of passage lengths. Several stated that if you did not understand a passage that "you could try to do the easy items" ("Based on Table 1 . . .") and then move on to something that you knew better. Students stated that, "similar to the reading results," being familiar with or interested in a topic made the questions easier.

The shorter versions of passages were part of an initial examination of using a multistage adaptive testing format. That approach was eliminated from consideration when it was decided that the enhanced ACT must be available in comparable form for both paper and online testing. The findings about the use of marking up the passage and drawing lines connecting the *x* and *y* axes to the specific data point should be further probed when another study is conducted on identical problems in both paper and online testing. Overall, students who answered correctly exhibited the targeted cognitive processes, particularly using legends, keys, graph label axes, and table headers, to locate the appropriate information so that higher-level cognitive processes, such as evaluating a scientific claim, could be demonstrated, when necessary.

3.3.4 Enhanced ACT Cog Labs 2024

Another set of cognitive labs was completed in 2024. The goal of this study was to provide evidence of cognitive processes elicited by items that target skills that were previously on the legacy ACT as well as new skills targeted by the enhanced ACT across ELA, math, and science.

The research plan included collecting data in four different metropolitan areas: two in the Midwest, one in the Great Plains, and one in the Southwest. Students were first recruited through school contacts and prior participation in ACT research. All students were given a screening survey so that targeted demographics could be reached. Care was taken to include students who normally receive accommodations such as extra time because of a 504 plan or IEP. The screening survey results will be analyzed below.

The research plan involved a think-aloud protocol for all four multiple-choice sections of the test. The English, math, reading, and science teams selected items and passages to investigate with students. English, reading, and science counterbalanced the order of the passages to ensure that all items would be answered. Math counterbalanced the items by ensuring that items that had four foils on the A form had five foils on the B form, and vice versa. The A and B protocols also counterbalanced (Table 3.6) the order of subjects to minimize fatigue effects or the possibility that running behind schedule would adversely affect the collection of data for any one test section. Each section of the test was 22 minutes and stopped when the student finished the question they were working on when the timer went off.

Table 3.6. Counterbalanced Protocols for 2024 Cog Lab

Order	Protocol A	Protocol B
1	English Passage 1A (argumentative, 10 items) English Passage 1B (personal narrative, 8 items)	Reading Passage 1B (paired, 4 items) Reading Passage 1A (VQI, 5 items)
2	Math (odd-numbered items have four options, 11 items)	Science Passage 1B (biology CV, 7 items) Science Passage 1A (engineering RS, 7 items)
3	Reading Passage 1A (VQI, 5 items) Reading Passage 1B (paired, 4 items)	English Passage 1B (personal narrative, 8 items) English Passage 1A (argumentative, 10 items)
4	Science Passage 1A (engineering RS, 7 items) Science Passage 1B (biology CV, 7 items)	Math (odd-numbered items have five options, 11 items)
5	Survey and interview	Survey and interview

After the think-aloud, participants completed a survey and a guided interview. At the completion of the two-hour study, students were given a $50 electronic Amazon gift card.

Content experts identified the targeted skills for each item and created a Q-matrix coding sheet for proctors and secondary coders to identify what adaptive skills the students used. For some items, these were labeled as the optimal (most efficient) skills to determine the correct answer. Suboptimal skills that might take more time were also listed. Anticipated errors or other maladaptive behaviors (e.g., random guessing) were also included in the Q-matrix worksheet. Additional spaces on the Q-matrix were available for proctors and coders to identify other adaptive skills (skills that moved meaningfully toward the correct answer) or maladaptive skills (behaviors and cognitive processes that did not move meaningfully toward the correct answer). It was anticipated that students who used adaptive skills would answer the items correctly at a much higher rate than students who did not. From previous cog labs for both the ACT and WorkKeys assessments, it was anticipated that some students would use some targeted and some maladaptive skills, which corresponds to what Gotwals & Songer (2010) refer to as "the messy middle," students who inconsistently apply skills when they are at intermediate stages of a learning progression.

Video of the student's face was synced with a second video of the test booklet so that all sessions could be double coded. Proctors coded in real time as the students talked through their cognitive processes. Because it can be difficult to record all the processes in the Q-matrix in real time, each student's responses were double coded based on the video by a content specialist in that area (ELA, math, or science). All proctors and coders were trained on the process prior to participating in the collection or coding of data. When the two coders did not agree, a senior researcher for that content area watched the video and made the final decision used for analysis.

3.3.4.1 Demographics and Screening Survey Findings

Students signed up by completing a screening survey that included their availability to participate in the cog lab session. A total of 75 students completed the screening survey. Students were sent an informed consent form, a nondisclosure agreement (NDA), and a list of the available cog lab sessions in their area. Students participating in the cog lab study—as well as the parent(s) or caregiver(s) of students 17 or younger—were required to sign the NDA and consent forms.

Ultimately, 27 participants completed cog lab sessions. Some students were unable to find an available time in their area or were no-shows for a scheduled cog lab. Of these 27 participants, 11 (40.7%) students identified as female, 1 identified as nonbinary/third gender (3.7%), and 1 (3.7%) preferred not to identify a gender. In terms of accommodations, 3 students (11.1%) reported having an accommodation for standardized testing (extra time because of ADHD or another reason). Attempts were made to balance student demographics across the two study forms in terms of grade, gender, and race. Table 3.7 presents a comparison of student demographics across the study forms.

Table 3.7. Cog Lab Participant Characteristics by Study Form

Variable	Form A *n*	Form A %	Form B *n*	Form B %
Grade	12	—	14	—
Grade: 10th	6	50.0	4	28.6
Grade: 11th	1	8.3	5	35.7
Grade: 12th	5	41.7	5	35.7
Gender	12	—	14	—
Gender: female	5	41.7	6	42.9
Gender: male	6	50.0	7	50.0
Gender: non-binary/third gender	1	8.3	0	0.0
Gender: prefer not to say	0	0.0	1	7.1
Hispanic origin	12	—	14	—
Hispanic origin: yes	0	0.0	2	14.3
Hispanic origin: no	12	100.0	12	85.7
Race	12	—	14	—
Race: Asian	3	25.0	0	0.0
Race: Asian, Black/African American, Two or more races	0	0.0	1	7.1
Race: Asian, White	0	0.0	1	7.1
Race: Asian, White, Two or more races	0	0.0	1	7.1
Race: Black/African American	0	0.0	3	21.4
Race: Black/African American, Native Hawaiian/Other Pacific Islander	0	0.0	1	7.1
Race: Black/African American, White	1	8.3	0	0.0
Race: White	8	66.7	7	50.0
Receive testing accommodations	12	—	14	—
Accommodations: yes	3	25.0	0	0.0
Accommodations: no	9	75.0	14	100.0

Results from the screening survey are reported for the 75 respondents (note that not all respondents answered all survey questions) as well as for the 27 students who completed the cog lab and are a subset of the 75 initial respondents. Respondents were asked about test mode when they took state proficiency tests in Grades 3 through 8, with the results shown in Table 3.8. The pattern was the same across groups. For the 75 respondents, the largest percentage of students (45.7%, n=32) took these tests online, followed by a mix of online and paper (38.6%, n=27) and paper only (15.7%, n=11). Similarly, of the 27 participants who also completed the cog labs, the largest percentage (46.2%, n=12) took these tests online, followed by a mix of online and paper (42.3%, n=11) and paper only (11.5%, n=3). One person did not respond to this question.

Students were then asked whether they prefer to take tests like the ACT or state proficiency tests on paper or online, based on the test format (multiple choice, essays, or short constructed response). The results in Table 3.9 show that both groups prefer to take all three types of tests online, followed by paper and no preference.

Table 3.8 Test Mode for State Testing in Grades 3–8

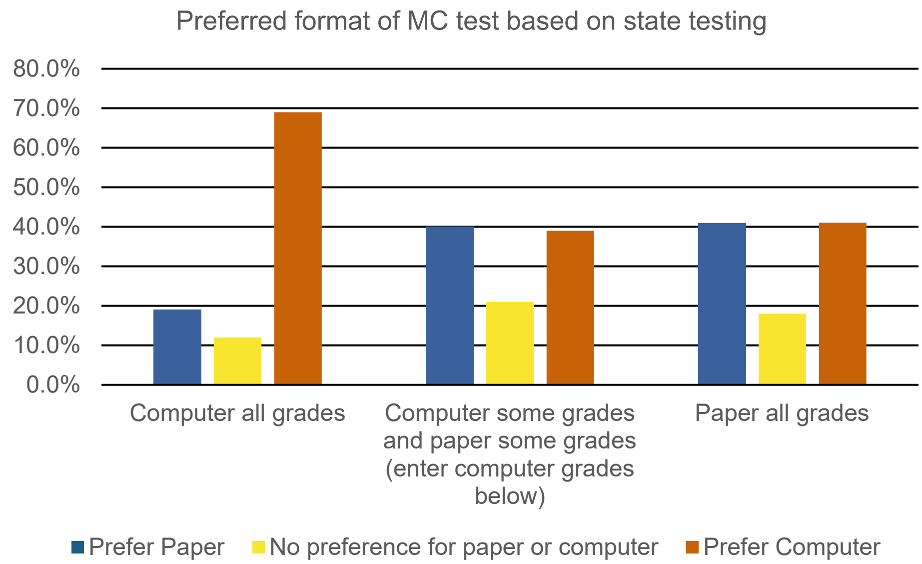
Test mode	Screening survey *n*	%	Cog lab participants (*n* = 27) *n*	%
Online all grades	32	45.7	12	46.2
Online some grades and paper some grades	27	38.6	11	42.3
Paper all grades	11	15.7	3	11.5

Table 3.9. Test Mode Preference by Test Format

	Screening survey participants (*n* = 75)						Cog lab participants (*n* = 27)					
	Online		No preference		Paper		Online		No preference		Paper	
Test format	*n*	%	*n*	%	*n*	%	*n*	%	*n*	%	*n*	%
Multiple-choice tests like the state test and the ACT	38	53.5	11	15.5	22	31.0	13	48.1	4	14.8	10	37.0
Tests with essays	53	74.6	8	11.3	10	14.1	22	81.5	2	7.4	3	11.1
Tests with written short answers	46	64.8	12	16.9	13	18.3	19	70.4	4	14.8	4	14.8

The enhanced ACT test will be available in both paper and online formats. Format preference was affected by how students had taken state testing previously.

Figure 3.21. Preferred Format of Multiple-Choice (MC) Testing by State Testing Format in Grades 3–8



3.3.4.2 English Findings

Generally, English items performed as expected relative to the constructs being tested. Students who used one or more adaptive skills when working through an item were able to answer that item correctly 69.4% of the time across both passages. Students who used *only* adaptive skills answered items correctly 89.9% of the time, while students who used one or more maladaptive skills answered correctly only 8.9% of the time. Students who used at least one adaptive skill were correct at a rate 60.5% higher than students who used at least one maladaptive skill, and students who only used adaptive skills were correct at a rate 80.0% higher than students who used a maladaptive skill. Overall, students were correct on 66.1% of the items they attempted. These results, along with the results of other test sections, are summarized in Table 3.25 at the conclusion of this chapter. One English item, a grammar item testing punctuation of an appositive phrase, was particularly difficult for students, with many students using at least one adaptive skill but still getting the item wrong; this item will be discussed further below.

3.3.4.2.1 Item and Passage Selection

Two new essay types were selected for the cognitive study: an argumentative essay and a short essay. The argumentative essay was standard length (approximately 340 standard words) and was accompanied by ten items. This social sciences essay, which argued that Congress should retire the penny, featured five Production of Writing items that targeted skills in developing and supporting arguments and five additional items across all three reporting categories—Production of Writing (POW), Knowledge of Language (KLA), and Conventions of Standard English (CSE)—that did not directly call for analyzing the essay's claims. The short essay (approximately 185 standard words) was a narrative about navigating a universally accessible hiking path. There were eight items in the short unit, representing all three reporting categories. Items in both essays were selected to elicit evidence in support of the most significant changes to the English section: adding argumentative content, shortening some essays, and adding stems to all items.

3.3.4.2.2 Argumentative Content

Argumentative items require students to utilize knowledge and skills such as identifying claims, choosing relevant and appropriate supporting evidence, and making choices about organizing arguments logically. All participants noted that they had learned the claim-evidence-reasoning model of argumentation in English classes either before or in high school, which corresponds to expert panel feedback indicating that evaluating and constructing arguments is a focus in ELA classrooms.

The argumentative essay chosen for the cog lab presented a classic argumentative framework: the first paragraph outlined the problem and stated a call to action, the middle paragraphs provided reasoning and support, and the last paragraph concluded the argument. As such, the essay provided ample opportunities for a variety of argumentative items, including determining the best support for a particular claim, analyzing the relevance of specific evidence to the author’s argument, sequencing an argument logically, providing a conclusion that restates the essay’s primary claims, and evaluating the essay’s purpose. The other five items in this unit were standard item types, including items focused on precision in language, punctuation, and organization.

Students who used the targeted skills were much more likely to answer correctly than those who did not. On average, students demonstrated appropriate skills when working through the argumentative items, with 91.4% using at least one adaptive skill when attempting to answer. 72.6% of those students answered correctly, as opposed to 14.7% answering correctly when using at least one maladaptive skill. When students used only adaptive skills, they answered correctly 89.0% of the time. Overall performance of the argumentative items was similar to that of non-argumentative items, and no students noted surprise or confusion at seeing this type of content on the English test.

Table 3.10. English Passage A (Argumentative) Data

Item	Attempted	Used one or more adaptive skills	Adaptive only correct (%)	Used one or more maladaptive skills	Maladaptive correct (%)	Difference adaptive correct and maladaptive correct (%)	Overall correct (%)	ARG items
1	27	27	59.3%	6	0.0%	59.3%	59.3%	—
2	27	20	95.0%	8	25.0%	70.0%	77.8%	ARG
3	27	27	85.2%	5	40.0%	45.2%	85.2%	ARG
4	27	26	30.8%	18	0.0%	30.8%	29.6%	
5	26	21	81.0%	8	12.5%	68.5%	65.4%	
6	24	24	54.2%	7	0.0%	54.2%	54.2%	ARG
7	23	21	66.7%	6	0.0%	66.7%	60.9%	
8	21	19	68.4%	8	12.5%	55.9%	61.9%	ARG
9	20	21	76.2%	2	0.0%	76.2%	80.0%	
10	17	16	56.3%	6	0.0%	56.3%	52.9%	ARG
Average: total	23.9	22.2	66.7%	7.4	8.1%	58.6%	62.8%	—
Average: non-ARG only	24.6	23.2	61.2%	8.0	2.5%	58.7%	57.7%	—
Average: ARG only	23.2	21.2	72.6%	6.8	14.7%	57.9%	68.1%	—

The difference in average performance between the non-ARG and ARG items is largely driven by Item A4 in this small sample. This item tested comma usage with an appositive phrase, an item type that can appear in any essay. Most students either misunderstood the function of the phrase being tested or applied the rule incorrectly, maladaptive behaviors that are unrelated to the passage format (argumentative) or passage length. Though 96.3% of the students who attempted the item used at least one adaptive skill (most often reading the phrase in context to assess how it functioned), a high percentage of students also used at least one maladaptive skill, and only 30.8% of students got the item correct. The enhanced ACT will continue to test this skill, as punctuating appositive phrases improves clarity and aligns to common state standards; however, there will be fewer of these items per form.

3.3.4.2.3 Passage Type: Long Versus Short

Short essays will feature item types from all reporting categories. The eight items attached to the short essay consisted of two POW items, three KLA items, and three CSE items. Because of a formatting issue that affected how students answered one of the KLA items (B6), only seven of the eight items are included in this analysis.

Table 3.11. English Passage B (Short Essay)

Item	Attempted	Used one of more adaptive skills	Adaptive only correct (%)	Used one or more maladaptive skills	Maladaptive correct (%)	Difference adaptive correct and maladaptive correct (%)	Overall correct (%)	
1	24	24	83.3%	5	40.0%	43.3%	83.3%	CSE
2	25	25	56.0%	8	0.0%	56.0%	56.0%	CSE
3	23	23	91.3%	2	0.0%	91.3%	91.3%	POW
4	23	23	60.9%	9	0.0%	60.9%	60.9%	KLA
5	22	22	77.3%	5	0.0%	77.3%	77.3%	KLA
6	—	—	—	—	—	—	—	—
7	19	17	70.6%	6	16.7%	53.9%	63.2%	CSE
8	17	17	64.7%	6	16.7%	48.0%	64.7%	POW
Average	21.9	21.6	72.2%	5.9	9.8%	62.4%	71.2%	—

Note**:** Item B6 was excluded because a formatting error influenced how students answered the item.

Of the students who answered the items in the short essay unit, 98.7% used at least one adaptive skill; 72.2% of those students answered correctly. In contrast, only 9.8% of students who used one or more maladaptive skills answered that item correctly. Overall, on the short essay unit, a higher percentage of students used at least one adaptive skill and answered items correctly at a higher rate than they did on the long essay. While there may be several factors at play in this difference, including variance caused by item A4 and the higher number of complex POW items in the long unit, early pretest analyses have indicated that short essay units may trend slightly easier overall than long essay units. This will continue to be monitored as enhanced ACT materials are developed and pretested.

Table 3.12. Comparison of Long and Short Essays

Passage	Used one or more adaptive skills	Adaptive only correct (%)	Used one or more maladaptive skills	Maladaptive correct (%)	Difference adaptive correct and maladaptive correct (%)	Overall correct (%)
A (long)	22.2	66.7%	7.4	8.1%	58.6%	62.8%
B (short)	21.6	72.2%	5.9	9.8%	62.4%	71.2%

3.3.4.2.4 Stems

All cog lab items featured stems. Students did not comment on the addition of stems to formerly stemless item types or express confusion with the new standard stem wording for CSE items. Moreover, 100% of participants selected “strongly agree” or “agree” for the statement “I was able to understand what the question was asking me to do, even if I did not know how to answer the specific question” as it pertained to the English section. However, analysis of student responses showed that some stem language was found to either impede initial understanding of the question or potentially drive maladaptive behaviors; specific items are addressed below.

Item B5, Short Essay

This stem asked students to create “stylistic parallelism” within a sentence. Though 77.3% of students who used at least one adaptive skill answered correctly, many students noted initial confusion with this phrasing. Upon reading the answer options, most were able to determine that the correct answer needed to create a sentence with parallel structure. This was an experimental stem style designed by the ELA development team to expand KLA item options by asking students to make specific choices for stylistic effect. As such, the cog lab provided important feedback, and we were able to ascertain that most students were not considering style when answering this item but were arriving at the right answer by evaluating sentence structure. Despite the stem phrasing, most students quickly recognized how to answer the item by working through the answer options.

Item A2, Long Essay

While the “stylistic parallelism” item was an experimental, nonstandard stem, we also discovered a potential problem with a more standard stem style. Used in a topic development item in the argumentative essay set, the stem read: “The writer wants to pose a rhetorical question that helps support the claim in the preceding sentence. Which choice best accomplishes that goal?” In this case, the use of the word “rhetorical” in the stem was meant to convey that all the options function in the same way, allowing students to evaluate which choice is best in relation to the claim rather than evaluating the type of question presented in each answer choice. In many cases, the stem had the opposite effect, as students focused more on the rhetorical nature of each answer choice than on evaluating the choices in terms of claim and support. These students chose their answer based on which choice felt like the *most* rhetorical question while ignoring the claim it was meant to support. This may be an example of a case in which students substitute an easier question for a more difficult one (Kahneman, 2011). Identifying when students are likely to substitute an easier question can be used to improve the item development feedback loop (Thomas, 2020) and should result in higher quality items with fewer items failing in pretest.

Items A10, Long Essay, and B6, Short Essay

Based on feedback from the external panel that ACT should more directly explore audience in English items, ELA staff created two items for the cog lab that addressed audience. The first item (A10) was a topic development item that asked students to evaluate how well the author met a defined purpose for a specific audience. The second item (B6) was a style item that asked students to choose wording that might appeal to a specific audience. Most students were able to answer A10 correctly by primarily focusing on purpose alone; only 41% of students who attempted to answer the item referred to the audience when answering. Twenty-seven percent of students who attempted to answer item B6 referred to the audience when answering, but as noted above, a formatting error skewed the data for this item, and it was not included in the larger analysis. Regardless, student responses to these two items indicate that further refinement of audience items is required before operational use on the enhanced ACT English test.

3.3.4.2.5 Recommendations

It was recommended that ACT proceed with the proposed changes to the enhanced ACT English test as the items were able to elicit evidence of the cognitive processes of the targeted skills for argumentative essays as well as essays of both lengths. Additionally, the cog lab results indicate that students are engaging in complex multistep processes with multiple-choice items, which addresses the perceived weakness that multiple-choice items cannot assess higher-level thinking skills.

3.3.4.3 Math Findings

The items on the math test performed as expected according to the targeted math constructs. Students who used one or more adaptive skills when attempting an item were able to correctly answer that item 78.4% of the time, while students who used one or more maladaptive skills (guessing, using incorrect math, etc.) were correct only 20.0% of the time (similar to random guessing). There was one item on which some students used maladaptive skills but answered correctly at rates substantially higher than random guessing. Students who used only adaptive behaviors answered items correctly on 92.1% of attempts, which was 72.1% higher than students who used one or more maladaptive behaviors. Overall, students were correct on 54.1% of the items that they attempted.

3.3.4.3.1 Item Selection

Two types of items were chosen for the think-aloud cognitive lab. The first type were items where every answer option should be read and considered before answering. The second type were items where students may consider working backward from the answer options to learn information or find their answer. These items covered a variety of math topics primarily in algebra, functions, and statistics; however, this final set did not include any geometry items.

Table 3.13. Math Item Performance

Item	Attempted	Used one or more adaptive skills	Adaptive only correct (%)	Used one or more maladaptive skills	Maladaptive correct (%)	Difference adaptive correct and maladaptive correct (%)	Overall correct (%)
1	26	11	90.9	11	9.1	81.8	42.3
2	27	21	71.4	12	25.0	46.4	59.3
3	26	17	64.7	15	13.3	51.4	46.2
4	27	17	100.0	11	45.5	54.5	70.4
5	27	26	100.0	2	100.0	0.0	100.0
6	26	8	75.0	22	13.6	61.4	23.1
7	26	17	88.2	4	50.0	38.2	84.6
8	27	15	40.0	22	13.6	26.4	26.0
9	25	11	100.0	14	21.4	78.6	52.0
10	24	12	75.0	9	11.1	63.9	45.8
11	18	13	46.2	8	12.5	33.7	38.9
Avg	25.4	15.3	78.4	11.8	20.0	58.4	54.1

3.3.4.3.2 Items With Five Answer Options Versus Items With Four Answer Options

The most significant change to the items for the enhanced ACT math test is the change from five answer options to four answer options. This change led to some concerns about items getting easier with fewer options. The likelihood of a student randomly guessing correctly increases from 20% to 25%. We also wanted to investigate whether students were more likely to backsolve items with only four options rather than five. The decrease from five options to four could lighten the reading load as well as reducing the percentage of students who engage in nonsolution behaviors (Wise & Smith, 2011) such as random guessing, as previous cog labs have demonstrated that some students give up when items are longer.

The results from the study show that items with five options were answered correctly 44.8% of the time and items with four options 53.5% of the time. Those numbers are consistent with results from our field test studies. Students who used targeted skills got four-option items correct 79.7% of the time and five-option items correct 69.1% of the time. Students who took maladaptive approaches got correct answers 19.0% of the time with four options and 18.6% of the time with five options (refer to Table 3.14).

Table 3.14. Comparison of Items with Four and Five Answer Options

Item	Total attempts	Used one or more adaptive skills	Adaptive only correct (%)	Used one or more maladaptive skills	Maladaptive correct (%)	Difference adaptive correct and maladaptive correct (%)	Overall correct (%)
4 answer option pairs	127	68	79.7	58	19.0	60.7	53.5
5 answer option pairs	125	56	69.1	70	18.6	50.5	44.8

Each student received a mix of items with four options and five options. No student made a comment about this mix of items while doing the think-aloud or during the interviews and surveys afterward. An item having four options or five options did not seem to substantially change how students worked and interacted with the items in terms of choosing a targeted solution path, working backward from the options, using answer elimination techniques, or using a calculator.

Students demonstrated adaptive skills 58.3% of the time on four-option items and 54.4% of the time on five-option items. The study supports the idea that students still provide evidence of what they know and can do when answering a math item with four options instead of five options. Taken together, evidence supports the idea that response processes do not change drastically in evaluating four-option items compared to five-option items, thus supporting the decision to reduce to four response options.

3.3.4.3.3 Calculator Usage

Students used a calculator on 26.0% of attempts on four-option items and on 28.8% of attempts on five-option items. Students answered correctly 54.6% of the time on four-option items and 50.0% of the time on five-option items when using a calculator. This means that students answered correctly about half of the time and incorrectly about half of the time when using a calculator. These results are evidence that a calculator alone does not provide a pathway to a correct answer on these math items. For some items, students used a calculator when it was maladaptive and wasted time on paths that could not lead to a correct solution.

3.3.4.3.4 Recommendations

It was recommended that ACT proceed with the proposed changes to enhanced ACT math as the items were able to elicit evidence of the cognitive processes of the targeted skills regardless of the number of answer options. Additionally, the cog lab results indicate students engaging in complex multistep processes with multiple-choice items, which addresses a perceived weakness that multiple-choice items cannot assess higher-level thinking skills.

3.3.4.4 Reading Findings

The items in the reading section performed as expected relative to the targeted reading constructs. Students who used one or more adaptive skills when working through an item were able to answer that item correctly 70.0% of the time across both passages, a rate 62.0% higher than students who used at least one maladaptive skill (8.0%). Students who used only adaptive skills answered items correctly 92.2% of the time, a rate 84.2% higher than students who used a maladaptive skill. Overall, students were correct on 68.4% of the items they attempted.

3.3.4.4.1 Item and Passage Selection

Two passages were selected for this study: a paired passage set about Cherokee storytelling traditions and a VQI passage and graphic about informal giving trends. Given concerns from the external expert panel that paired units may be more difficult, passages were selected to be as similar as possible in terms of genre and complexity; both were sociology passages at the same text complexity level. The cog lab focused on items from the Integration of Knowledge (IKI) reporting category, which includes three subcategories: Synthesis of Multiple Texts (SYN), Visual and Quantitative Information (VQI), and Arguments (ARG). The item distribution across both passages was as follows: three VQI items, three SYN items, and three non-IKI items assessing the whole passage or large parts of the passage (refer to Tables 3.15 and 3.16).

Table 3.15. Reading Item Performance, Passage A: Informal Giving (VQI)

Item	Attempted	Used one of more adaptive skill	Adaptive only correct (%)	Used one or more maladaptive skills	Maladaptive Correct (%)	Difference adaptive correct and maladaptive correct (%)	Overall correct (%)	Item type
1	26	25	36.0%	15	0.0%	36.0%	34.6%	TST
2	26	26	57.7%	6	0.0%	57.7%	57.7%	PPV
3	25	25	72.0%	8	12.5%	59.5%	72.0%	VQI
4	24	23	69.6%	8	37.5%	32.1%	66.7%	VQI
5	25	25	84.0%	5	20.0%	64.0%	84.0%	VQI
Average	25.2	24.8	63.7%	8.4	11.9%	51.8%	62.7%	—

Table 3.16. Reading Item Performance, Passage B: Cherokee Storytelling (Pair)

Item	Attempted	Used one of more adaptive skill	Adaptive only correct (%)	Used one or more maladaptive skills	Maladaptive Correct (%)	Difference adaptive correct and maladaptive correct (%)	Overall correct (%)	Item type
1	27	26	80.8%	6	0.0%	80.8%	77.8%	CLR
2	26	26	88.5%	3	0.0%	88.5%	88.5%	SYN
3	25	24	83.3%	5	0.0%	83.3%	80.0%	SYN
4	21	21	47.6%	10	10.0%	37.6%	52.4%	SYN
Average	24.8	24.3	76.3%	6	4.2%	72.1%	75.8%	—

3.3.4.4.2 Passage and Item Type: VQI Versus Paired

The enhanced ACT includes a domain sampling plan that would allow for students to be scored on either a VQI passage set or a paired passage set on any given form. In the cog lab, students were presented with one VQI passage with five items and one paired passage with four items.

VQI items and SYN items both call for integrating and analyzing information from multiple sources, though the sources are different. VQI items ask students to understand and analyze information presented in visual or quantitative format alongside a text, while SYN items ask students to integrate knowledge across two related texts. VQI and SYN items performed similarly in the cog lab. In the VQI unit, 98.6% of students who answered the VQI items used one or more adaptive skills, and 75.3% of students who used an adaptive skill answered correctly. In the paired unit, 98.6% of students who answered the SYN items used one or more adaptive skills, and 74.6% of students using an adaptive skill answered correctly (refer to Table 3.17).

Table 3.17. Comparison of VQI and SYN Item Performance

Items	Attempted	Used one of more adaptive skills	Adaptive correct (%)	Used one or more maladaptive skills	Maladaptive correct (%)	Difference adaptive correct and maladaptive correct (%)	Overall correct (%)
VQI	24.7	24.3	75.3%	7.0	23.8%	51.5%	74.3%
Pair	24.0	23.7	74.6%	6.0	5.6%	69.1%	75.0%

After working through all cog lab items, students were asked whether either passage set seemed harder than the other. In response, 44.4% of cog lab participants stated that the paired passage set was more difficult to work through, 29.6% indicated that the VQI passage set was more difficult, and 22.2% noted that the passage sets felt the same. One student (3.7%) did not answer directly, stating that the English, reading, and math sections were all hard.

Students noted a variety of reasons for feeling that one passage set was more difficult than the other, but a few themes did emerge: Students tended to identify the paired passages as signaling compare-and-contrast work, while they tended to view the graphic in the VQI unit as an aid to their understanding of the passage. While VQI and SYN items both test a variety of skills within the IKI reporting category, some students may identify the passages as automatically requiring different knowledge and skills.

While the above themes did emerge, preference for passage type was not limited to the skills students believed they would be required to demonstrate with each passage type. Several of the students who felt that the pair was harder pointed to unfamiliar Indigenous terms or the presence of a long paragraph in one of the passages, both of which speak to the passage being difficult for a reason other than the paired structure. Likewise, some of the students who thought the VQI passage set was harder pointed to the content of the passage or the density of the information rather than the inclusion of a graphic.

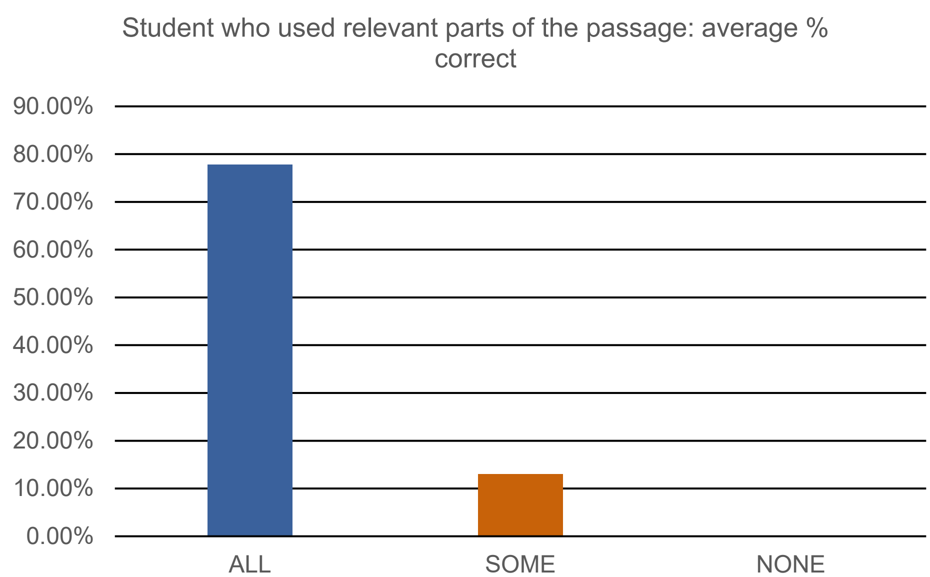
Some students also registered surprise at seeing a graphic in a reading unit, even though they understood how to answer the items using the passage and graphic. This echoes comments from the previous VQI cog lab. This may be remedied, at least in part, through updates to test prep materials and communication to key stakeholders at the state, district, and student levels.

3.3.4.4.3 Whole-Passage Items

The reading section also featured three whole-passage items. While these are not new item types, the cog lab presented an opportunity to verify that these items target skills found in reading standards calling for textual analysis that can best be achieved through engagement with multiparagraph passages. These standards include analyzing how ideas develop over the course of a text; explaining how individuals, ideas, or events interact over the course of a text; and analyzing how an author's choices regarding how to structure specific parts of a text contribute to the overall meaning of the text. The three whole-passage items included in the cog lab assessed different skills; one was from the KID reporting category, and two were from the CAS category. On average, the items performed as expected according to the construct. Of the students who answered the whole-passage items, 97.5% used one or more adaptive skills. Students who used at least one adaptive skill answered correctly 58.4% of the time. One of the items, which required students to analyze the text's structure, was particularly difficult, with only 34.6% of students answering correctly.

Students who used all the relevant parts of the passage to answer these items were correct far more frequently than students who used only some of the relevant parts of the passage or none of the passage (refer to Figure 3.22).

Figure 3.22. Percent Correct Based on Amount of Relevant Passage the Student Used



3.3.4.4.4 Recommendations

It was recommended that ACT proceed with the proposed changes to the enhanced ACT reading test as the items were able to elicit evidence of cognitive processes reflecting the targeted skills. Specifically, the items showed that multiparagraph passages can elicit cognitive processes that require integrating information across multiple paragraphs. Additionally, the cog lab results indicate that students engage in complex multistep processes with multiple-choice items, which addresses a perceived weakness that multiple-choice items cannot assess higher-level thinking skills.

3.3.4.5 Science Findings

The items on the science test performed as expected according to the targeted science constructs. Students who used one or more adaptive skills when attempting an item were able to correctly answer that item 73.8% of the time across both passages, a rate that was 44.3% higher than students who used one or more maladaptive skills (29.5%). Students who used only adaptive behaviors answered items correctly on 99.5% of attempts, which was 70.0% higher than students who used one or more maladaptive behaviors. Overall, students were correct on 68.6% of the items that they attempted. There was one item that required scientific background knowledge (BGK) that many students used maladaptive skills to answer but were able to guess correctly. This item will be addressed below.

3.3.4.5.1 Item and Passage Selection

For Passage A, several items were selected that were expected to be easier based on the construct map but used specific engineering and design thinking skills and terminology. Another item, which involved finding the difference between two decimal values in a table, was selected to investigate whether students would use a calculator. Consequently, it was expected that the overall difficulty of the items selected for Passage A would be slightly lower than that of the items selected for Passage B. Additionally, the science terminology used in Passage B included several terms that a student would not have studied specifically even though the general process would have been covered in a biology class.

Three items that required science content knowledge were selected, including one that involved a common misconception (meiosis). These items required the student to integrate prior science knowledge with information from the passage to answer correctly. Several items were also selected because they required integrating information from two graphics to answer the item. Student results for each passage and item are shown in Tables 3.18 and 3.19.

Table 3.18. Science Engineering Passage (A) Data

Item	Attempted	Used one or more adaptive skills	Adaptive only correct (%)	Used one or more maladaptive skills	Maladaptive correct (%)	Difference adaptive correct and maladaptive correct (%)	Overall correct (%)	Secondary code (if present)
1	26	23	60.9	16	37.5	23.4	61.5	Two graphics
2	24	23	100	1	100	0	100.0	—
3	24	21	38.1	18	11.1	27.0	33.3	Two graphics
4	22	20	90.0	7	42.9	47.1	81.8	
5	20	18	77.8	11	45.5	32.3	70.0	BGK
6	20	18	100	4	50.0	50.0	90.0	EDT
7	20	17	100	5	40.0	60.0	85.0	EDT
Mean	22.3	20.0	80.0	8.86	33.9	46.1	73.7	—

Table 3.19. Science Conflicting Viewpoints Biology Passage (B) Data

Item	Attempted	Used one or more adaptive skills	Adaptive only correct (%)	Used one or more maladaptive skills	Maladaptive Correct (%)	Difference adaptive correct and maladaptive correct (%)	Overall correct (%)	Secondary code (if present)
1	27	26	76.9	9	22.2	54.7	74.1	—
2	27	19	73.7	17	58.8	14.9	74.1	BGK
3	26	25	88.0	6	33.3	54.7	84.6	—
4	26	24	58.3	12	0	58.3	53.8	—
5	25	23	78.3	10	30.0	48.3	73.0	—
6	25	20	60.0	15	26.7	33.3	52.0	—
7	22	17	29.4	18	11.1	18.3	31.8	BGK
Mean	25.4	22.0	68.2	12.4	26.4	41.7	64.0	—

3.3.4.5.2 Passage Type: Conflicting Viewpoints (CV) Versus Research Summary (RS)

Two of three potential passage types were included in this cog lab. A Research Summary (RS) is a medium-length passage (approximately 275 standard words) that describes one or more related experiments along with experimental results. This passage included two tables and one clustered bar graph. This RS passage described simulated experiments using solar panels to reduce the need for fossil-fuel-powered generators. Conflicting Viewpoints (CV) passages are the longest passage type (approximately 350 standard words) and present two or more theoretical models that address the same scientific phenomenon. This passage contained two viewpoints that differed on the explanation of a biologic phenomenon. The overall percentage correct for all items attempted was 73.7% for RS Passage A and 64.0% for CV Passage B.

As anticipated, Passage A was in general less difficult than Passage B. Passage B also took more time for students to complete. Several students who had Passage B first (Protocol B), were unable to complete Passage A. This may have contributed to the relative difficulty of Passage A as those students frequently struggled with Passage B.

During the structured interview, students were asked about the science passages and which passage was easier. These results were mixed; however, there were several comments similar to those for the reading passages: Some students preferred to have information in the text clarified with graphics, while others preferred the compare-and-contrast nature of the CV. Some students commented that the passage difficulty was more related to the scientific terminology and familiarity with the topic. The mix of responses suggests that maintaining a mix of passage formats and topics would allow for students to better maximize their strengths and show what they know and can do in science.

3.3.4.5.3 Use of Targeted Science Skills

On both passages, students who used one or more targeted skills were able to successfully answer items, especially those items that relied on transferable science skills without science BGK as shown in Table 3.20.

Table 3.20. Student Performance Across Both Science Passages

Item type	Adaptive skills correct (%)	Maladaptive skills correct (%)	Difference (adaptive – maladaptive)	Used only adaptive correct (%)	Difference (only adaptive – maladaptive)	Overall correct (%)
Overall	73.8	29.5	44.3	99.5	70.0	68.6
Transferable science skill	76.7	35.8	27.4	99.4	63.6	70.9
BGK	61.1	42.5	18.6	100	57.5	47.8

The items that required more cognitively complex skills and the use of multiple graphics were the most difficult items on Passage A. Item 1A required finding one piece of data (the highest temperature) in Table 1 and using that information to determine which data to use in Figure 1. Item 3A required that students reorder the sites from lowest to highest based on data from Table 1 and then use that information to interpret trends in Figure 1. Several students ignored

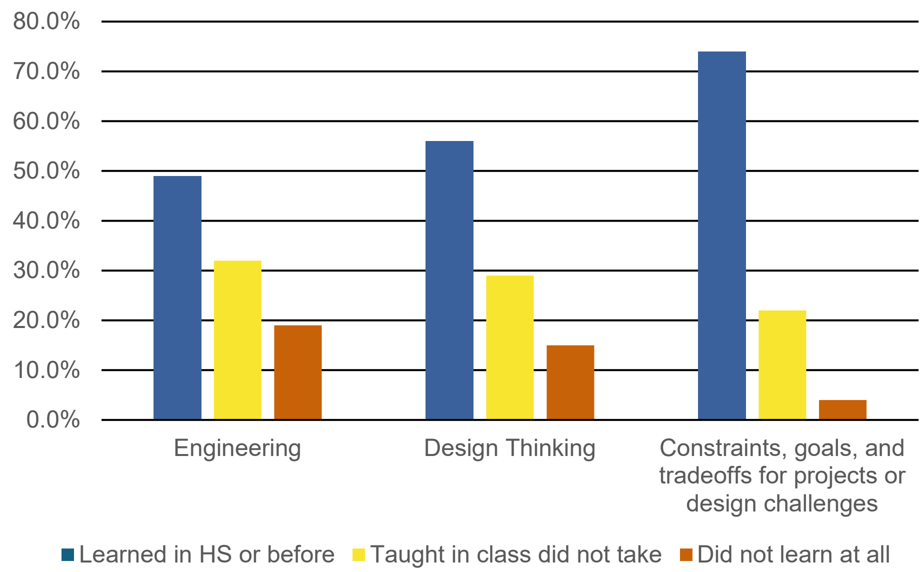
Table 1 and focused only on the trend in Figure 1. This is consistent with earlier findings about graphic literacy: Many test takers focus on only one graphic when items require using information from two graphics (Langenfeld, et al., 2020; Thomas & Langenfeld, 2017a). Eventually, a multidimensional IRT or Q-matrix may be available to identify what skills a student likely possesses based on wrong answer selections as well as correct answers. This data also supports claims that these items can elicit complex multistep solution processes, addressing the common complaint that multiple-choice items are unable to elicit evidence of complex cognitive processes.

The large difference between students who used at least one adaptive skill and those who used maladaptive skills (44.3%) provides strong evidence that the science items and passages can be used to support claims that students who answer correctly preferentially possess and use the targeted skills. The 70.0% difference between students who use only adaptive behaviors and those who use maladaptive behaviors strengthens that claim. Evidence that some students use both adaptive and maladaptive skills is consistent with the claims of other researchers (Gotwals & Songer, 2010; Gotwals et al., 2012), who found that students in the “messy middle” of a learning progression often exhibit both higher-level skills and lower-level errors. This may help ACT to validate a learning progression framework or performance level descriptors that would meet the goal of placing students along a learning progression from the NRC Report *Designing Assessments for the Next Generation Science Standards* (National Research Council, 2014).

3.3.4.5.4 Engineering and Design Thinking (EDT)

Students were generally successful (87.5%) at answering the items that targeted the new EDT skills in Passage A. All who used the targeted skills in the EDT subdomains were able to answer the items correctly. It appears that, for at least the EDT skills tested by these items, students have been able to learn and apply these skills. In the follow-up survey, students were asked when they learned topics in engineering and design thinking, as shown in Figure 3.23.

Figure 3.23. Student Survey Responses About Learning Engineering



Over 70% of students say that they have learned about constraints, goals, and tradeoffs, which are key targeted skills in EDT in both state standards and the enhanced ACT science test. The overwhelming majority were able to apply these skills; however, nearly half of students do not consider themselves to have learned engineering or design thinking. This suggests that communication about the EDT change should encourage students to focus on the skills they are using rather than the language of engineering and design thinking as it is used in state standards. Conversely, to state stakeholders, we will want to communicate the standards language and that students are able to demonstrate mastery of these EDT skills. Further studies such as the National Curriculum Survey, posttest questionnaires, or focus groups with teachers or students should be used to verify that students are learning the skills that the enhanced ACT will target in EDT even if students think of this as a design challenge or application of science rather than engineering and design thinking.

3.3.4.5.5 BGK Versus Other Items

Items that require integrating background science content knowledge are given a secondary code of BGK. In response to previous peer reviews, alignments, and RFPs, the targeted percentage of BGK items on forms has been increased for the enhanced ACT. Because these items require students to integrate information from the passage as well as their content knowledge, the items potentially represent an additional dimension to the model of science college readiness. A recent analysis of BGK items conducted by the science team found that many items that appear to be of high quality based on the construct assessed and structure of the question would fail because of poor discrimination indices (biserial or point biserial) in pretesting. This stems from a smaller gap between the performance of high, middle, and low scorers on these items. Additionally, items that test topics low scorers are more likely to take in high school (Earth and space science, environmental science) may behave in the opposite way anticipated for a unidimensional construct because most honors or AP science students do not take these subjects in high school. As shown in Table 3.20, the correct percentage for students who attempted the BGK items was substantially lower than it was for transferable science skills (47.8% versus 70.9%). However, there was great variability in these items, so they will be discussed individually.

Item A5 EDT Unit

This item involved a combined scientific argument with justification and BGK. To answer correctly, students must use their understanding of the experimental design to determine that the installations are collecting solar energy. They must then identify that solar energy is a renewable resource. Of the students who attempted this item, 70.0% answered correctly. Some students selected the option “renewable resources do not produce harmful emissions.” Test developers expect that students may conflate a goal of using renewable energy and what renewable energy sources are; this makes for a foil that differentiates between students who understand what renewable energy is and what its desired effects are and students who do not.

Item B2 Biology Unit

This item brought to light important issues in question wording: “Based on the passage, dosin is composed of what subunits?” Many students spent substantial time looking for information that stated directly what dosin is made of. It was not clear that students needed to use both the passage and their science BGK to answer the question. A student in the follow-up interview stated that on the reading test, when items use “based on” language, you must find something explicitly in the text. So, he was expecting to find something in the text explaining what dosin was made of as one of the choices. The item required students to identify that the passage states that dosin is a protein and then apply the BGK that proteins are made of amino acid subunits. For items that do not link the science BGK to the passage using reasoning (similar to Item A5), it is suggested that we include “Based on the passage and your science knowledge…” as part of the item stem. Test prep should address the usage of “based on” and “according to” to help teachers and students to understand the specific tasks required.

This item also raises issues about plausible answer options for this kind of item. Of students who attempted this item, 63% used one or more maladaptive behaviors (mostly guessing) and answered correctly 54.7% of the time. Although students are likely exposed to all four of these terms in a required biology class, they are not treated equally. Amino acids are stressed as building blocks of proteins in multiple units, including biochemistry, transcription, and genetics units. Although students learn that starches are polymers of monosaccharides, the scientific name is rarely used and either simple sugars or examples like glucose are used in most textbooks, labs, and teaching materials. The focus of instruction on RNA, the other chemical mentioned in the passage, is on what it does—mRNA (messenger RNA) or tRNA (transfer RNA)—and not what it is made of. Further study is needed of items that have been pretested or were operational in the pool that use specific content knowledge like this without the scientific argumentation aspects of the other items used in this study. Further study of answer options that are nonfunctional should occur as well.

Item B7 Biology Unit

There are many topics that students struggle to understand that have been identified as misconceptions (Helm, 1980; Eaton et al., 1984), preconceptions (Novak, 1977), alternate conceptions (Driver & Easley, 1978), p-prims (DiSessa, 1993), and stepping-stone understandings (Harrison & Treagust, 1996), to name only a few. The difference between meiosis and mitosis is a good example (Ozcan et al., 2012), and understanding this difference is necessary to correctly answering this item. For this item, 59.1% of students engaged in both adaptive and maladaptive behaviors, meaning that they found relevant information in the text and then either drew the wrong conclusion or lacked the BGK to correctly answer the item. This indicates that the item was eliciting skills in both the transferable science practices and content domains. Expectedly, the item was very difficult (31.8% correct); however, the item would have difficulty meeting discrimination indices because the confusion between mitosis and meiosis was spread across all ability groups, which is similar to findings of other research on this misconception (Lewis et al., 2000). Topics such as these that contain common misconceptions frequently appear on concept inventories (Tsui & Treagust, 2007) and may be problematic to include based on test specifications because they will not follow the general pattern of items in which individuals in the high, middle, and low ability groups are progressively less likely to answer the item correctly.

3.3.4.5.6 Impacts and Changes to BGK Items

Items that require BGK in addition to transferable science skills will be classified in one of two ways. Items that use a claim-evidence-reasoning model of scientific argument (similar to A5 and B7) will not change as the answer choices indicate a need to link content knowledge to evidence and claim and should provide the context students need to activate appropriate BGK. Items that do not use a claim-evidence-reasoning argument model (similar to B2) will be worded "Based on the passage/Experiment 1 *and your science knowledge . . ."* in order to clearly indicate that students must use prior knowledge and guide them to activate relevant BGK in addition to the information in the passage.

3.3.4.5.7 Grade Level

As expected, students in Grade 10 appeared to be less prepared than students in either Grades 11 or 12 (refer to Table 3.21.)

Table 3.21. Science Performance by Grade Level

Grade	Average % correct	Used 1+ adaptive skill & correct	Used 1+ maladaptive skill & correct	Completely or mostly understood question	Understood question Likert (out of 4)
10	62.0%	78.9%	49.3%	69.1%	3.27
11	80.3%	97.0%	31.8%	86.3%	3.40
12	68.8%	91.7%	29.1%	83.3%	3.53

The results of this small sample were greatly skewed for the Grade 12 students by the two BGK items on the biology unit. On Questions B2 and B7, Grade 10 students used maladaptive skills including guessing on 60.0% and 88.9% of attempts, respectively. On item B7, several students in Grade 12 stated that they could not remember the difference between meiosis and mitosis and consequently guessed. For the seniors in an honors track in science, it had been three years since they had taken biology. Several of the seniors who had taken either the SAT or the ACT commented that they could not find what dosin was made of in the passage. One specifically stated that on reading tests for both the SAT and ACT, he knew that “based on the passage” items should require finding information in the passage.

Table 3.22. Passage and Topic Effects by Grade

	Passage A		Passage B	
Grade	Percent Correct	Completely or mostly understood question (%)	Percent Correct	Mostly or completely understood
10	78.0	66.2	56.3	72.1
11	80.8	100	80.0	92.5
12	78.0	83.0	60.9	85.5

The lower rate of “mostly” or “completely understand the question” responses for Grade 10 students in Table 3.22 may suggest that there is a large learning jump for many students in science moving from Grade 10 to Grade 11. This is likely because many students take chemistry, which integrates more data analysis and math than most high school biology courses, during their junior year. This should continue to be monitored for potential impacts on projections from the PreACT suite into the ACT, and it should likely be researched further using the National Curriculum Survey or some other tool. It also suggests that students who take more of the suggested science core classes (3 credits including biology as well as either chemistry or physics) will likely be better prepared for the enhanced ACT science test. Again, further research relating course-taking patterns to performance on the enhanced ACT test will be needed. It is valuable to note that even for tenth grade, the student either completely or mostly understood the question almost seventy percent of the time, so the questions were clear to most students. On Passage A, which primarily focused on understanding the experimental design and interpreting the data, students across grades performed strongly on items that required using only one graphic.

3.3.4.5.8 Calculators

Unlike the legacy ACT science directions, the instructions did not indicate whether students were allowed or forbidden to use a calculator on the science test. There were two items for which a calculator could be useful, A2 and A4. For A2, students were required to find two data points in the table and find the difference ($44.3 - 36.8 = 7.5$). Of the students who used the calculator for this item, 45.8% answered it correctly, as did all the students who did not use a calculator. On A4, students needed to find two data points (approximately \$325,000 and \$100,000) and in addition to finding the difference, justify whether one experimental site had a higher economic benefit and by how much. On this item, only two students used a calculator.

On item A1, three students used the calculator to pursue a maladaptive path, and all three were incorrect on the item. So, having the calculator may lead some students to pursue paths that will waste time and lead to wrong answers, paths they might not have pursued if they had had to make the calculations by hand. One student used the calculator as a straight edge to trace from the graph back to the *y*-axis to find the appropriate value.

The current boundary statements for items associated with CCRS IOD-402 (and similar CCRS skills) require that students should be able to complete the math easily without a calculator. Most state standards include some form of mathematical reasoning or data analysis that is consistent with our current approach. The external panel concluded that for the items that were on the forms that they reviewed, the math could be done without calculators. Although some students may be helped by using the calculator on a few items, allowing calculators may hinder some students who try maladaptive paths with a calculator that they would not have pursued had they been calculating by hand. Further study on this is needed.

3.3.4.5.9 Word Count and Reading of Science Passages

In general, students who read the science passages before trying to answer questions were able to answer more questions correctly in less time. It was more efficient to make sense of the passage, frequently with underlining or other markup of the passage, and then attempt to answer the questions. Students did not read every data point when first interacting with a table or graph but rather spent time looking at labels and axes, which was similar to the science eye-tracking data from the 2017 cog labs. These findings led to an adjustment of word count rules to more accurately reflect the time needed for students using adaptive skills to make sense of the scientific phenomena being presented. This also allows for adding more context to the passage, which was recommended by the expert panel, to help students activate prior knowledge and facilitate more engagement with the scientific phenomena.

Further study of the use of tools for online testing may be needed, as students frequently drew lines on graphics, circled points in tables, and underlined key information in the passage and then referred to these marks as they answered more than one item, especially on the Conflicting Viewpoints passage. Since the current testing interface does not retain highlighting of the passage text when a student moves on to the next item, research on how this affects students who are using adaptive behaviors will be needed.

The planned changes for the enhanced ACT science test were supported by the cog lab findings. The inclusion of EDT and additional BGK items will allow the enhanced ACT science test to better align with most state standards.

3.3.4.6 Post-Study Survey and Interview Findings

After the think-aloud protocol, students completed a brief survey. The survey measured perceptions of the ACT questions compared to the student's abilities, when the student learned various skills, whether the test would be easier or harder on paper or online, and whether the student had already taken the ACT and SAT tests.

For each subject test, participants were asked whether they understood what the question was asking them to do, even if they did not know how to correctly answer the question (refer to Table 3.23). They were also asked whether they were able to show what they know and can do on each section of the test (refer to Table 3.24). For all subjects, more than 85% of students responded that they understood what the question was asking them to do. Over 70% of students agreed that they were able to show what they knew and could do on each section of the test (Table 3.24).

Table 3.23. Student Ability to Understand What the Question Was Asking by Test Section

I was able to understand what the question was asking me to do, even if I did not know how to answer the specific question on the ____ test.

Section	% Strongly Disagree	% Disagree	% Agree	% Strongly Agree
English	0.00	0.00	44.44	55.56
Math	0.00	14.81	44.44	40.74
Reading	0.00	3.70	51.85	44.44
Science	3.70	7.41	48.15	40.74

Table 3.24. Students' Rating of Their Ability to Show What They Know and Can Do by Test Section

I was able to show what I know and can do in the subject on the ____ test.

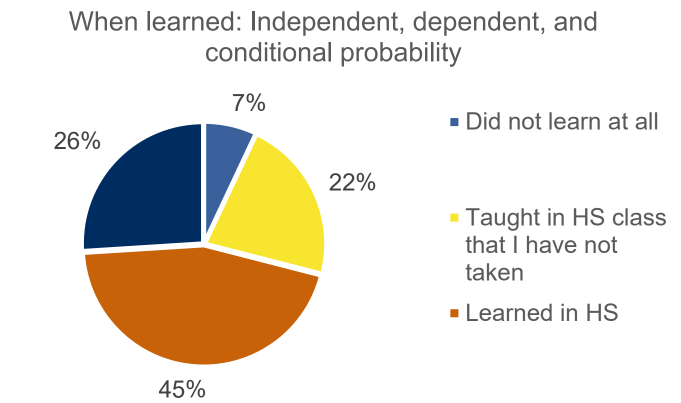
Section	% Strongly Disagree	% Disagree	% Agree	% Strongly Agree
English	0.00	11.11	25.93	62.96
Math	3.70	25.93	33.33	37.04
Reading	0.00	11.11	29.63	59.26
Science	0.00	25.93	29.63	44.44

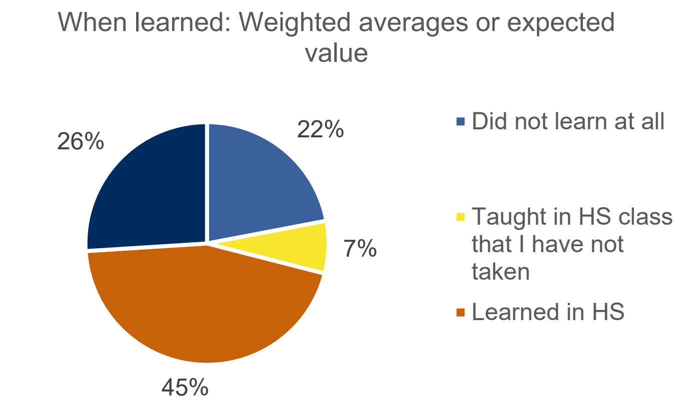
The lower, although still acceptable, results of students' beliefs that they were able to show what they know and can do in math and science were impacted by grade level. The sample of items tended to include a higher percentage of advanced topics compared to a full test battery, which impacted students' perceptions, particularly those of students in Grade 10.

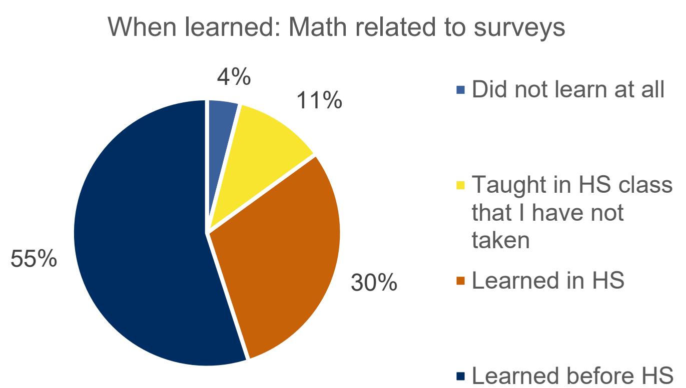
Students were asked, "If you were to take a test with items like these on a computer would you think the test was . . ." with the following results: 44% reported that the test would be easier on paper, 37% that it would be about the same on paper and on computer, and 19% that it would be easier on computer. The most common justifications for the test being easier on paper were that it was easier to annotate reading and science passages and do work on math problems. The most common justification for "about the same" was that the questions would be the same and use the same thought processes. The justifications for the test being easier on computer did not show common themes. These results, as well as the screening survey results, support the decision to offer the ACT in both paper and online formats. These data contrast slightly with the screening data about testing preference after students have seen the types of passages and items that occur on the ACT as opposed to those used in state testing.

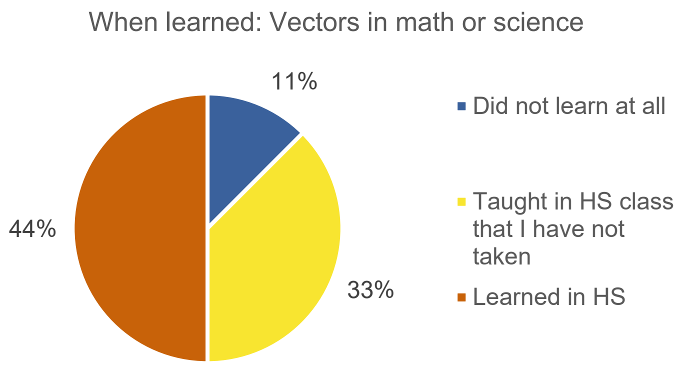
Figures 3.24 through 3.26 show when, if at all, students reported learning specific skills in math, argumentation with evidence, and aspects of engineering and design thinking.

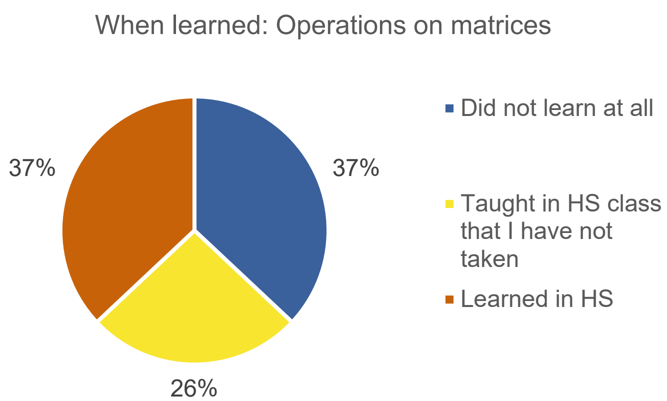
Figure 3.24. When Students Learned Topics in Math





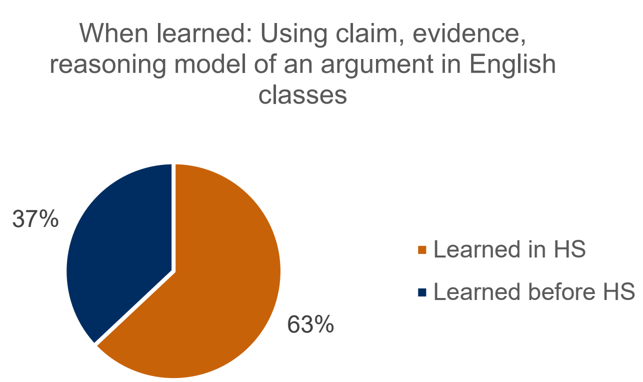


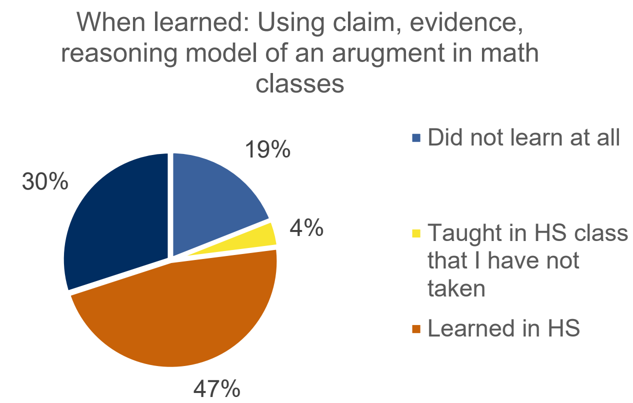


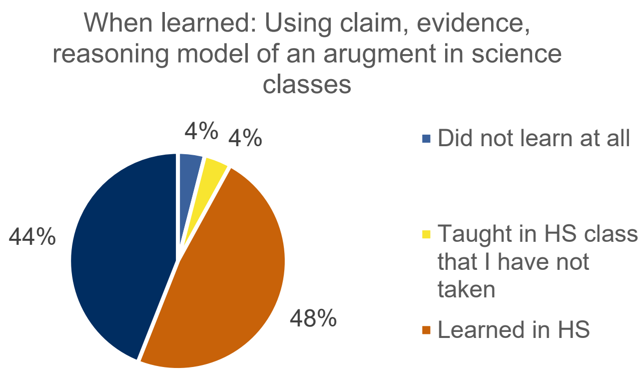


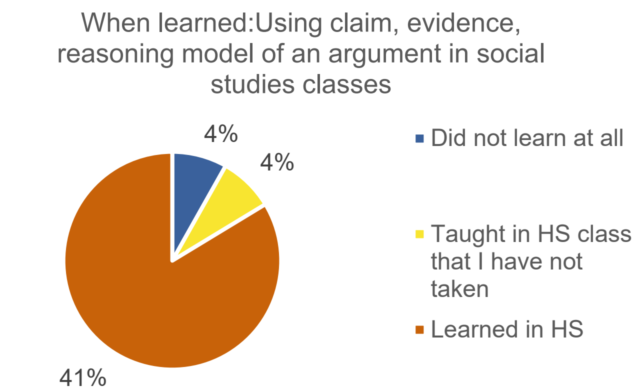
Although matrices and vectors are represented in most common math standards, it appears that few students are acquiring knowledge and skills on these topics. This supports the decision to slightly decrease the percentage of advanced topics, which include these, on the enhanced ACT math test. The high percentage of students who have learned skills in the statistics domain provides additional evidence to support the decision to include more statistics and probability items on the enhanced ACT math tests.

Figure 3.25. When Students Learned CER Model of Argument Across School Subjects



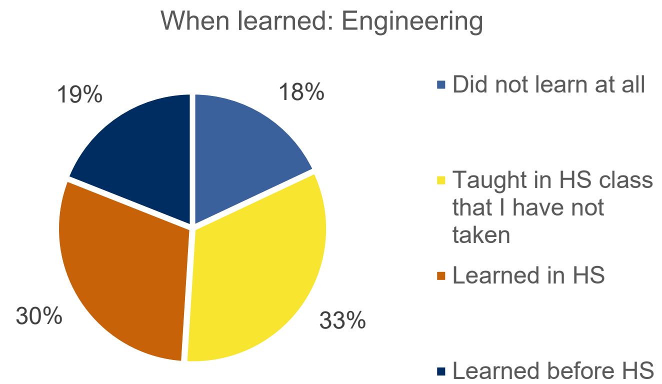


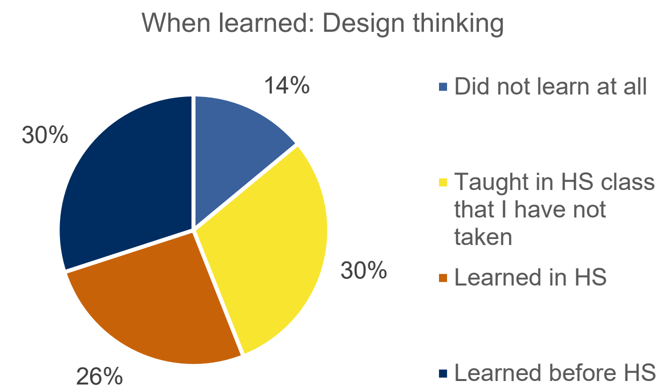


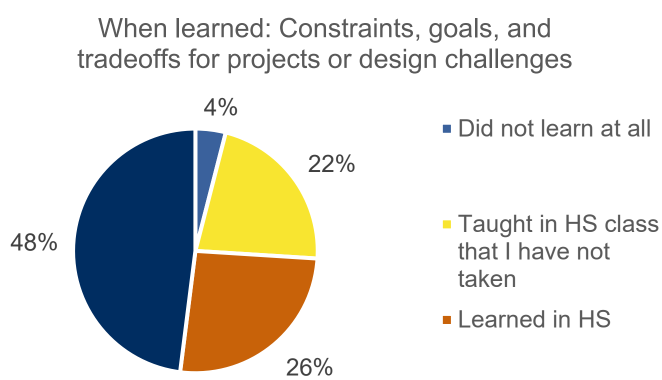


The prevalence of learning the claim-evidence-reasoning model of argument across all domains strengthens the need for including argumentative essays and items using this model as a key skill that is taught in high school and is necessary for college and career readiness.

Figure 3.26. When Students Learned About Topics in Engineering and Design Thinking







In Figure 3.26, over 70% of students stated that they learned constraints, goals, and tradeoffs, which are key targeted skills in EDT in both state standards and the enhanced ACT science test. The overwhelming majority of students in this study were able to apply these skills; however, nearly half of students do not think they have learned engineering or design thinking. Communication about the EDT change should encourage students to focus on the skills they are using rather than the language of engineering and design thinking as it is used in state standards. Conversely, to state stakeholders, ACT should communicate that students are able to demonstrate the mastery of the EDT skills in the standards.

3.3.4.6.1 Interview

As part of a brief interview after completing the survey, participants were asked whether anything was surprising or different than they expected. The responses were coded for up to three themes per participant. Additionally, participants were asked whether each test was easier or harder than expected, and for the reading and science sections, which passage was more difficult.

For English, the top three themes were that nothing was surprising or different (36.4%), that there were fewer grammar and punctuation questions than expected (15.2%), and that the section uses a side-by-side format with questions next to the corresponding sentences (12.1%). Almost half (48.2%) of participants reported that the English section was easier than expected, 25.9% reported that it was harder than expected, and 25.9% reported that it was about what they had expected. The comments about the side-by-side format may be a result of the difference in format with state testing in Grades 3 through 8.

For math, the top three themes were that there was some content on the test that students did not know how to do (25.7%), that there was nothing surprising or different (11.4%), and that the questions were confusing or unclear (11.4%). Two-thirds (66.7%) of participants reported that the math section was harder than they expected, 22.2% reported that it was easier than expected, and 11.1% reporting that it was what they had expected. As noted before, a large percentage of tenth graders and the combination of advanced topics and no geometry items likely contributed to these observations.

For reading, the top three themes were that there was nothing surprising or different (29.0%), that it was long or sometimes hard to read (19.4%), and that there were graphs on a reading test (12.9%). Students were split over the difficulty of this test, but the largest percentage (29.6%) of participants reported that the reading test was about what they had expected. Students were split on which passage they viewed as more difficult, with 44.4% saying the paired passages were more difficult, 29.6% saying the VQI was more difficult, and 22.2% reporting that both passages were equally difficult.

For science, the top three themes were that the test was like a reading test (11.8%), that nothing was surprising (11.8%), and that the test was difficult to process and required rereading the passage (11.8%). Thirty-seven percent (37.0%) of participants reported that the science test was harder than expected, 29.6% reported that it was easier than expected, and 33.4% said it was similar to what they expected. Students were evenly split on which passage was more difficult, with 25.9% of students choosing the engineering passage, 25.9% choosing the conflicting viewpoints passage, and 3.7% reported that both were hard.

3.4 Conclusions

Across all parts of the ACT, students who use one or more of the targeted skills answer correctly much more frequently than do students who use maladaptive skills. As Table 3.25 shows, students who use only targeted skills answered items correctly over 80% of the time in both the ELA and STEM domains.

Table 3.25. Use of Targeted Skills Across Domains

Domain	Used one or more targeted skills correct	Used maladaptive skill correct	Difference between targeted skill and maladaptive skills	Used only adaptive percent correct	Difference between only adaptive and maladaptive skills	Overall average correct
English	70.3%	13.2%	57.1%	88.0%	74.8%	67.5%
Math	73.2%	19.8%	53.4%	84.0%	64.2%	54.8%
Reading	70.4%	18.6%	51.8%	81.0%	62.3%	68.0%
Science	73.4%	21.8%	51.7%	94.6%	72.8%	67.7%
ELA	70.4%	15.9%	54.4%	84.5%	68.6%	67.8%
STEM	73.3%	20.8%	52.6%	89.3%	68.5%	61.2%

Overall, the findings did not reveal any significant concerns and support the proposed changes for the enhanced ACT. This study is limited in that it had a larger percentage of tenth graders than was desired (33.7%, n = 10), so many participants were unable to provide comparisons to prior ACT or SAT test experiences. Additionally, although the sample was diverse in terms of race, ethnicity, gender, and accommodation status, it did not purposely include students with specific accommodations. It is recommended that ACT conduct additional focus groups, including one specifically for students with accommodations.

Chapter 4: Additional Psychometric and Validity Evidence

4.1 Introduction

In this chapter, we summarize additional psychometric and validity evidence supporting interpretations of scores from the enhanced ACT® test. First, we describe the June 2024 Linking Study that was used to evaluate construct invariance and to link enhanced ACT section test scores to legacy ACT section test scores. Next, we describe the October 2024 mode comparability study that we used to evaluate whether scores between paper and online administrations are equivalent and to obtain interchangeable scores across modes of testing. We then investigate the concurrent validity evidence examining relationships of enhanced ACT test scores with other measures of academic performance, as well as predictive validity evidence examining relationships of simulated enhanced ACT test scores with college outcomes. Finally, we summarize research on Composite score comparability, examining differences between Composite score interpretations for the legacy ACT and enhanced ACT, with and without the inclusion of science in the Composite score.

Much of the evidence presented in this chapter is drawn from other research reports, including the Enhanced ACT Linking Study Report (Li et al., 2025) and Initial Evidence Supporting Interpretations of Scores from the Enhanced ACT Test (Allen & Cruce, 2025). We argue that the evidence presented in this chapter supports the interpretation of scores from the enhanced ACT as measures of high school academic achievement and college readiness and that scores from the enhanced ACT may be used interchangeably with scores from the legacy ACT for nearly all purposes. We conclude that scores from the enhanced ACT can be used for informing college admissions decisions, awarding college scholarships, placing students into programs and courses, identifying students in need of academic support, and measuring academic achievement at the school and district level for accountability systems.

4.2 June 2024 Linking Study

The June 2024 linking study was conducted in an operational testing environment where participants received college-reportable scores. Data were collected during the online administration using a random groups design, where one full-length form based on the legacy ACT test blueprint (hereafter referred to as Legacy Form) and two forms based on the enhanced ACT test blueprint (hereafter referred to as Enhanced Form 1 and Enhanced Form 2) were spiraled among students within test centers, including students who tested with 1.5 extended time accommodations. Over 180 test centers across the country participated in the study. The final cleaned dataset included 6,882 students: 2,298 for the Legacy Form, 2,280 for Enhanced Form 1, and 2,304 for Enhanced Form 2. Table 4.1 provides a summary of the linking study sample.

Table 4.1. June 2024 Linking Study Sample Summary

Variable		Enhanced ACT	Legacy ACT
Number of students		4,584	2,298
Gender (%)	Another gender	0.2	0.1
	Female	54.1	54.7
	Male	44.8	44.6
	Missing	0.9	0.6
Race/ethnicity (%)	Asian	4.9	4.6
	Black	14.4	14.8
	Hispanic/Latino	13.6	12.9
	Native American	0.7	0.7
	Native Hawaiian/OPI	0.1	0.1
	Two or more races	4.8	5.4
	White	57.9	58.3
	Missing	3.7	3.3
Grade level (%)	10	12.0	13.2
	11	68.8	69.6
	12	12.1	10.5
	Other	7.2	6.6
ACT score mean (*SD*)	Composite	21.6 (5.5)	21.6 (5.5)
	English	20.8 (6.4)	20.7 (6.3)
	Math	20.9 (5.6)	20.9 (5.4)
	Reading	22.4 (6.7)	22.4 (6.6)
	Science	21.8 (5.8)	21.8 (5.7)
Other data (%)	Took post-test survey	91.6	92.9
	Took prior ACT test	53.6	54.2

Note: OPI = other Pacific Islander; *SD* = standard deviation.

The goal of the linking study was to evaluate construct invariance and to link section scores from the enhanced ACT to section scores from the legacy ACT. We examined the impact of the enhancements on the construct of measurement and student performance before we conducted statistical linking. Subsequently, we examined the psychometric properties of the enhanced forms and compared the results with those of the legacy form.

4.2.1 Construct Equivalence

One of the requirements for an effective linking study is a high degree of similarity in test features, including constructs, populations, inferences, and measurement characteristics (Kolen & Brennan, 2014). The enhanced ACT was designed to yield the same inferences as the legacy ACT by measuring the same constructs. The test specifications for the enhanced ACT are very similar to those of the legacy ACT but have different lengths and/or timing across the four test sections.

To evaluate construct equivalence, confirmatory factor models were estimated for the three forms used in the linking study. Fit indices and factor loadings of the enhanced forms were similar to those of the legacy form for each model, reflecting the similarities in the internal structure and the test blueprints, supporting both unidimensionality and the structures for scoring and reporting, and indicating consistency in the construct of measurement.

Table 4.2. Model Fit Statistics and Average Factor Loadings

Section	Form	χ^2	*df*	*p*-value of χ^2 test	RMSEA	CFI	Average standardized loading
English	Legacy Form	7,008.428	2,700	0.00	0.03	0.95	0.55
	Enhanced Form 1	1,749.684	740	0.00	0.02	0.97	0.52
	Enhanced Form 2	1,710.667	740	0.00	0.02	0.97	0.51
Math	Legacy Form	4,150.435	1,710	0.00	0.03	0.97	0.55
	Enhanced Form 1	1,739.304	779	0.00	0.02	0.97	0.50
	Enhanced Form 2	2,197.379	779	0.00	0.03	0.95	0.49
Reading	Legacy Form	1,771.695	740	0.00	0.03	0.97	0.55
	Enhanced Form 1	594.894	324	0.00	0.02	0.99	0.52
	Enhanced Form 2	514.656	324	0.00	0.02	0.99	0.51
Science	Legacy Form	2,096.352	740	0.00	0.03	0.97	0.56
	Enhanced Form 1	1,528.193	527	0.00	0.03	0.95	0.50
	Enhanced Form 2	1,193.889	527	0.00	0.02	0.97	0.51

Note: *df* = degrees of freedom; RMSEA = root mean square error of approximation; CFI = comparative fit index. All of the chi-square statistics are significant (*p*-value = .00).

4.2.2 Evaluation of Psychometric Properties

We conducted analyses to investigate the impact of test changes on measurement precision and test score reliability, as well as to examine the practical impact on score interpretations and the decision consistency of individual scores.

4.2.2.1 Conditional Standard Error of Measurement, Standard Error of Measurement, and Reliability

In 1989, the ACT test was scaled to have approximately equal conditional standard errors of measurement (CSEMs) along the score scale. The extent that the CSEM is constant across scales varies among legacy forms because of the gradual changes of the test specifications and form difficulty over time. However, the claim that the ACT multiple-choice section tests have a roughly constant CSEM of about 2 still holds for legacy ACT forms (ACT, 2024).

The scale score CSEMs and standard errors of measurement (SEMs) of the enhanced ACT forms were examined to evaluate whether measurement errors were still roughly equal across scores and whether the magnitude of the error exceeded those reported in the ACT score report (i.e., 2 for each multiple-choice section test and 1 for the Composite score). Scale score CSEMs, SEMs, and reliability were estimated based on a four-parameter beta compound binomial model as described in Kolen et al. (1992). Figure 4.1 presents the CSEM plots for each section test. The CSEMs for Enhanced Forms 1 and 2 follow the same pattern as the Legacy Form for all four sections, with the measurement precision varied across the scale scores.

Despite Enhanced Forms 1 and 2 having a similar pattern across the scale scores as the Legacy Form, both enhanced forms have increased CSEMs because of the reduction in test length, and they show greater deviation from the reported SEM (2) than the Legacy Form does.

Figure 4.1. Conditional Standard Errors of Measurement for Each Section Test

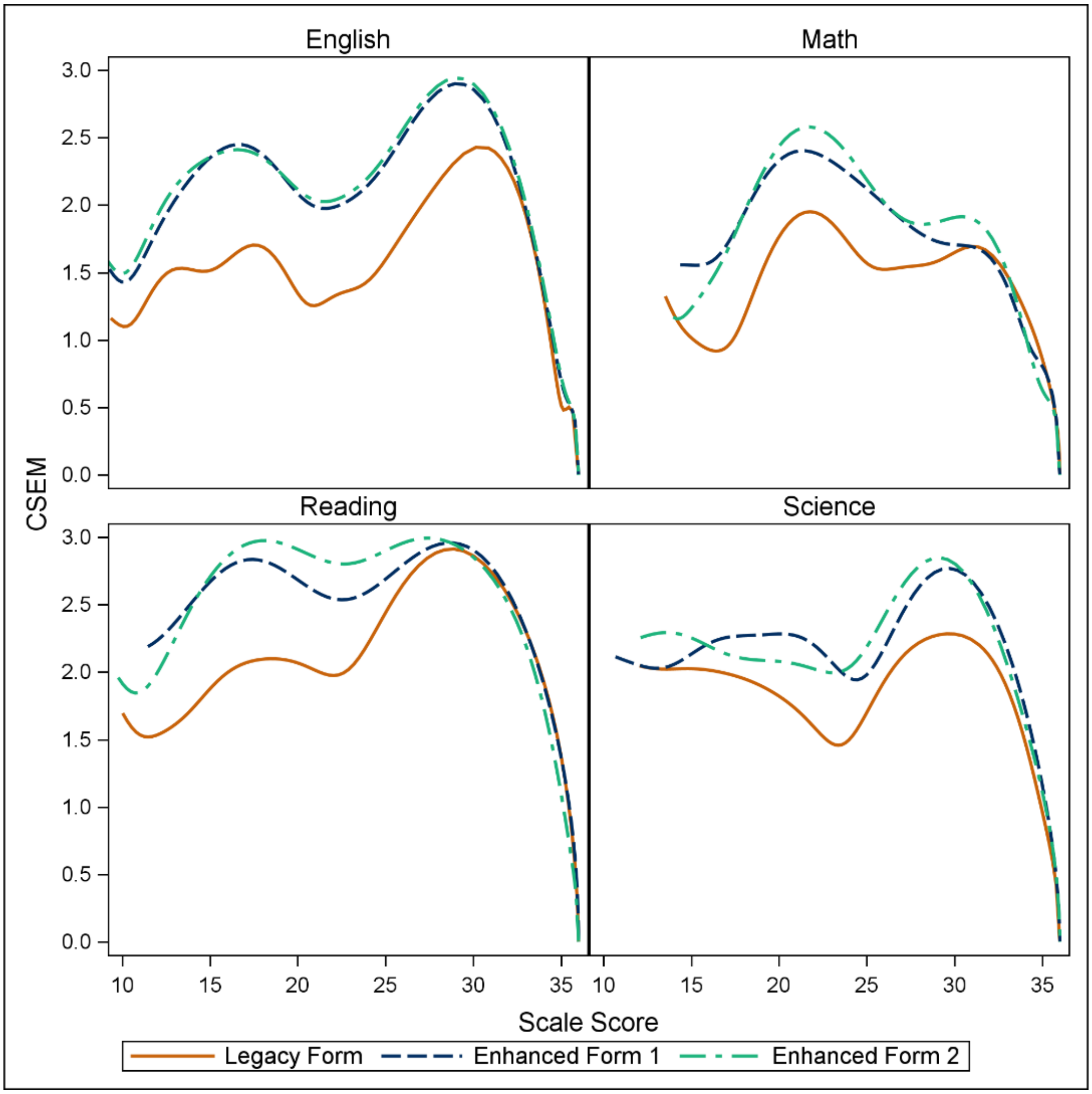


Table 4.3 shows the SEMs and reliability estimates. We use the abbreviations EMRS (English, math, reading, science) to refer to Composite scores with science and EMR (English, math, reading) to refer to Composite scores without science. Because of the reduction in test length, the SEMs increased, and the scale score reliability estimates of the enhanced ACT forms decreased. The reliability estimates for the enhanced forms reached an acceptable level (i.e., >.80). However, the SEMs are large in some instances, most notably for the reading scale scores that reached 2.66 for Enhanced Form 1 and 2.75 for Enhanced Form 2 and for the EMR score that reached 1.32 and 1.35 for Enhanced Forms 1 and 2, respectively.

Table 4.3. Reliability Estimates and Standard Errors of Measurement for Each Section Test and Composite Scores

Score	Reliability			SEM		
	Legacy Form	Enhanced Form 1	Enhanced Form 2	Legacy Form	Enhanced Form 1	Enhanced Form 2
English	0.94	0.88	0.88	1.58	2.23	2.26
Math	0.93	0.88	0.88	1.43	1.92	1.94
Reading	0.89	0.84	0.83	2.22	2.66	2.75
Science	0.89	0.85	0.85	1.88	2.22	2.25
STEM	0.95	0.93	0.92	1.18	1.47	1.49
EMR	0.97	0.94	0.94	1.03	1.32	1.35
EMRS	0.97	0.96	0.96	0.90	1.14	1.16

***Note*:** SEM = standard error of measurement; STEM is a combination of students' math and science scores; EMR = English, math, and reading; EMRS = English, math, reading, and science.

4.2.2.2 Effective Weights for Composite Scores

Effective weights, which measure the proportion of the variability of the Composite score that can be attributed to each section test, were estimated for the EMRS and EMR scores. The effective weights are identical or very similar between the Legacy Form and Enhanced Forms 1 and 2.

4.2.2.3 Classification Accuracy of ACT College Readiness Benchmarks and WorkKeys NCRC Predictions

Classification consistency refers to how often students are placed in the same category when a test is repeated. Since tests are usually given only once, classification consistency is typically estimated from a single test, assuming certain distributions in errors and true scores. Using the method described by Livingston and Lewis (1995), we estimated the classification consistency of the ACT College Readiness Benchmarks (18 for English, 22 for math, 22 for reading, and 23 for science). The classification consistency rates are fairly high, with the rates for the two enhanced forms ranging from 0.83 to 0.89, though slightly lower than those of the Legacy Form, which range from 0.87 to 0.92.

Using the same methodology, we also estimated the classification consistency of the Progress Toward the ACT® WorkKeys® National Career Readiness Certificate® (NCRC®) indicators. This indicator classifies Composite scores into five categories corresponding to the five WorkKeys NCRC levels (below bronze, bronze, silver, gold, and platinum). The classification consistency indices were 0.77 for EMR and 0.80 for EMRS for the Legacy Form, 0.71 for EMR and 0.75 for EMRS for Enhanced Form 1, and 0.71 for EMR and 0.74 for EMRS for Enhanced Form 2.

4.2.3 Linking Methodology

The item response theory (IRT) true score equating method was used to link scores from Enhanced Forms 1 and 2 to the Legacy Form. First, items on all three forms were calibrated.

Then, the item parameters were transformed to be on the same scale as the legacy ACT item pool, using the Stocking and Lord (1983) procedure to obtain scale transformation intercepts and slopes based on parameters on the Legacy Form. Finally, IRT true score linking was conducted to obtain raw to scale score conversions for the enhanced ACT forms.

Figure 4.2 shows the test characteristic curves (TCCs) of the three study forms for each section test. The solid line denotes the Legacy Form, and the dashed lines denote the two enhanced forms. The *x*-axis represents the theta score, and the *y*-axis represents the proportion correct. Although the enhanced forms tend to be slightly more difficult than the Legacy Form, the differences are small. This observation is consistent with the proportion correct raw score distributions, depicted by Figure 4.3.

Figure 4.2. Test Characteristic Curves

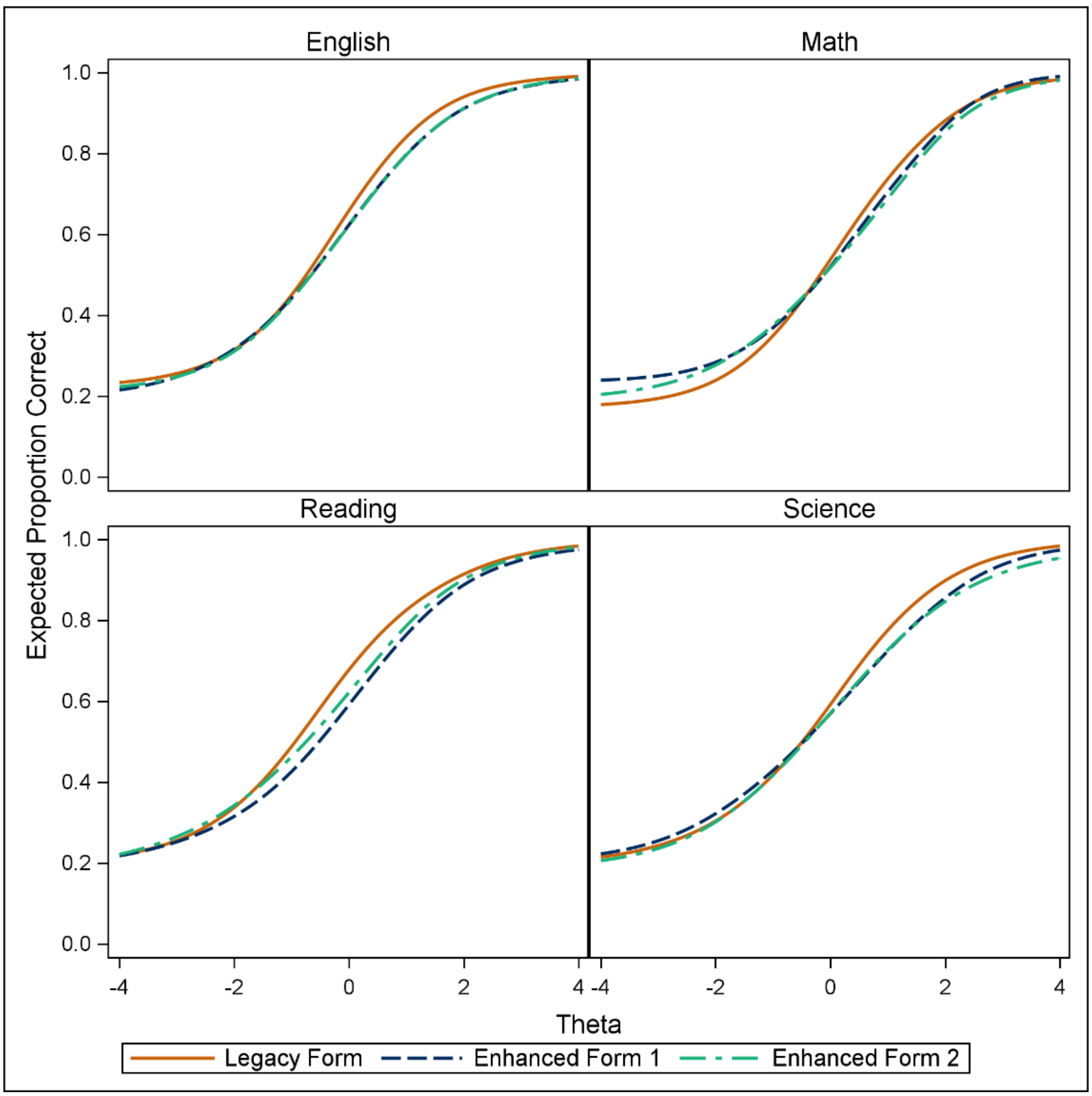


Figure 4.3. Relative Cumulative Distributions of Proportion Correct Scores

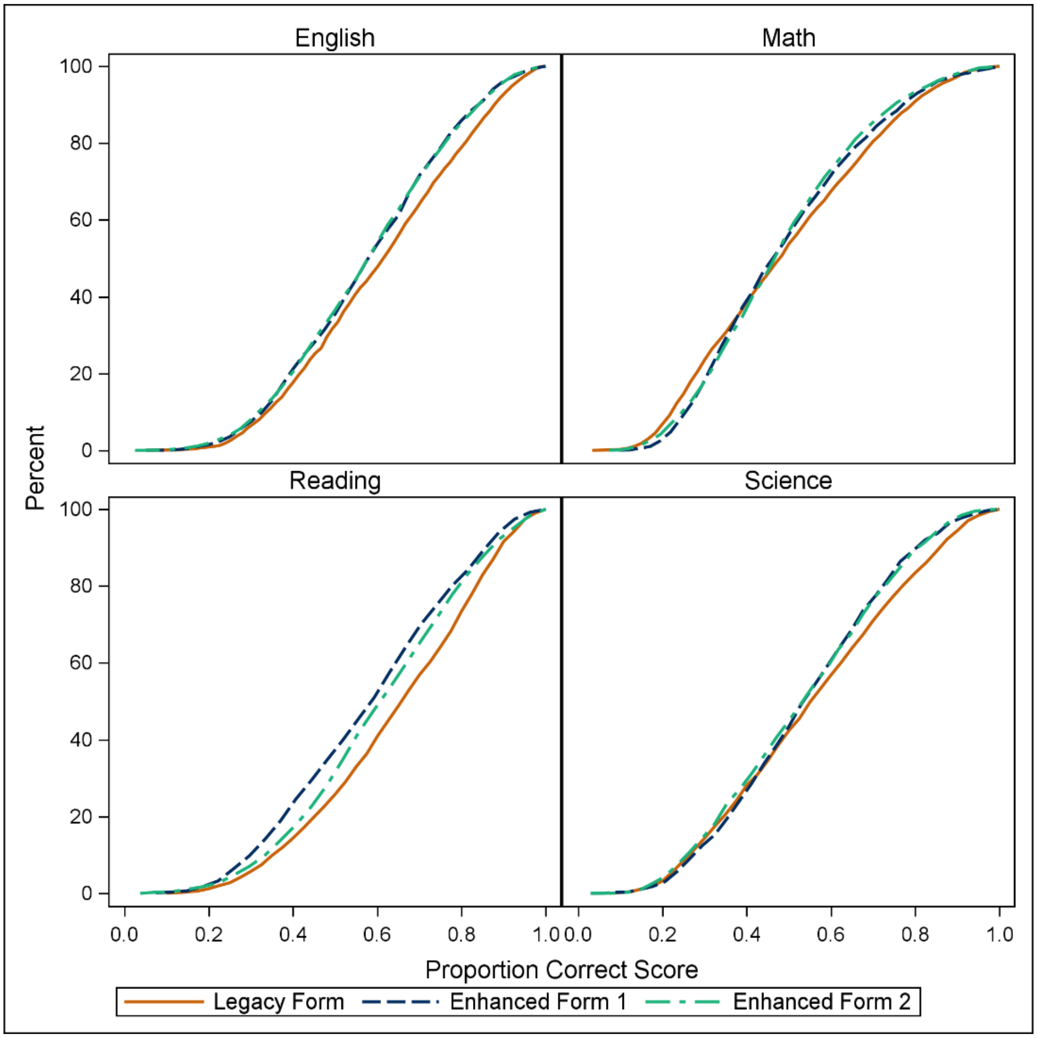
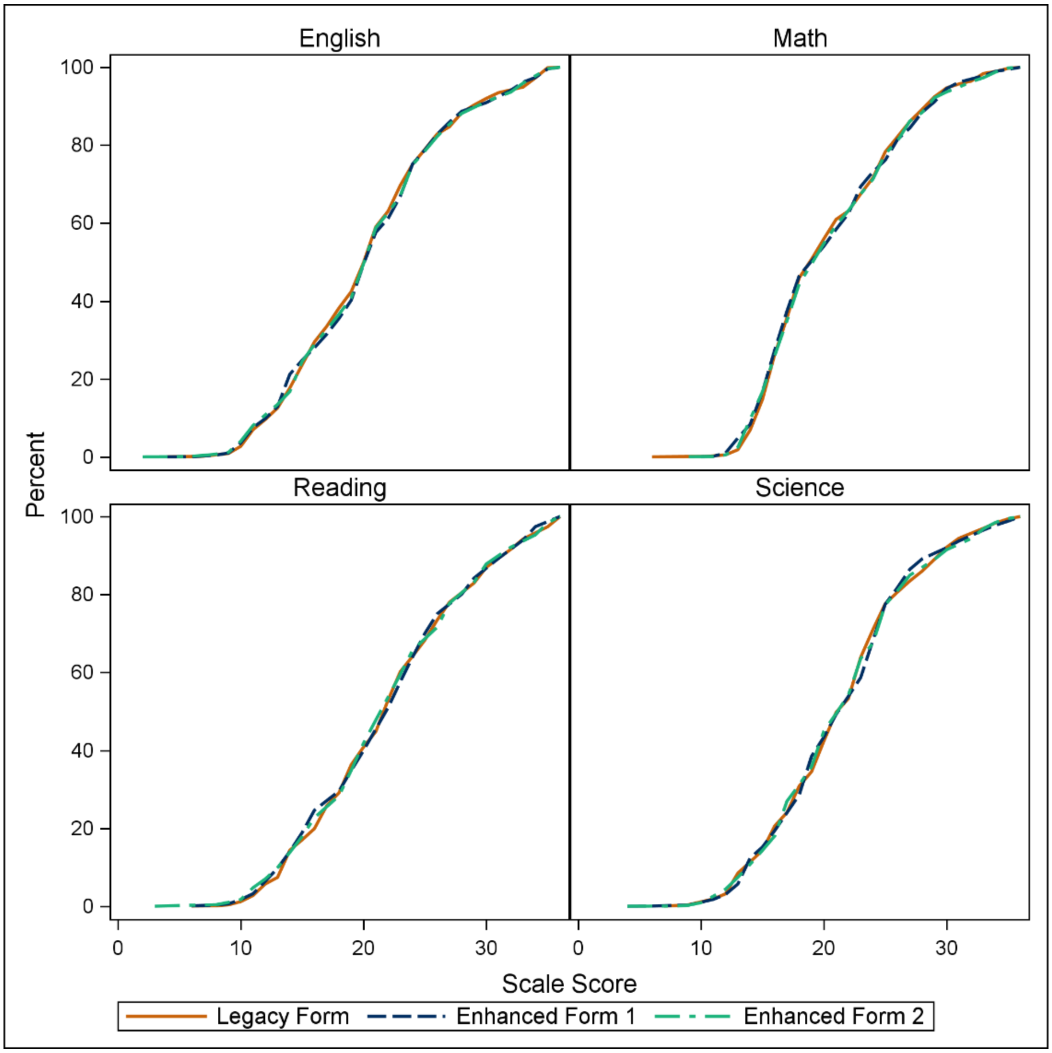


Figure 4.4 presents the relative cumulative scale score distributions after equating. As expected, the scale score descriptive statistics (refer to means and standard deviations in Table 4.1) and distributions were very similar after equating.

Figure 4.4. Relative Cumulative Distributions of Scale Scores



To further evaluate the results of the June 2024 linking study, we examined the population invariance property of Enhanced Forms 1 and 2. Population invariance refers to the degree that linking or equating results are consistent across different groups of examinees. The evaluation was conducted by comparing conversion tables obtained by using different subgroups in the equating sample with conversion tables using the total equating sample. Equipercentile linking with post-smoothing was used for this evaluation. Conversion tables from each subgroup were compared with those from the total group. The weighted root mean squared difference (WRMSD) statistic was calculated for each comparison to summarize the conversion table differences. Similar analyses were also conducted for 11 recently equated legacy ACT forms to serve as a baseline. All but three of the WRMSD statistics of the enhanced ACT forms fell within the ranges observed for the legacy ACT forms, indicating a similar extent of population invariance for the linking between the Enhanced and Legacy Forms as that of the equating of legacy ACT test forms. More details on the evaluation of population invariance is provided in the June 2024 linking study report (Li et al., 2025). These findings provided evidence for the robustness of the linking results.

In summary, the enhanced ACT measures the same constructs and reports the section test scores on the same scales as the legacy ACT. While the reduced test length led to expected decreases in measurement precision, the impact on score interpretation was minimal. Because section test scores from the enhanced ACT are linked to section test scores from the legacy ACT using the IRT true score equating method, with evidence for the robustness of the linking results, scores on the enhanced ACT and the legacy ACT can be used interchangeably. The enhanced ACT therefore supports continued interpretations of ACT section scores that were established from the legacy ACT, including interpretations of the ACT College Readiness Benchmarks.

4.3 October 2024 Mode Comparability Study

The October 2024 mode comparability study was conducted in an operational testing environment where participants received college-reportable scores. Data were collected during the October national online administration and October state and district (school-day) testing, using a random groups design. Study participants from selected national test centers took an enhanced form of the ACT test, with students randomly assigned to either paper testing or online testing via the UTD test delivery software. Study participants from selected schools in participating districts took the same enhanced form, with students randomly assigned to either paper testing or online testing via the TestNav test delivery software. Testing accommodations were provided, but students who tested with accommodations that require a specific test mode were not included in the randomized mode assignments. The final cleaned dataset included 4,791 students. Table 4.4 summarizes the study sample.

Table 4.4. October 2024 Mode Comparability Study Sample Summary

Student Group		School-day testing		Saturday testing	
		Online % (TestNav)	Paper %	Online % (UTD)	Paper %
Race/ethnicity (%)	Black/African American	15.9	14.5	19.7	20.1
	American Indian/Alaska Native	0.2	0.6	0.3	0.3
	White	54.7	52.5	50.2	48.2
	Hispanic/Latino	18.7	18.7	18.0	20.2
	Asian	2.0	2.0	4.9	4.3
	Native Hawaiian/Other Pacific Islander	0.1	0.3	0.3	0.3
	Two or more races (%)	5.0	5.0	4.9	4.8
	Prefer not to respond (%)	2.1	1.4	1.8	2.0
Gender (%)	Male	46.4	47.1	40.2	39.3
	Female	50.9	50.3	57.7	58.3
	Another gender	0.2	0.5	<0.01	0.1
	Prefer not to respond	1.7	1.4	2.1	2.3

Analyses were conducted to evaluate the potential impact of test delivery platform (UTD or TestNav) on examinee experience, item performance, and score comparability. The results of these analyses did not reveal any significant evidence of platform effects. Therefore, data from the school-day testing and Saturday testing were combined for the mode comparability study.

4.3.1 Mode Comparability Analyses

The mode comparability analyses were designed to evaluate whether section test scores of the enhanced ACT from paper and online administrations are equivalent and to obtain interchangeable scores across modes for score reporting. Mode analyses were conducted at two levels for the section tests to examine score comparability: score equivalency and construct equivalency. Score equivalency indicates that observed score distributions from the online and paper testing are very similar for the two randomly equivalent groups. Construct equivalency indicates that online and paper testing are measuring the same underlying abilities or attributes. Because the items were the same for online and paper testing, if score equivalency holds, then construct equivalency is partly supported. Additional comparisons to determine whether online and paper testing measure the same construct were conducted to ensure construct equivalency. Evaluation of score and construct equivalency was conducted at the test level and the item level, and analyses involved comparing the similarity of test score distributions and consistency in the construct measurement for the English, math, reading, and science section tests across modes, examining the similarity of their item score distributions between online and paper testing, comparing factor analysis results across modes and examining differential item functioning (DIF) between paper and online scores.

4.3.1.1 Test-Level Comparison

The similarity of test score distributions were compared for the online and paper testing groups. This included comparisons of means, standard deviations, and relative cumulative frequency distributions for the English, math, reading, and science section tests of the enhanced ACT. Obtaining interchangeable scores across modes for score reporting is one of the goals of the mode comparability study. Adjustments should be made if scores do not appear equivalent across online and paper testing so that scale score equivalency could be achieved. Results in this section provide a look at the unadjusted results.

Scale scores were compared across modes by applying the online version conversions obtained from the June 2024 linking study to both online and paper groups in the October 2024 mode comparability study. Raw scores were also compared. Figures 4.5 and 4.6 depict the raw score and scale score mean differences across modes for each section test. Table 4.5 presents the scale score mean differences, effect sizes, and *t*-tests of mean differences across modes for each section test. Results showed that the mean scores tended to be higher for online testing than paper testing for all section tests, with the largest differences observed in reading and the smallest differences in science. The mean differences were all statistically significant at $\alpha = .01$ for all section tests except science.

Figure 4.5. Raw Score Mean Comparison

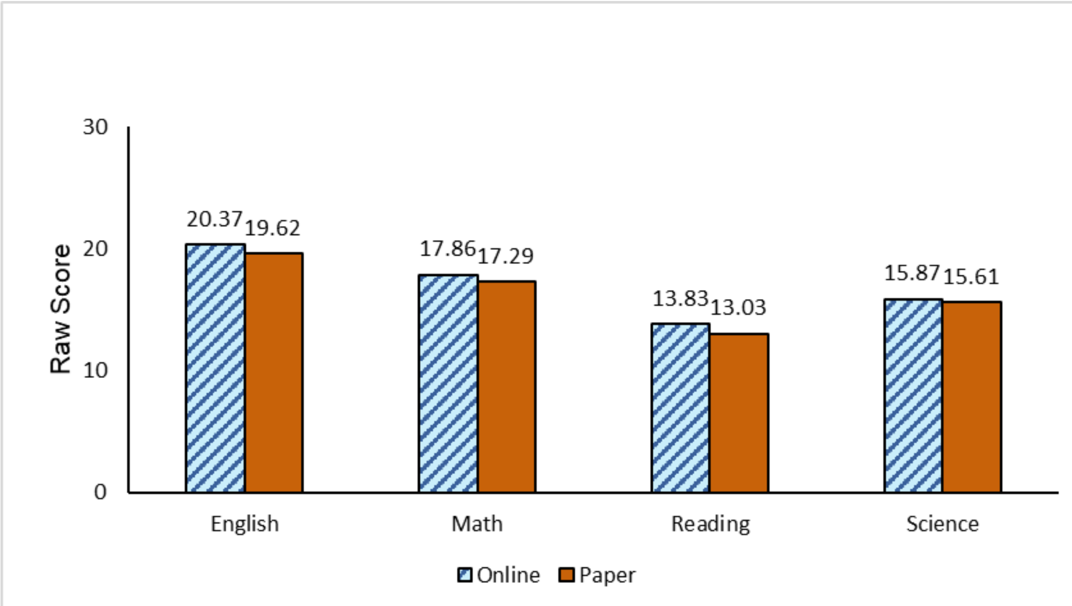


Figure 4.6. Scale Score Mean Comparison

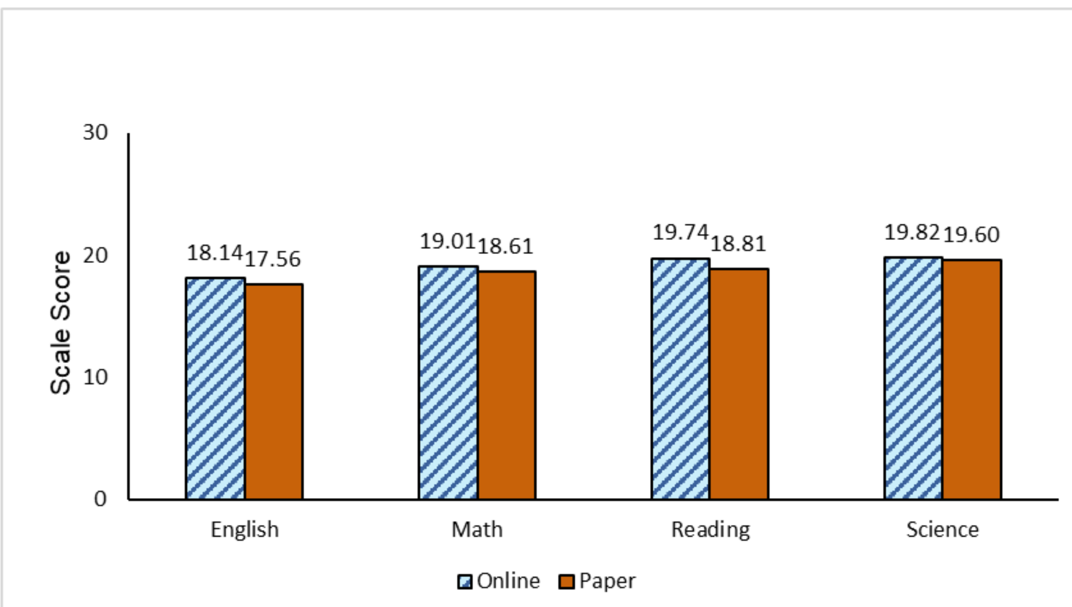
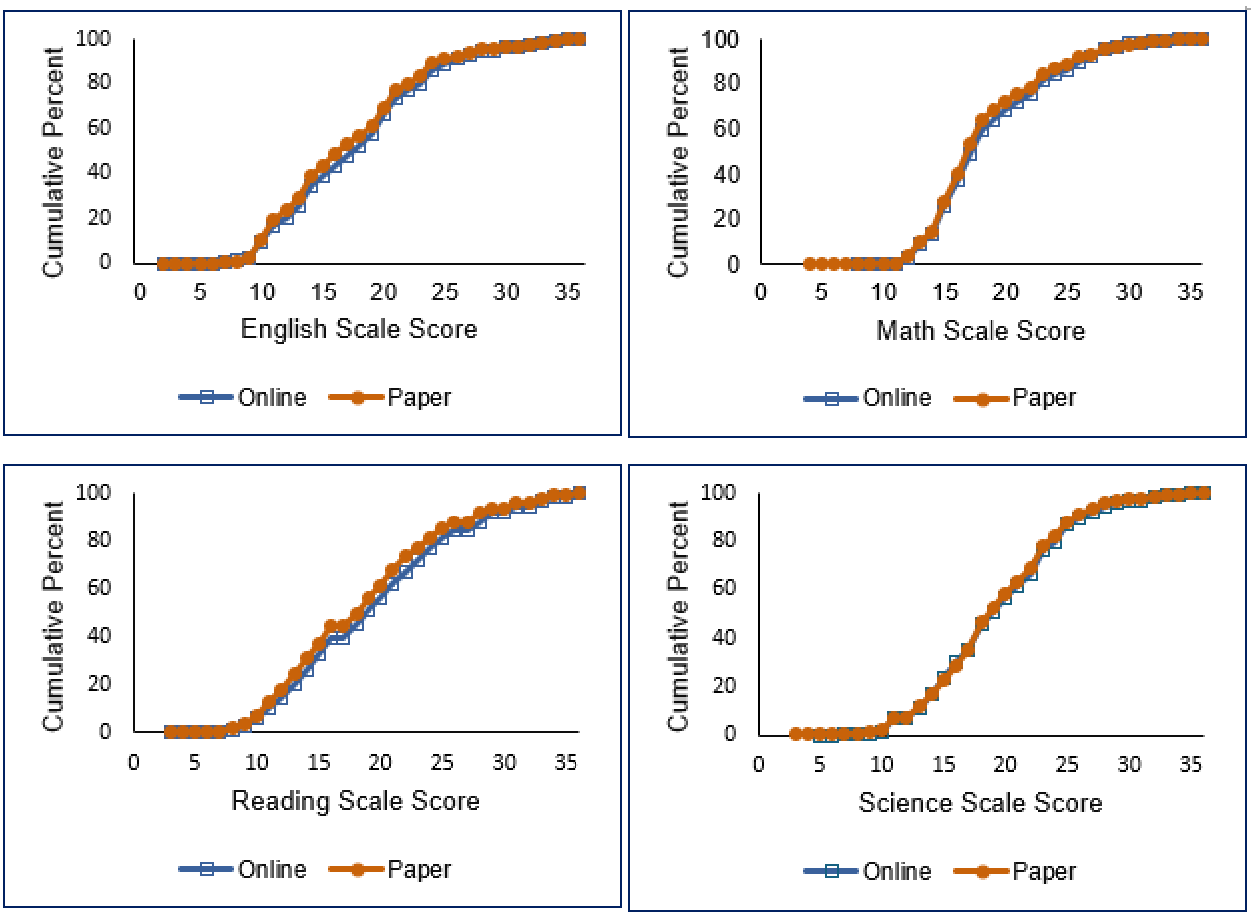


Table 4.5. Scale Score Mean Differences, Effect Sizes and *t*-Tests of Mean Differences (Online and Paper)

Test section	Mean difference	Effect size	Test *p*-value
English	0.578	0.096	<0.001
Math	0.396	0.082	0.005
Reading	0.922	0.140	<0.001
Science	0.217	0.041	0.154

Proportion correct raw score and scale score frequency distributions and cumulative frequency distributions were also compared across online and paper testing. Consistent with the mean score differences showed in Figures 4.5 and 4.6 and Table 4.5, score distributions for math and science were quite similar for both modes, and scores tended to be higher for the online group than for the paper group for English and reading.

Figure 4.7. Relative Cumulative Frequency Distributions of Scale Scores



The Kolmogorov-Smirnov (KS) test of equivalency of distributions was conducted for the scales of each section test across modes. The KS tests showed that the scale score distributions across modes were statistically significant at $\alpha = .01$ for English and reading.

Table 4.6 contains scale score correlations, effective weights, and Cronbach's alpha reliability estimates for the section tests across online and paper testing. Results for both modes were similar, providing evidence that the same constructs are measured through online and paper testing of the enhanced ACT.

Table 4.6. Scale Score Correlations, Effective Weights, and Cronbach's Alpha

Correlation	Online				Paper			
	Eng.	Math	Rdg.	Sci.	Eng.	Math	Rdg.	Sci.
English	—	0.69	0.80	0.74	—	0.69	0.76	0.70
Math	—	—	0.64	0.75	—	—	0.66	0.74
Reading	—	—	—	0.72	—	—	—	0.71
Science	—	—	—	1.00	—	—	—	1.00
Effective weights	0.27	0.20	0.30	0.23	0.26	0.21	0.30	0.23
Cronbach's alpha	0.87	0.85	0.85	0.81	0.87	0.84	0.84	0.81

4.3.1.2 Item-Level Comparison

The similarity of item score distributions for the English, math, reading, and science section tests of the enhanced ACT were compared between online and paper testing. The percentage of study participants who answered each item correctly (the item *p*-value), the correlation between the item score and the total test score (the item point-biserial value), and the percentages of study participants omitting any item (item omission rates) were analyzed for each section test.

The comparison of item *p*-value showed that items tended to be easier for the online testing group. The differences in item *p*-value were the largest for the reading test, especially for items that appeared later in the test. The comparison of item point-biserial correlations across online and paper testing did not reveal a consistent pattern of item performance. While some items demonstrated higher discrimination for the online testing group, others showed stronger performance for the paper testing group. Overall, the average item point-biserial correlations of the English and reading tests tended to be slightly higher for the online testing group. The item omission rates were consistently lower for the online testing group than for the paper testing group for the latter part of the tests, which partially explained why item *p*-values were higher for the online testing group. The differences in item omit rate were smallest for the math test.

4.3.1.3 Confirmatory Factor Analysis (CFA) and Differential Item Functioning (DIF)

Confirmatory factor analysis (CFA) was used to test whether the scoring and reporting structure of online and paper testing of the enhanced ACT are consistent. Additionally, differential item functioning (DIF) analysis was used to investigate whether items performed differently across the two modes at the same proficiency level on the enhanced test and, if so, whether the source of the difference could be identified.

CFA results showed that all models achieved acceptable mode fit for all section tests across both modes. Average standard loadings for all section tests are similar cross the two modes. The Mantel-Haenszel common odds-ratio (MH) procedure was implemented for the DIF analyses, where the paper testing group was treated as the focal group, and the online testing group was treated as the reference group. Only a few items were flagged for DIF, and all of them were determined to be appropriate measures of the intended construct across modes.

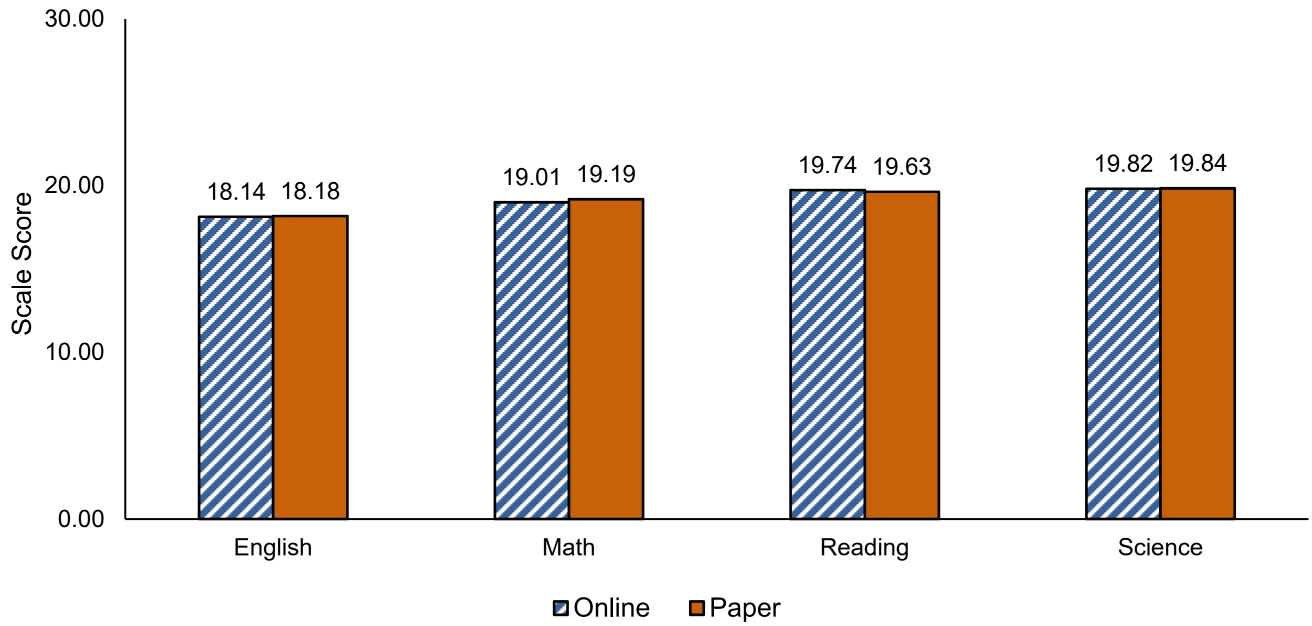
4.3.2 Mode Adjustments

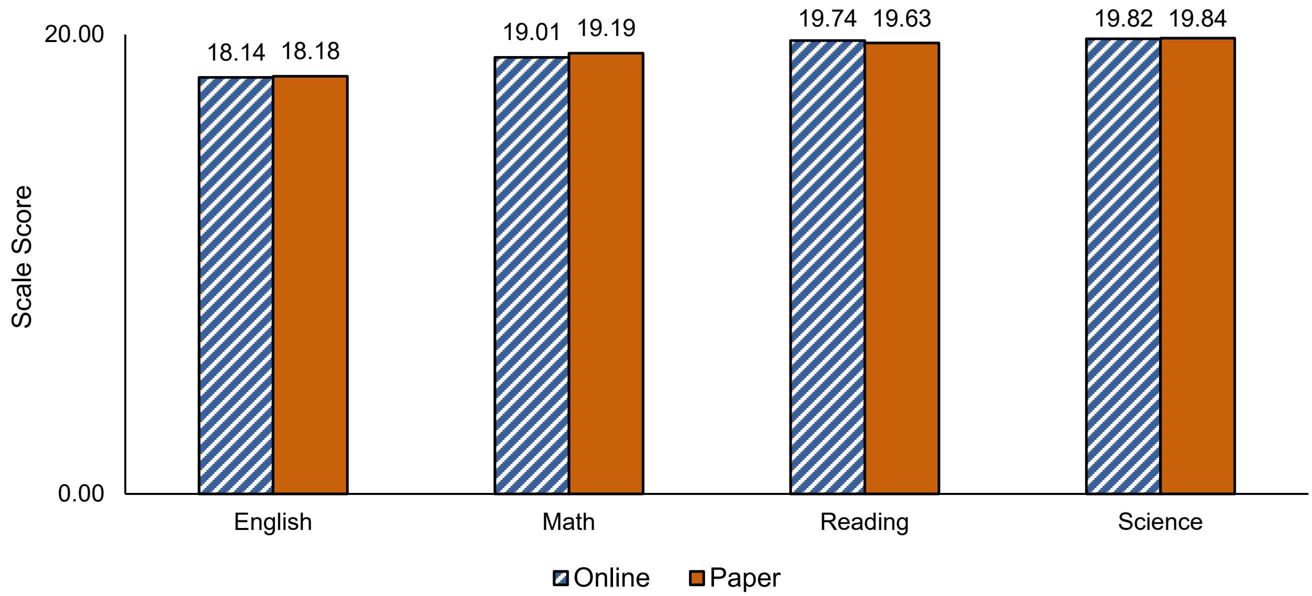
Results of the mode comparability analyses indicated that observed score distributions from online and paper testing are very similar for the two randomly equivalent groups and that the paper and online modes are measuring the same attributes. Results also suggested that using a mode adjustment would improve score comparability. Therefore, we used an equating methodology to ensure that the scores for students participating in the study were comparable, regardless of testing conditions.

The IRT true score equating method was used to link scores from paper testing to online testing. First, items on the enhanced ACT on paper and online were calibrated. Then, we used the Stocking and Lord (1983) procedure to transform the item parameters from paper testing to fit the same scale as that for the online ones. Finally, IRT true score equating was conducted to obtain raw to scale score conversions for the enhanced ACT on paper.

As expected, the scale score distributions of four section tests of the enhanced ACT were nearly identical across modes after the mode adjustment. Figure 4.8 depicts the scale score mean comparison across modes for English, math, reading, and science.

Figure 4.8. Scale Score Mean Comparison After Mode Adjustment





The October 2024 mode comparability study examined test and item level differences for the enhanced ACT administered online and on paper. Results showed very small differences for the paper and online groups' scores. Scores tended to be slightly higher for the online group, with the largest differences observed for reading. Equating methodology was applied to adjust for the differences so that reported scores from online and paper testing were comparable.

4.4 Concurrent Validity Evidence

The validity of interpretations of ACT test scores are supported by evidence of relationships of ACT scores with other measures of current academic performance. Arguments of concurrent validity for the enhanced ACT are supported by relationships of ACT scores from the enhanced ACT test with high school grades and prior test scores from the legacy ACT test. We compared correlations for two groups of students from the June 2024 linking study: those who were randomly assigned to the enhanced ACT and those who were randomly assigned to the legacy ACT. Refer to Table 4.1 for more information on the June 2024 sinking study sample.

In this section, we also examine interrelationships of ACT section test scores and compare the results for the enhanced ACT and legacy ACT, again using data collected from the June 2024 linking study.

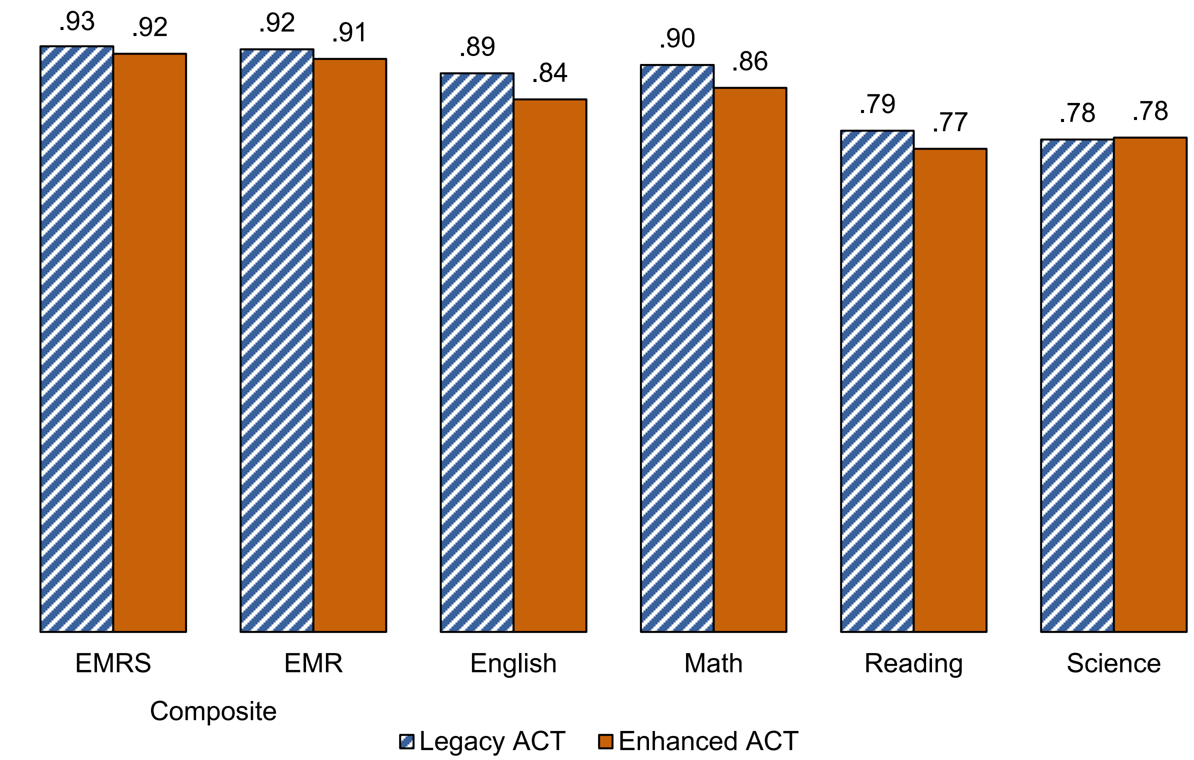
4.4.1 Correlations With Prior ACT Scores

Among the students who participated in the June 2024 linking study, 54% had previously taken the legacy ACT. We examined the extent that relationships with prior ACT test scores were different for students who took the enhanced ACT instead of the legacy ACT.

Figure 4.9 shows the correlations of the June 2024 ACT scores with prior ACT scores. Note that the students took their prior ACT tests from September 2022 through May 2024, so all tests were the legacy ACT. We again use the abbreviations EMRS (English, math, reading, science) to refer to Composite scores with science and EMR (English, math, reading) to refer to Composite scores without science.

We found that the Composite score correlations were very similar for the enhanced ACT and legacy ACT. Further, the Composite correlations were very similar with and without the science score. For the enhanced ACT, the Composite retest correlations were .91 (EMR) and .92 (EMRS). For the legacy ACT, the Composite retest correlations were .92 (EMR) and .93 (EMRS). This suggests that Composite scores are consistent across tests, even when the first test is the legacy ACT and the second test is the enhanced ACT.

Figure 4.9. Correlations of ACT Test Scores With Prior ACT Test Scores: Enhanced and Legacy ACT



For science, the retest correlation was .78 for both the enhanced ACT and the legacy ACT. For English, math, and reading, the retest correlations were higher for the legacy ACT relative to those for the enhanced ACT. However, the differences in correlations were only statistically significant for English and math. Slightly lower correlations were expected for the enhanced ACT because of its shorter test length (and lower reliability). Therefore, it is surprising that the retest correlations for science are the same for the enhanced ACT and legacy ACT. This may be explained by the number of scored items for the enhanced science test (34 items) being only 15% less than the number of scored items for the legacy science test (40 items).

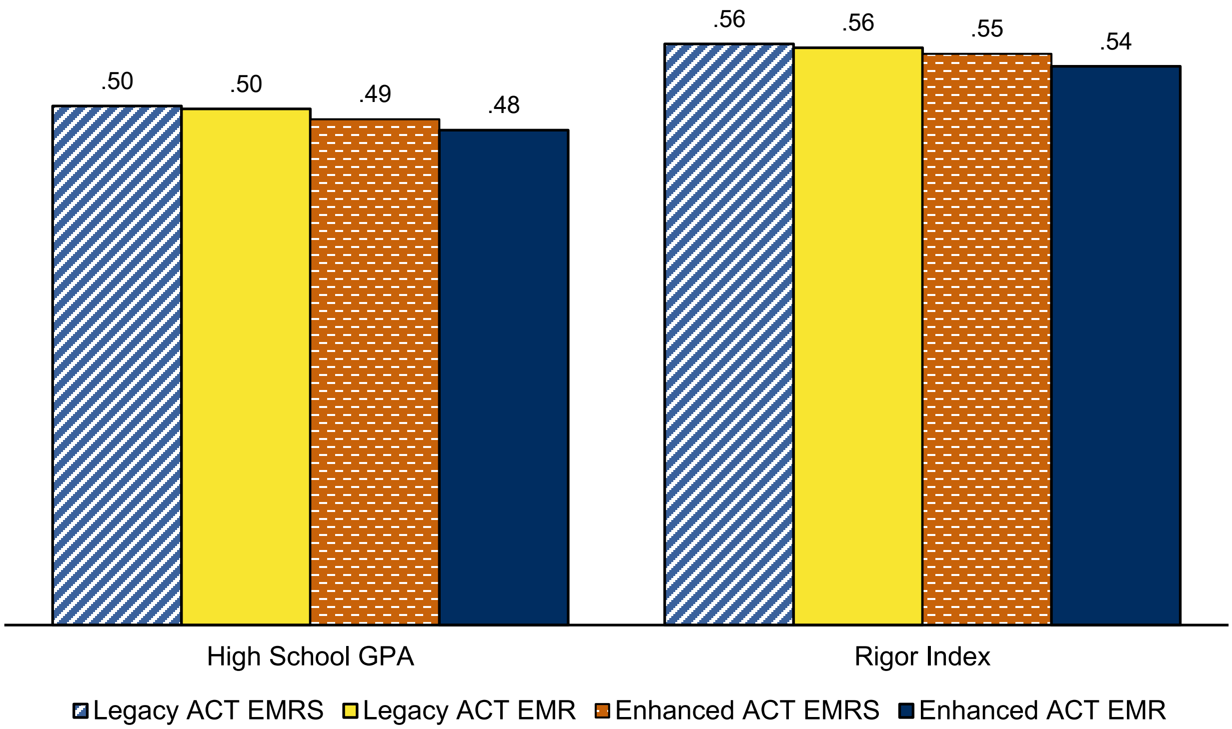
4.4.2 Correlations With High School Coursework and Grades

First, we compared the correlations between ACT Composite scores and high school grades. Through ACT's registration system (MyACT®), students have the option of self-reporting the courses they have taken in high school and the grades they have earned on the traditional 4-point scale. High school GPA (HSGPA) was calculated by averaging the students' self-reported grades earned in up to 23 core high school courses taken in English, math, social studies, and the natural sciences. In addition to comparing the correlations between Composite scores and HSGPA, we compared the correlations between ACT Composite scores and the ACT Rigor Index. The ACT Rigor Index is similar to HSGPA, but it awards more points for taking more difficult courses. The ACT Rigor Index is based on student-reported grades in up to 30 different high school courses (including foreign languages and arts, in addition to the core subject areas) and incorporates indicators for advanced coursework and students' plans for taking upper-level STEM courses (chemistry, physics, advanced math, and calculus). Relative to HSGPA, the ACT

Rigor Index has higher correlations with ACT test scores and college degree attainment (Allen & Mattern, 2019).

Figure 4.10 shows the correlations of the ACT Composite scores with the two measures of academic achievement. The correlations are based on students who participated in the June 2024 linking study and reported their high school coursework and grades. This included 89% of the students who took the enhanced ACT and 90% of the students who took the legacy ACT.

Figure 4.10. Correlations of ACT Composite Scores With Measures of High School Academic Achievement: Enhanced and Legacy ACT



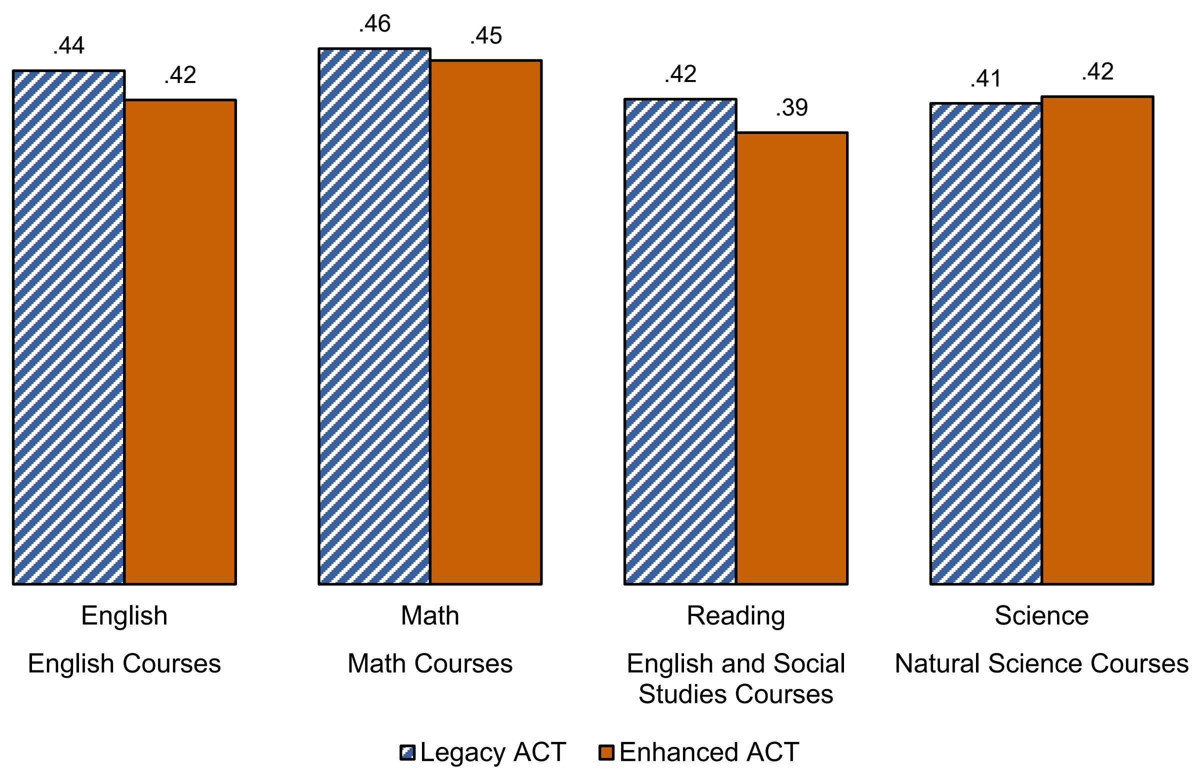
The correlations were slightly lower for the enhanced ACT relative to the legacy ACT. For HSGPA, the correlation was .49 for the enhanced EMRS score and .50 for the legacy EMRS score. For the enhanced EMR score, the HSGPA correlation was slightly lower (.48) than both the legacy EMR score (.50) and the enhanced EMRS score (.49). The differences in EMRS and EMR correlations with HSGPA between the enhanced ACT and legacy ACT were not statistically significant.

For the rigor index, the correlation was .56 for both the enhanced and legacy EMRS score. For the enhanced EMR score, the rigor index correlation was slightly lower (.54) than both the legacy EMR score (.56) and the enhanced EMRS score (.55). The differences in EMRS and EMR correlations with the rigor index between the enhanced ACT and legacy ACT were not statistically significant.

Second, we compared the correlations between students' ACT section scores and their respective high school grades in the four core subject areas (i.e., English, math, social studies, and natural science) (refer to Figure 4.11). Subject-area GPA was calculated for students who used MyACT to report their course grades in at least three English, math, social studies, and natural science courses. Correlations were calculated for ACT English, math, and science scores with English, math, and natural science course GPAs, respectively. However, we correlated the ACT reading score with the average of the social studies and English course GPAs.

We found that the correlations were slightly lower for the enhanced ACT relative to the legacy ACT for English and reading. However, none of the differences in correlations were statistically significant. For science, the correlations were the same for the enhanced and legacy ACT. These results suggest that the enhanced ACT and legacy ACT are similarly reflective of subject-area GPAs.

Figure 4.11. Correlations of ACT Section Test Scores With High School Subject GPA: Enhanced and Legacy ACT



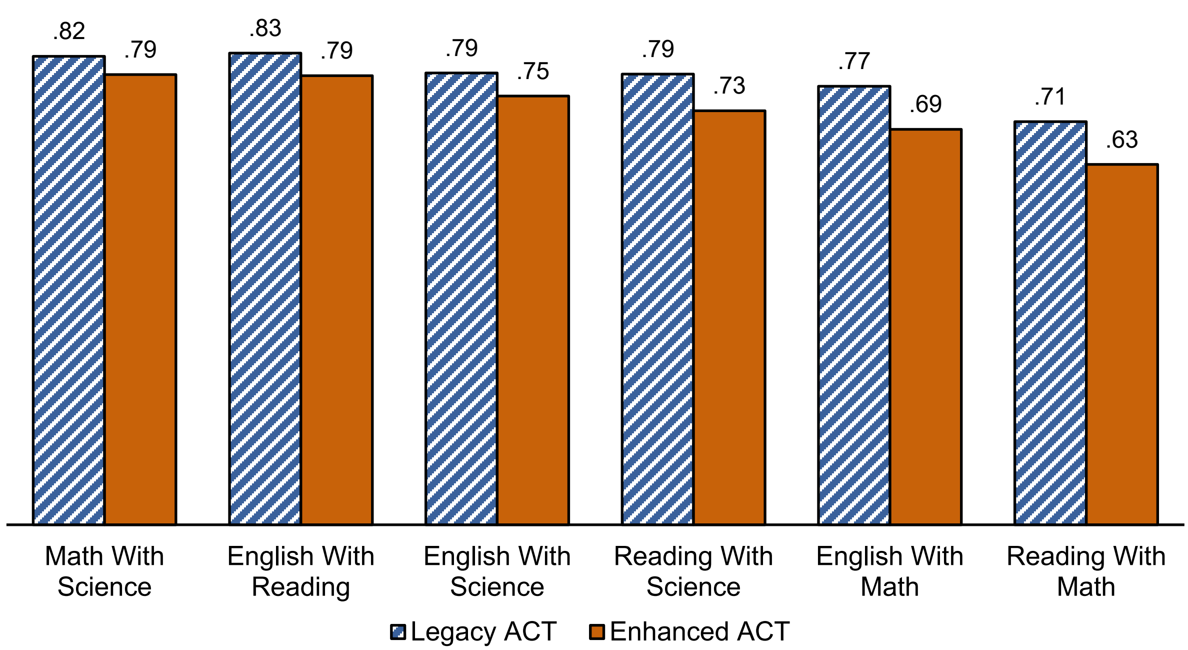
4.4.3 Interrelationships of ACT Scores

The intercorrelations of ACT section test scores provides evidence of convergent or divergent relationships among the constructs being measured. For example, higher English and reading correlations are evidence of convergent relationships, while lower English and math correlations are evidence of divergent relationships. Using the scores of all students who participated in the June 2024 linking study, we compared the intercorrelations of ACT section test scores for the enhanced ACT to those of the legacy ACT. We anticipated that the overall pattern of relationships among the section test scores would be comparable between the enhanced ACT and the legacy ACT, although we expected somewhat lower correlations for the enhanced ACT because of the anticipated lower reliability of these test scores.

Figure 4.12 provides the correlations among the ACT section test scores for students who were randomly assigned to the enhanced version of the ACT and for students who were randomly assigned to the legacy version of the ACT. As the figure shows, the correlations among the ACT section scores are somewhat lower for the enhanced ACT but exhibit similar patterns as the legacy ACT correlations. The lower correlations for the enhanced ACT are expected because of the lower reliability observed for the enhanced ACT test scores.

For both the enhanced and legacy ACT tests, the highest correlations (convergent relationships) were observed for English with reading and for math with science. For both versions of the ACT, the lowest correlations (divergent relationships) were observed for reading with math and for English with math. This suggests that the relationships among the measured constructs are comparable irrespective of the enhancements made to the ACT.

Figure 4.12. Correlations Among Section Test Scores: Enhanced and Legacy ACT



In addition to the observed correlations, disattenuated correlations (from which the impact of measurement error is removed) among the section test scores were assessed. Relative to the observed correlations, the disattenuated correlations are more similar between the legacy form and enhanced forms that were used for the June 2024 linking study (for more details, refer to Li et al., 2025).

4.5 Predictive Validity Evidence

The validity of interpretations of ACT test scores are also supported by evidence of relationships of ACT scores with subsequent measures of academic performance, such as first-year cumulative college GPA and college degree attainment. As such, predictive validity evidence for ACT scores requires time for the students who have taken the ACT to enter college. Because most students who took the enhanced ACT as part of the June 2024 linking study just completed 11th grade, predictive validity evidence specific to the enhanced ACT is not yet available (as of spring 2025).

4.5.1 Simulation Study Design

Historical data for students who took the legacy ACT and enrolled in college can be used to examine how the predictive validity of ACT section and Composite scores might change when those scores are based on a shorter test that matches the number of items scored for the enhanced ACT.

To achieve this, we used ACT test data and college outcomes data that had been collected for students who took the legacy ACT test. For each test form, we randomly selected the items from each section test to simulate shortened tests that match the enhanced ACT's number of scored items.[1] For example, for each English test form, we randomly selected 40 of the 75 test items from the full English test. For each test section, we then calculated the number correct and converted the number correct to a scale score so that the mean and standard deviation of the scale score matched the mean and standard deviation of the scale score from the legacy (full) test. Using the scale scores generated from the shortened test, we then calculated the Composite scores (EMRS and EMR) for the shortened test.

To compare the correlations of ACT Composite scores with first-year GPA between the shortened ACT and legacy ACT, we used a dataset of 1,111,776 students who completed high school between 2006 and 2017 and enrolled at one of 527 2- and 4-year colleges that provided ACT with students' first-year college GPA. To compare the correlations of ACT Composite scores with degree attainment between the shortened ACT and legacy ACT, we used a dataset composed of a random sample of 116,918 ACT-tested students who completed high school in 2017 and entered the subsequent fall into one of 2,578 2- and 4-year colleges for which degree attainment status was tracked over 6 years by the National Student Clearinghouse. More details on these samples are provided in a full research report (Allen & Cruce, 2025).

We hypothesized that the correlations between the scores of the shortened ACT with college outcomes would be comparable, albeit smaller, than the correlations that result from the scores of the legacy ACT. We caution, however, that scoring a shortened version of the legacy ACT for

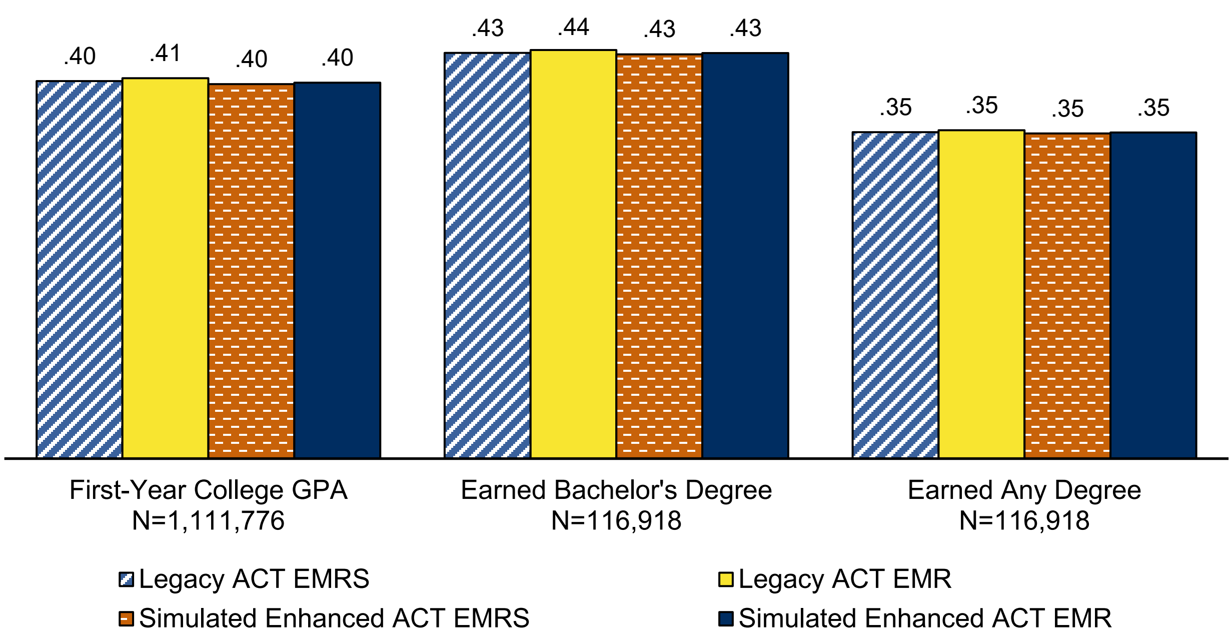
[1] We also explored other methods for selecting a subset of *p* items from the full set of *n* items including (a) select the first *p* items from each section test, (b) select the *p* items with the highest correlations with first-year college GPA, and (c) select the *p* items with the lowest correlations with first-year GPA. Results were similar across item selection methods.

students who took the legacy ACT does not fully capture the differences between the enhanced ACT test and the legacy ACT test for several reasons. First, the creation of the shortened ACT test forms and their scale scores does not follow the method of form construction and equating that is typically followed in creating the reporting scale. Second, the enhanced ACT has other design changes in addition to having fewer scored items than the legacy ACT. Additionally, the experience of taking the enhanced ACT test, with fewer test questions and more time per test question, cannot be fully simulated by simply scoring fewer items for students who took the legacy ACT. Therefore, the predictive validity evidence presented in this report is provisional and will require additional analyses based on actual postsecondary outcomes data for students who take the enhanced ACT.

4.5.2 Correlations With First-Year GPA and Degree Attainment

Figure 4.13 shows the correlation of the ACT Composite score with three college outcomes: first-year college GPA, having earned a bachelor's degree within 6 years of first enrolling in college, and having earned any undergraduate degree or certificate within 6 years of first enrolling in college. As seen in the figure, the correlation between the ACT Composite score and each of the three outcomes was essentially the same for the simulated enhanced ACT (shortened legacy ACT) and legacy ACT. Removing science from the shortened ACT Composite score did not change the relationship in any meaningful way.

Figure 4.13. Correlations of ACT Composite Scores With First-Year GPA and Degree Attainment: Legacy ACT and Simulated Enhanced ACT



The results suggest that shortening the ACT will have very little effect on the relationship between the Composite score and outcomes such as first-year college GPA and timely undergraduate degree attainment. Additional analyses presented in the full research report (Allen & Cruce, 2025) suggest that shortening the ACT has very little impact on how well the

Composite scores and section test scores predict success in different types of first-year courses. However, we again caution that this simulation of the enhanced ACT test does not fully capture the differences between the enhanced ACT test and the legacy ACT test. Therefore, the predictive validity evidence presented here is provisional and will require additional analyses based on actual postsecondary outcomes data for students who take the enhanced ACT test.

4.6 Composite Score Comparability

Beginning with ACT tests taken online on national test dates in April 2025, the ACT science test will be optional, and reported Composite scores will be based on English, math, and reading scores. Beginning in fall 2025, this change will be applied to all ACT test administrations. We examined the impact of removing science from the Composite score for six different ACT-tested groups:

- Students who took the enhanced ACT as part of the June 2024 linking study
- School day testers who took the ACT as part of state or district testing programs in 11th grade in spring 2024 (This group of students is most representative of the general population of 11th-grade high school students.)
- Students who took the ACT and completed high school in 2024 (This is the ACT-tested graduating cohort and is a mix of students who only took the ACT as part of a school-day testing program and students who tested on a National [Saturday] test date. This group represents all U.S. students who take the ACT.)
- Students who took the ACT and applied to at least one of 14 select colleges and universities as part of the prospective student cohorts of 2021, 2022, or 2023
- Students who took the ACT, completed high school in 2022, and enrolled in college during the 2022–2023 college academic year (This group represents U.S. students who take the ACT and go on to college immediately after high school.)
- Students who took the ACT, completed high school in 2022, and enrolled in a 4-year college during the 2022–2023 college academic year (This group represents U.S. students who take the ACT and go on to a 4-year college immediately after high school.)

More details on the six groups are provided in the full research report (Allen & Cruce, 2025).

As measures of general academic achievement, the constructs represented by the Composite scores with and without science (i.e., EMRS and EMR) have a lot of overlap. Not only are three out of the four section test scores included in both measures, but also students' English, math, and reading scores are highly correlated with their science score (refer to Figure 4.12). As a result of the overlap in the constructs being measured by EMR and EMRS, these two scores are also very highly correlated with one another. Across the six different ACT-tested student groups, the correlation between EMR and EMRS is 0.99, suggesting that very little unique information about the student is lost when science is removed from the Composite score.[2] Given the very strong relationship between EMR and EMRS, we would expect EMR and EMRS to be similarly related to other concurrent and subsequent measures of academic achievement. Evidence from the analyses summarized in this chapter supports this presumption; we found differences that were never any larger than one-hundredth of a point when comparing EMR and EMRS correlations with concurrent measures such as prior ACT test scores (refer to Figure 4.9), HSGPA and the high school rigor index (refer to Figure 4.10), and with subsequent measures such as first-year college GPA and college degree attainment (refer to Figure 4.13).

4.6.1 Concordances of Composite Scores

Although EMR and EMRS are very highly correlated, it is possible for differences to exist across the 1–36 score scale. We used equipercentile concordance to examine the comparability of EMRS and EMR scores across the score scale. For each possible EMRS score, we identified the EMR score that has the closest percentile rank. Table 4.7 presents the results for each group. Note that concordance estimates are not provided for EMRS score points with a cumulative frequency of fewer than 50 examinees because the estimates are not reliable when based on a small number of examinees.

For EMRS scores from 14 to 36, the most comparable EMR score was equal to the EMRS score. Although EMRS scores below 14 were typically most comparable to EMR scores that were 1 point lower, Composite scores in this range are generally not used to support high school graduation requirements or decisions regarding college admission, scholarships, and placement within college courses. The results were very consistent across the six ACT testing groups and suggest that EMR scores support the same normative interpretations as EMRS scores.

Table 4.7. Concordances of ACT Composite Scores With and Without Science, by ACT Tested Population

EMRS score	Comparable EMR Score: June 2024 enhanced ACT	Spring 2024 school day	2024 high school cohort	Selected college applicants	2022 college enrollees	2022 4-year college enrollees
1	*	*	*	*	*	*
2	*	*	*	*	*	*
3	*	2**	2**	*	*	*
4	*	3**	3**	*	*	*
5	*	4**	4**	*	*	*
6	*	5**	5**	*	5**	*
7	*	6**	6**	*	6**	6**

[2] This phenomenon is not particular to the absence of the science score. A measure of general academic achievement based on any three of the four section test scores is very highly correlated with a measure of general academic achievement based on all four section test scores, implying that any of the four would add very little unique information once the other three section test scores have been included in the measure of general academic achievement.

EMRS score	Comparable EMR Score June 2024 enhanced ACT	Spring 2024 school day	2024 high school cohort	Selected college applicants	2022 college enrollees	2022 4-year college enrollees
8	*	7**	7**	*	7**	7**
9	*	8**	8**	*	8**	8**
10	*	9**	9**	9**	9**	9**
11	*	10**	10**	10**	10**	10**
12	11**	11**	11**	11**	11**	11**
13	13	12**	13	12**	13	13
14	14	14	14	14	14	14
15	15	15	15	15	15	15
16	16	16	16	16	16	16
17	17	17	17	17	17	17
18	18	18	18	18	18	18
19	19	19	19	19	19	19
20	20	20	20	20	20	20
21	21	21	21	21	21	21
22	22	22	22	22	22	22
23	23	23	23	23	23	23
24	24	24	24	24	24	24
25	25	25	25	25	25	25
26	26	26	26	26	26	26
27	27	27	27	27	27	27
28	28	28	28	28	28	28
29	29	29	29	29	29	29
30	30	30	30	30	30	30
31	31	31	31	31	31	31
32	32	32	32	32	32	32
33	33	33	33	33	33	33
34	34	34	34	34	34	34
35	35	35	35	35	35	35
36	36	36	36	36	36	36

*Estimates are not available for EMRS score points with cumulative frequencies < 50.
**These comparable scores are different.

It is important to note that the concordance results presented in Table 4.7 are based on analyses of students who took either the legacy or enhanced ACT test and thus only inform EMRS/EMR score comparability within each version of the ACT. Therefore, there is strong evidence of score comparability for the enhanced ACT EMRS score and the enhanced ACT EMR score and also for the legacy ACT EMRS score and the legacy ACT EMR score. However, the concordances in Table 4.7 do not directly address score comparability for the legacy ACT EMRS score and enhanced ACT EMR score.

To establish score comparability for the legacy ACT EMRS score and enhanced ACT EMR score, we estimated an additional concordance using data from the June 2024 linking study. Examinees who took the enhanced ACT and legacy ACT are random equivalent groups, so concordance can be estimated using the equipercentile method. We found that EMR from the enhanced ACT concords with EMRS from the legacy ACT for EMRS scores of 14 or higher.

Concordance estimates for EMRS scores below 14 were not obtained because the cumulative frequency of scores was less than 50 in that score range, resulting in unreliable estimates.

4.6.2 Composite Score Differences

There are some practical differences between the ACT Composite score with and without science that will result in some students having an EMR score that is either higher or lower than their EMRS score would have been. The first difference is in the rounding logic that gets applied to the score when moving from a Composite score that is based on four section tests to a Composite score that is based on three section tests. Both EMR and EMRS are calculated by rounding the average of the section test scores to the nearest whole number. Under EMRS, the mean of the four section test scores can take one of four possible values after the decimal: .00, .25, .50, and .75. Given that about one-quarter of all students fall into each scenario, about 25% have a mean score that does not require rounding, 25% have a mean score that is rounded down, and 50% of students have a mean score that is rounded up. Under EMR, the mean of the three section test scores can take one of three possible values after the decimal: .00, .33, and .67. Given that about one third of all students fall into each scenario, about 33% have a mean score that does not require rounding, about 33% have a mean score that is rounded down, and about 33% of students have a mean score that is rounded up. This shift from EMRS to EMR means that fewer students will have a mean score that has been rounded up.

The second practical difference between EMR and EMRS is the presence or absence of the science score within the computation of the Composite score. In particular, when students' science score is higher than their other three section test scores, the absence of the science score from the Composite means that these students are far more likely to have an EMR score that is lower than their EMRS score would have been. Conversely, when students' science score is lower than their other three section test scores, the absence of the science score from the Composite means that these students are far more likely to have an EMR score that is higher than their EMRS score would have been.

To examine the extent to which students' EMR score might differ from their EMRS score, we calculated a difference measure as

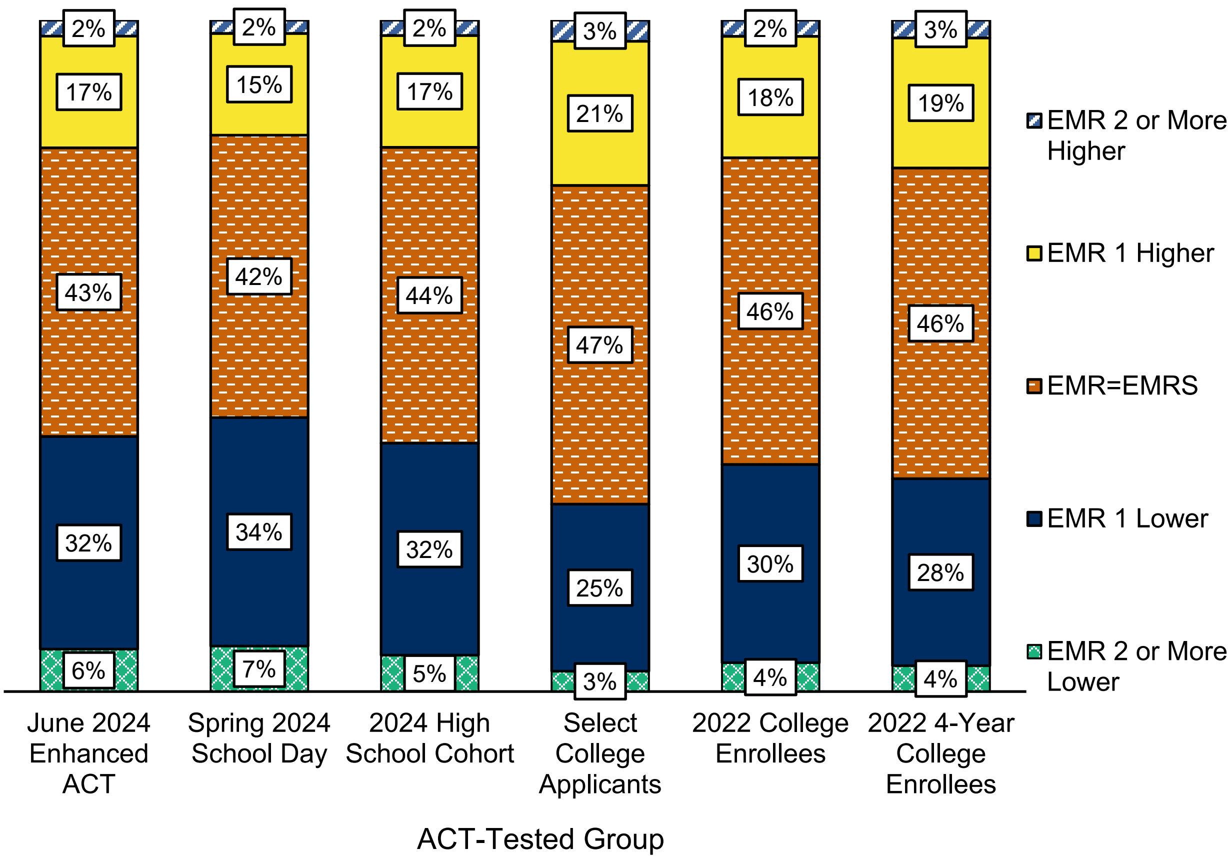
$$\text{Difference} = \text{EMR} - \text{EMRS}.$$

We classified the difference scores as the following:

- EMR score is **2 or more points higher** than the EMRS score.
- EMR score is **1 point higher** than the EMRS score.
- EMR score is the **same** as the EMRS score.
- EMR score is **1 point lower** than the EMRS score.
- EMR score is **2 or more points lower** than the EMRS score.

Figure 4.14 shows the percentage of students in each difference score category. Across the six groups, 42–47% of the students had no change in Composite score. The percentage of students whose EMR score was within 1 score point of their EMRS score ranged from 91% for the school-day testing population to 94% for the college applicants and college enrollees.

Figure 4.14. Differences in ACT Composite Scores, With and Without Science

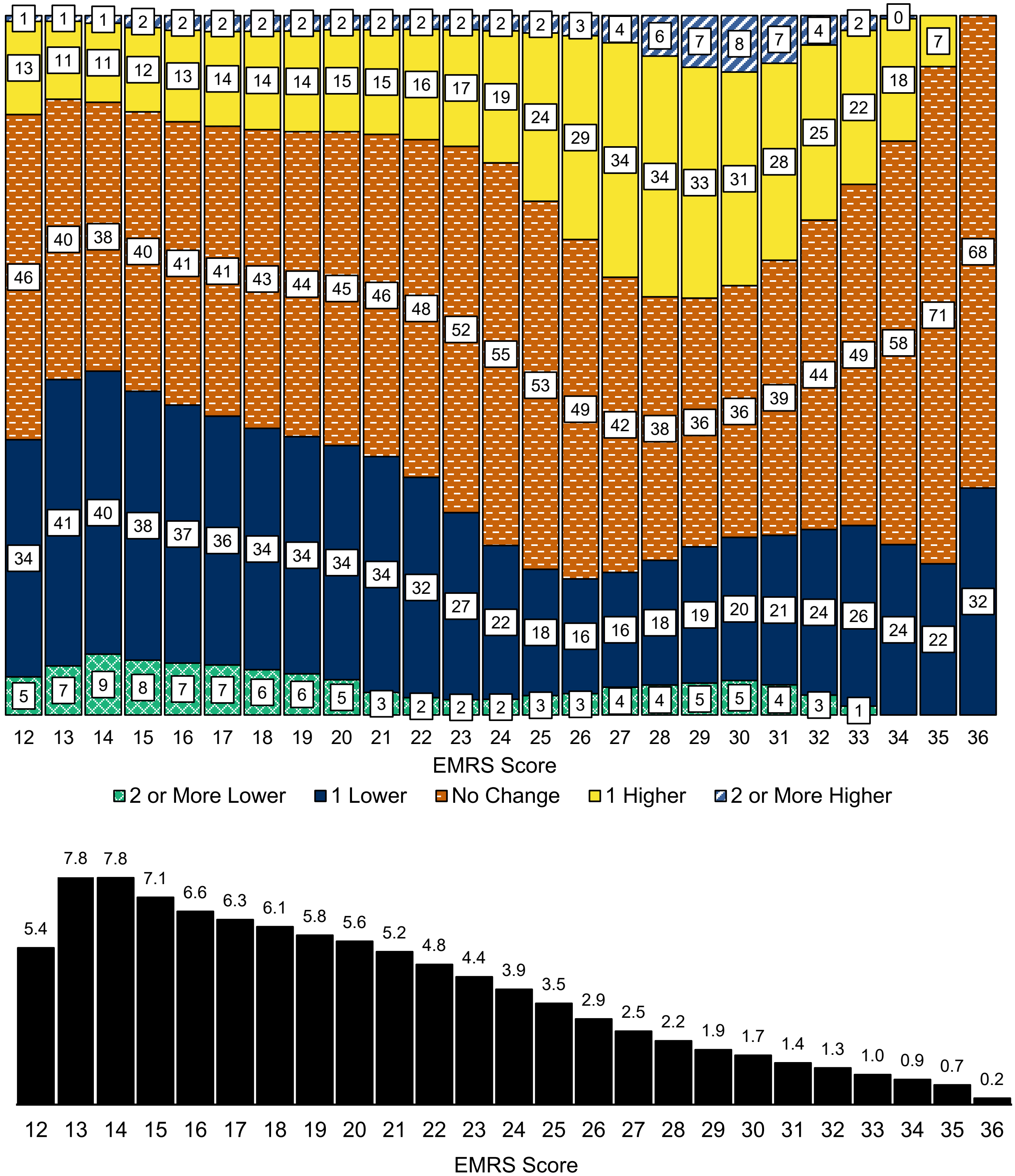


For students whose Composite scores differed, it was more common for their EMR score to be lower than their EMRS score. For example, while 37% of the students in the 2024 high school graduating cohort had an EMR score that was lower than their EMRS score, only 19% had an EMRS score that was lower than their EMR score. Although some of this imbalance between the share of students whose EMR score is lower than (as opposed to higher than) their EMRS score can be explained by the differences in the rounding logic between the calculations of EMR and EMRS, the imbalance is also because of more students having a science test score that is high relative to their other three section test scores.

Figure 4.15 shows the percentage of students in each difference score category by their EMRS score for the 2024 high school graduating cohort. As seen in the figure, at every point along the EMRS score scale, over 50% of students have an EMR score that is either the same as or higher than their EMRS score. However, students at the lower range of the score scale have greater chances than other students of having an EMR score that is lower than their EMRS score, whereas students at the higher range of the score scale have greater chances than other students of having an EMR score that is higher than their EMRS score. This means that many students in the lower score range have a science score that is high relative to their other three section test scores, whereas many students in the higher score range have a science score that is low relative to their other three section test scores. This difference in the role of science across the EMRS score scale is perhaps more evident when examining mean science scores for students at each point on the EMRS score scale. As seen in Table 4.8, at the lower range of the EMRS score scale, the average science score is higher than the students' EMRS score, suggesting that science is often one of the higher section test scores for these students. Conversely, at the higher range of the EMRS score scale, the average science score is lower than the EMRS score, suggesting that science is often one of the lower section test scores for these students.

Also provided in Figure 4.15 is a histogram showing the percentage of the 2024 high school graduating cohort by their EMRS score. As is evident by the juxtaposition of the histogram with the chart showing the percentage of students in each difference score category, the score range in which most students are located is also the area in which students are most likely to have an EMR score that is lower than their EMRS score. This explains the overall imbalance shown in Figure 4.14 between the share of students whose EMR score is lower than (as opposed to higher than) their EMRS score.

Figure 4.15. **Differences in ACT Composite Scores With and Without Science by EMRS Score**



Note. Both the stacked bar chart and histogram display percentages of students. Because of space constraints, EMRS scores below 12 are not provided in the figure. About 3.2% of the 2024 high school graduating cohort received an EMRS score of 11 or lower.

Table 4.8. Mean Science Score by EMRS Score, 2024 High School Cohort

EMRS score	Mean science score
14	14.8
15	15.6
16	16.5
17	17.4
18	18.4
19	19.3
20	20.2
21	21.1
22	22.0
23	22.8
24	23.6
25	24.3
26	25.1
27	26.0
28	26.9
29	27.9
30	28.9
31	30.1
32	31.3
33	32.4
34	33.6
35	34.7
36	35.7

For students in the 2024 high school cohort, we examined the Composite score differences for gender and racial/ethnic student groups (Figure 4.16). Over 91% of students within each student group had an EMR score that fell within +/– 1 point of their EMRS score (which is the standard error of measurement for EMRS). Moreover, within each group, over half of all students had an EMR score that was either the same as or higher than their EMRS score. For students whose EMRS and EMR scores were different, it was more common for the EMR score to be lower than the EMRS score, no matter which student group is considered. However, the extent of the imbalance between the share of students whose EMR score is lower than (as opposed to higher than) their EMRS score differs by gender and by race/ethnicity. This suggests that some groups of students are more likely than others to have science scores that are high relative to their other three section test scores. Mean scores reported by student group for the 2024 high school graduating cohort provide further evidence to support this conclusion (refer to Table 4.9). Specifically, we found that the higher the student group's mean science score compared to their mean EMRS score, the greater the imbalance seen in Figure 4.16 for that student group; in contrast, the lower the student group's mean science score compared to their mean EMRS score, the better the balance seen in Figure 4.16 for that student group.

Figure 4.16. Differences in ACT Composite Scores With and Without Science, by Student Group

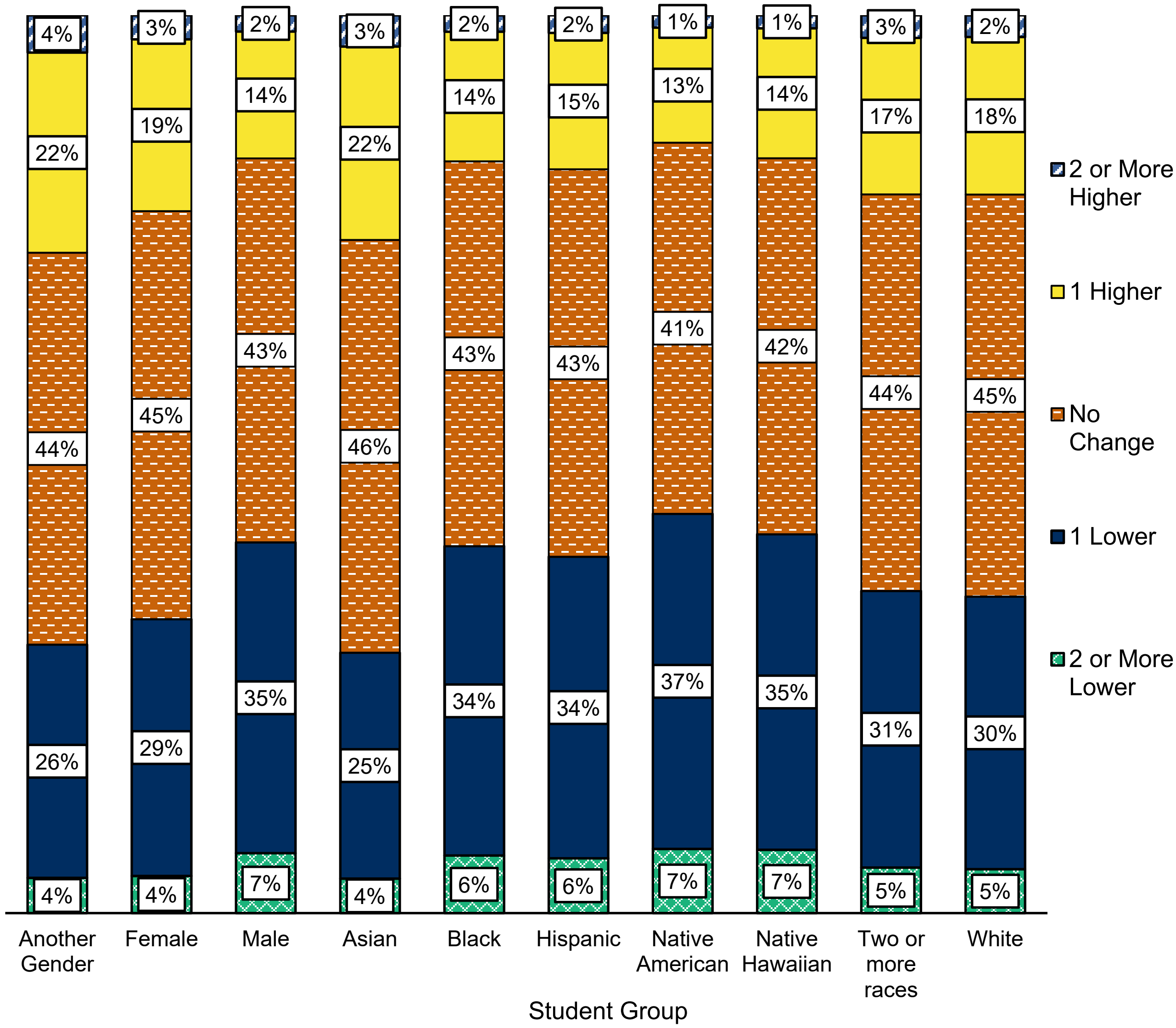


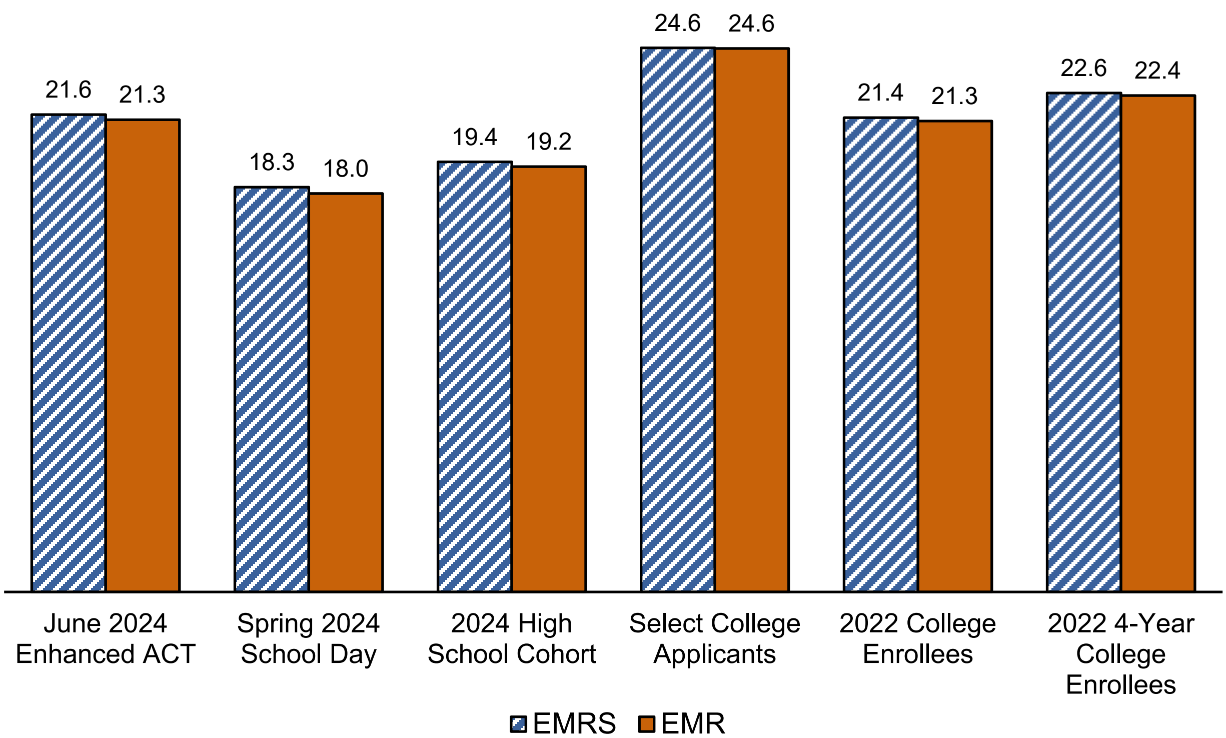
Table 4.9. Mean Science Score and EMRS Score by Student Group, 2024 High School Cohort

Student group	Science	EMRS
Another gender	20.7	21.1
Female	19.5	19.6
Male	19.8	19.4
Asian	23.7	24.1
Black	16.3	16.0
Hispanic	17.6	17.4
Native American	16.5	15.9
Native Hawaiian	16.8	16.3
Two or more races	19.8	19.7
White	21.0	20.9

4.6.3 Mean Scores

Given the individual student differences between EMR and EMRS previously noted, the two scores will have similar, but not identical, distributions. In fact, there is a 95.7% overlap in the two distributions. As suggested by the information presented in Figure 4.15, however, the distribution for EMR will have a flatter peak with more students at the lower tail when compared to the distribution for EMRS. The small differences in the EMR and EMRS score distributions mean that the summary statistics for these distributions will also be similar but not identical. The average EMRS and EMR scores are presented in Figure 4.17. For five of the six groups of ACT-tested students, the average EMR score was slightly lower than the average EMRS score. The difference was 0.3 score points for the 2024 school-day tested population and for the June 2024 enhanced ACT, 0.2 score points for the 2024 high school cohort and 2022 4-year college enrollees, and 0.1 score points for the 2022 college enrollees. For the group of applicants to the select colleges, the average EMRS and EMR scores were the same.

Figure 4.17. Average ACT Composite Scores



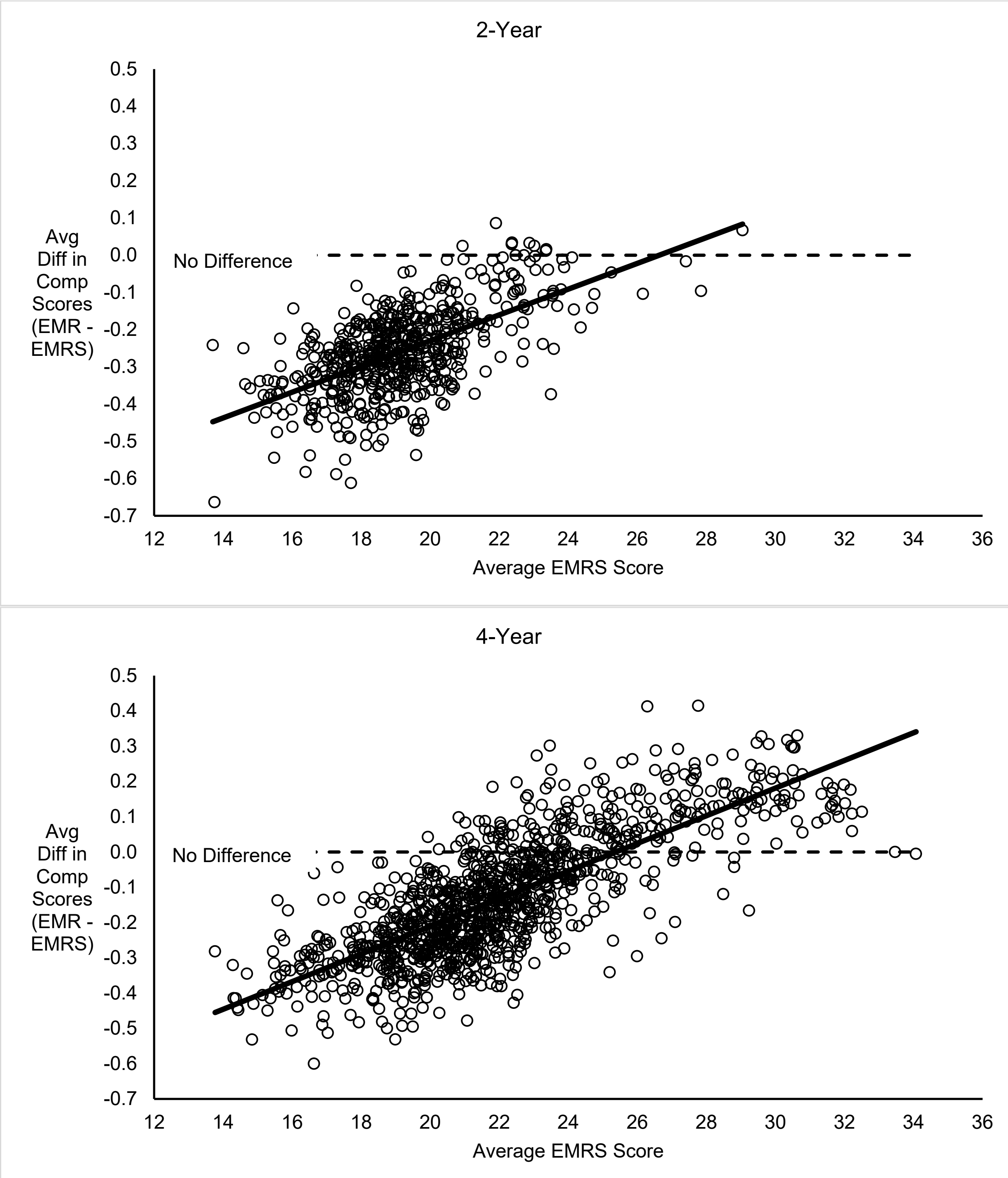
Given that some groups of students are more likely than others to have science scores that are high relative to their other three section test scores, we would anticipate that differences in the summary statistics for the ACT Composite score with and without science would vary across student groups. Table 4.10 provides the mean EMR and EMRS scores by student group for the 2024 high school graduating cohort. Across all student groups, the mean ACT Composite score without science is either slightly lower than or the same as the mean ACT Composite score with science, with differences between the two means ranging from 0.0 to 0.3 points. As anticipated, student groups that were more likely to have science scores that are high relative to their other three section test scores had larger mean differences between EMR and EMRS.

Table 4.10. Mean EMR Score and EMRS Score by Student Group, 2024 High School Cohort

Student group	EMR	EMRS
Another gender	21.1	21.1
Female	19.5	19.6
Male	19.1	19.4
Asian	24.0	24.1
Black	15.7	16.0
Hispanic	17.1	17.4
Native American	15.6	15.9
Native Hawaiian	16.0	16.3
Two or more races	19.5	19.7
White	20.8	20.9

In addition to examining differences by student groups, we used the college enrollee data to examine differences in average Composite scores with and without science by postsecondary institution. We considered 569 2-year colleges and 1,117 4-year colleges, each with at least 100 enrolled ACT-tested students. In Figure 4.18, we plot the difference in average Composite scores with and without science (i.e., EMR, EMRS, respectively) on the *y*-axis against each college's average EMRS score on the *x*-axis. As seen in the figure, for almost all 2-year colleges and for many 4-year colleges, the mean difference between EMR and EMRS is negative, meaning that the mean EMR score is lower than the mean EMRS score. However, we found that this mean difference between EMR and EMRS shrinks as the average ACT Composite score at the college increases. In fact, eventually the relationship flips direction such that some 4-year institutions with very high average ACT Composite scores (e.g., above 25) are actually more likely to observe *increases* in their average ACT Composite score when science is removed from the Composite. This variability between colleges in the direction and magnitude of the difference between EMR and EMRS averages is expected given the differences in students' science scores relative to their other three section test scores at different ranges of the EMRS score scale shown in Figure 4.15 and Table 4.8.

Figure 4.18. Average Difference in Composite Scores, by Postsecondary Institution



4.7 Conclusion

The enhanced ACT has several important changes from the legacy ACT, with notable reductions in test length and testing time and more time allowed per test question. The enhanced ACT also has other important changes to test design and test content. Because of these changes to the test, it is important to evaluate the evidence to make sure the scores for the enhanced ACT test will work well for all the different ways that people use these scores.

The design and implementation of the June 2024 linking study places section test scores from the enhanced ACT on the same scale as those from the legacy test. Data collected from the study also support the claim that the two tests measure the same constructs. The study also confirmed a small loss in reliability for the enhanced ACT relative to the legacy ACT, which was expected because of the enhanced ACT's shorter length. The loss of reliability is also reflected in larger standard errors of measurement and a small decline in classification accuracy.

The October 2024 mode comparability study examined test- and item-level differences for the enhanced ACT administered online and on paper. Results showed very small differences for the paper and online groups. Scores tended to be slightly higher for the online group, with the greatest differences observed for reading. Equating methodology was applied to adjust for the differences so that scores are comparable for online and paper testing.

We consistently found that measures of concurrent and predictive validity were very similar for the enhanced ACT and legacy ACT. While we expected a small loss in reliability for the enhanced ACT because of its shorter length, we found that validity coefficients were remarkably similar for the enhanced and legacy ACT. The evidence from a simulation study suggests that scores from the enhanced ACT will have predictive utility similar to scores from the legacy ACT.

Moreover, virtually the same normative interpretations of ACT Composite scores are supported, regardless of whether the Composite score is based on the enhanced ACT and excludes science, or whether the Composite score is based on the legacy ACT and includes science. The evidence supports the continued use of ACT scores from the enhanced ACT for college admissions, scholarships, high school accountability, program evaluation, course placement, and identifying students in need of academic support.

References

ACT. (2010). *Mind the gaps: How college readiness narrows achievement gaps in college success*. https://www.act.org/content/dam/act/unsecured/documents/MindTheGaps.pdf

ACT. (2012). ACT National Curriculum Survey 2012: Policy implications on preparing for higher standards. https://www.act.org/content/dam/act/unsecured/documents/NCS-PolicySummary2012.pdf

ACT. (2013a). *Relationship between ACT Composite score and Year 6 college cumulative GPA among students enrolled in a four-year postsecondary institution* (Information Brief No. 2013-6). https://www.act.org/content/dam/act/unsecured/documents/Info-Brief-2013-6.pdf

ACT. (2013b). *Relationship between ACT Composite score, high school GPA, and Year 6 college cumulative GPA among students enrolled in a four-year postsecondary institution* (Information Brief No. 2013-7). https://www.act.org/content/dam/act/unsecured/documents/Info-Brief-2013-7.pdf

ACT (2016). ACT National Curriculum Survey 2016. https://www.act.org/content/dam/act/unsecured/documents/NCS_Report_Web_2016.pdf

ACT (2020). ACT National Curriculum Survey 2020. https://www.act.org/content/dam/act/unsecured/documents/NCS_Report_Web_2020.pdf

ACT. (2024). *ACT technical manual (January 2024)*. https://www.act.org/content/dam/act/unsecured/documents/ACT_Technical_Manual.pdf

Adams, M. J. (2009). The challenge of advanced texts: The interdependence of reading and learning. In E. Hiebert (Ed.), *Reading more, reading better: Are American students reading enough of the right stuff?* (pp. 163–189). Guilford. https://www.childrenofthecode.org/library/MJA-ChallengeofAdvancedTexts.pdf

Allen, J., & Cruce, T. (2025). *Initial evidence supporting interpretations of scores from the enhanced ACT test* (Working Paper No. R2425). https://www.act.org/content/dam/act/unsecured/documents/R2425-Initial-Evidence-Interpretations-Enhanced-ACT-2025-03.pdf

Allen, J., & Mattern, K. (2019). Examination of indices of high school performance based on the graded response model. *Educational Measurement: Issues and Practice*, *38*(2), 41–52. https://doi.org/10.1111/emip.12250

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (Eds.). (2014). *Standards for educational and psychological testing*. American Educational Research Association. https://www.testingstandards.net/uploads/7/6/6/4/76643089/standards_2014edition.pdf

Anguiano-Carrasco, C., McVey, J., & Steedle, J. (2025, April 25). *Psychometric evaluation ofc culturally relevant items in a college admissions test* [Paper presentation]. NCME Annual Conference, Denver, CO, United States.

Arthur, A., & Thomas, J. (2023, April 13). *Using eyetracking to address reading rates and speededness on high stakes assessments* [eBoard presentation]. NCME Annual Conference, Chicago, IL, United States. http://dx.doi.org/10.13140/RG.2.2.29263.19365

Bennett, R. E. (2010). Cognitively based assessment of, for, and as learning (CBAL): A preliminary theory of action for summative and formative assessment. *Measurement*, *8*(2–3), 70–91. https://doi.org/10.1080/15366367.2010.508686

Biggs, J. B., & Collis, K. F. (1982). *Evaluating the quality of learning: The SOLO taxonomy (structure of the observed learning outcome)*. Academic Press.

Bryant, R. T. A. F. (2015). *College preparation for African American students: Gaps in the high school educational experience*. Center for Law and Social Policy. https://files.eric.ed.gov/fulltext/ED561728.pdf

Carver, R. P. (1983). Is reading rate constant or flexible? *Reading Research Quarterly*, *18*(2), 190–215. https://doi.org/10.2307/747517

Carver, R. P. (1990). *Reading rate: A review of research and theory*. Academic Press.

Carver, R. P. (1992). Reading rate: Theory, research, and practical implications. *Journal of Reading*, *36*(2), 84–95. https://www.jstor.org/stable/40016440

Chi, M. T., Feltovich, P. J., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science*, *5*(2), 121–152. https://doi.org/10.1207/s15516709cog0502_2

DiSessa, A. A. (1993). Toward an epistemology of physics. *Cognition and Instruction*, *10*(2/3), 105–225. https://doi.org/10.1080/07370008.1985.9649008

Driver, R., & Easley, J. (1978). Pupils and paradigms: A review of literature related to concept development in adolescent science students. https://doi.org/10.1080/03057267808559857

Eaton, F., Anderson, C. W., & Smith, E. L. (1984). Students' misconceptions interfere with science learning: Case studies of fifth-grade students. *The Elementary School Journal*, *84*(4), 365–379. https://doi.org/10.1086/461370

Flores, A. (2007). Examining disparities in mathematics education: Achievement gap or opportunity gap? *The High School Journal*, *91*(1), 29–42. https://doi.org/10.1353/hsj.2007.0022

Gallo, D. R. (1972). *Reading rate and comprehension: 1970–71 Assessment* (Report No. 02-R-09). National Assessment of Educational Progress. https://files.eric.ed.gov/fulltext/ED076934.pdf

Glaser, R., Chudowsky, N., & Pellegrino, J. W. (Eds.). (2001). *Knowing what students know: The science and design of educational assessment*. National Academies Press.

Gotwals, A. W., & Songer, N. B. (2010). Reasoning up and down a food chain: Using an assessment framework to investigate students' middle knowledge. *Science Education*, *94*(2), 259–281. https://doi.org/10.1002/sce.20368

Gotwals, A. W., Songer, N. B., & Bullard, L. (2012). Assessing students' progressing abilities to construct scientific explanations. In A. C. Alonzo & A. W. Gotwals (Eds.), *Learning progressions in science: Current challenges and future directions* (pp. 183–210).

Harrison, A. G., & Treagust, D. F. (1996). Secondary students' mental models of atoms and molecules: Implications for teaching chemistry. *Science Education*, *80*(5), 509–534. https://doi.org/10.1002/(SICI)1098-237X(199609)80:5<509::AID-SCE2>3.0.CO;2-F

Helm, H. (1980). Misconceptions in physics amongst South African students. *Physics Education*, *15*(2), 92–97. https://www.doi.org/10.1088/0031-9120/15/2/308

Johannesson, P., & Perjons, E. (2014). *An introduction to design science*. Springer International Publishing. https://doi.org/10.1007/978-3-319-10632-8

Kahneman, D. (2011). *Thinking, fast and slow*. Macmillan.

Kane, M. T. (1992). An argument-based approach to validity. *Psychological Bulletin*, *112*(3), 527–535. https://doi.org/10.1037/0033-2909.112.3.527

Kane, M. T. (2013). The argument-based approach to validation. *School Psychology Review*, *42*(4), 448–457. https://doi.org/10.1080/02796015.2013.12087465

Kliewer, K., Langenfeld, T., & Thomas, J. (2018, October 4). *Observing cognitive processes and identifying reading difficulties through eye tracking analysis* [Poster presentation]. Education Technology and Computational Psychometrics Symposium, Iowa City, IA, United States. https://www.researchgate.net/publication/375086423_Observing_Cognitive_Processes_and_Identifying_Reading_Difficulties

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

Kolen, M. J., Hanson, B. A., & Brennan, R. L. (1992). Conditional standard errors of measurement for scale scores. *Journal of Educational Measurement*, *29*(4), 285–307. https://www.jstor.org/stable/1435086

Langenfeld, T., & Thomas, J. (2016, August 8). *WorkKeys Graphic Literacy: Applying design science to assessment development* [Paper presentation]. ACT Design Science Conference, Iowa City, IA, United States. https://www.doi.org/10.13140/RG.2.2.20874.58561

Langenfeld, T., Thomas, J., & Gao, X. (2019, April 6). *Principled assessment design: Applications and tools for assessment updates* [Paper presentation]. NCME Annual Conference, Toronto, Ontario, Canada.

Langenfeld, T., Thomas, J., Zhu, R., & Morris, C. A. (2020). Integrating multiple sources of validity evidence for an assessment-based cognitive model. *Journal of Educational Measurement*, *57*(2), 159–184. https://doi.org/10.1111/jedm.12245

Lazarus, S. S., Johnstone, C. J., Liu, K. K., Thurlow, M. L., Hinkle, A. R., & Burden, K. (2022). *An updated state guide to universally designed assessments* (NCEO Report No. 431). National Center on Educational Outcomes. https://files.eric.ed.gov/fulltext/ED620541.pdf

Lewis, J., Leach, J., & Wood-Robinson, C. (2000). Chromosomes: The missing link—young people's understanding of mitosis, meiosis, and fertilisation. *Journal of Biological Education*, *34*(4), 189–199. https://doi.org/10.1080/00219266.2000.9655717

Li, D., Kapoor, S., Arthur, A., Huang, C., Cho, Y., Qiu, C., Wang, H., (2025). *The enhanced ACT linking study report* (ACT Research Paper No. R2515). ACT. https://www.act.org/content/dam/act/unsecured/documents/R2515-The-Enhanced-ACT-Linking-Study-Report-2025-03.pdf

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. https://doi.org/10.1111/j.1745-3984.1995.tb00462.x

Louisiana Board of Regents. (2019). *Louisiana prospers: Driving our talent imperative*. https://files.eric.ed.gov/fulltext/ED609737.pdf

Mattern, K., Radunzel, J., & Westrick, P. (2015). *Development of STEM readiness benchmarks to assist educational and career decision making* (ACT Research Report No. 2015-3). ACT. https://www.act.org/content/dam/act/unsecured/documents/ACT_RR2015-3.pdf

Mislevy, R. J., Almond, R. G., & Lukas, J. F. (2003). *A brief introduction to evidence-centered design* (ETS Research Report No. RR-03-16). ETS. https://onlinelibrary.wiley.com/doi/epdf/10.1002/j.2333-8504.2003.tb01908.x

Mislevy, R. J., & Riconscente, M. M. (2006). Evidence-centered assessment design. In T. M. Haladyna & S. M. Downing (Eds.), *Handbook of test development* (1st ed., pp. 75–104). Routledge. https://doi.org/10.4324/9780203874776

National Governors Association Center for Best Practices & Council of Chief State School Officers. (2010). *Common core state standards*.

National Research Council, Division of Behavioral and Social Sciences and Education, Board on Science Education, & Committee on a Conceptual Framework for New K-12 Science Education Standards. (2012). *A framework for K-12 science education: Practices, crosscutting concepts, and core ideas*. National Academies Press. https://nap.nationalacademies.org/catalog/13165/a-framework-for-k-12-science-education-practices-crosscutting-concepts

National Research Council. (2014). *Developing assessments for the next generation science standards*. National Academies Press. https://doi.org/10.17226/18409

Nelson, M. J., Brown, J. I., & Denny, M. J. (1960). *The Nelson–Denny reading test: Vocabulary, comprehension, rate*. Houghton Mifflin.

NGSS Lead States. (2013). *Next generation science standards: For states, by states*. National Academies Press.

Ozcan, T., Yildirim, O., & Ozgur, S. (2012). Determining of the university freshmen students' misconceptions and alternative conceptions about mitosis and meiosis. *Procedia: Social and Behavioral Sciences*, *46*, 3677–3680. https://doi.org/10.1016/j.sbspro.2012.06.126

Radunzel, J., & Noble, J. (2012). *Tracking 2003 ACT-tested high school graduates: College readiness, enrollment, and long-term success* (ACT Research Report No. 2012-2). ACT. https://files.eric.ed.gov/fulltext/ED542012.pdf

Sanchez, E. (2023). *Evidence of grade inflation since 2010 in high school English, mathematics, social studies, and science courses* (ACT Research Report No. R2300). ACT. https://www.act.org/content/dam/act/secured/documents/Evidence-of-Grade-Inflation-in-English-Math-Social-Studies-and-Science.pdf

Spichtig, A. N., Hiebert, E. H., Vorstius, C., Pascoe, J. P., Pearson, P., & Radach, R. (2016). The decline of comprehension-based silent reading efficiency in the United States: A comparison of current data with performance in 1960. *Reading Research Quarterly*, *51*(2), 239–259. https://doi.org/10.1002/rrq.137

Stocking, M. L., & Lord, F. M. (1983). Developing a common metric in item response theory. *Applied Psychological Measurement*, *7*(2), 201–210. https://doi.org/10.1177/014662168300700208

Taylor, S. E. (1965). Eye movements in reading: Facts and fallacies. *American Educational Research Journal*, *2*(4), 187–202. https://doi.org/10.2307/1161646

Thomas, J. (2020, April). *Using automated scoring to monitor and improve the assessment system* [Paper presentation]. NCME Annual Conference, San Francisco, CA, United States.

Thomas, J., Johanningmeier, S., Dreyer, S., & Zhang, C. (2025, April 26). *Using think aloud, surveys, and interviews to validate cognitive processes for an assessment* [Paper presentation]. NCME Annual Conference, Denver, CO, United States.

Thomas, J. & Langenfeld, T. (2017a, April 29). *Analyzing think-aloud and eye-tracking data to support score interpretations* [Poster presentation]. NCME Annual Conference, San Antonio, TX, United States. https://www.researchgate.net/publication/375085755_Analyzing_Think-aloud_and_Eye-tracking_Data_to_Support_Score_Interpretations

Thomas, J. & Langenfeld, T. (2017b, November 15). *Using multiple lines of evidence to validate a construct and assessment* [Poster presentation]. Education Technology and Computational Psychometrics Symposium, Iowa City, IA, United States. https://www.doi.org/10.13140/RG.2.2.11353.70240

Thomas, J. & Langenfeld, T. (2018). *Using eye tracking and pupillometry to validate a graphic literacy assessment* [Paper presentation]. International Test Commission Conference, Montreal, Québec, Canada.

Thomas, J., & Victoria, P. (2023, April 14). *Using eye tracking to validate cognitive processes in high stakes assessments* [Paper presentation]. NCME Annual Conference, Chicago, IL, United States.

Tsui, C. Y., & Treagust, D. F. (2007). Understanding genetics: Analysis of secondary students' conceptual status. *Journal of Research in Science Teaching*, *44*(2), 205–235. https://doi.org/10.1002/tea.20116

van Aken, J. E., & Romme, A. G. L. (2012). A design science approach to evidence-based management. In D. Rousseau (Ed.), *The Oxford handbook of evidence-based management* (pp.43–57). https://doi.org/10.1093/oxfordhb/9780199763986.013.0003

Wei, H., Cromwell, A. M., & McClarty, K. L. (2016). Career readiness: An analysis of text complexity for occupational reading materials. *The Journal of Educational Research*, *109*(3), 266–274. https://doi.org/10.1080/00220671.2014.945149

Westrick, P. (2017). *Precollege profiles of high-performing STEM majors* (ACT Research Report No. 2017-2). ACT. https://www.act.org/content/dam/act/unsecured/documents/R1628-high-performing-stem-majors-2017-05.pdf

Wise, S. L., & DeMars, C. E. (2010). Examinee noneffort and the validity of program assessment results. *Educational Assessment*, *15*(1), 27–41. https://doi.org/10.1080/10627191003673216

Wise, S. L., & Smith, L. F. (2011). A model of examinee test-taking effort. In J. A. Bovaird, K. F. Geisinger, & C. W. Buckendahl (Eds.), *High-stakes testing in education: Science and practice in K–12 settings* (pp. 139–153). American Psychological Association. https://doi.org/10.1037/12330-009

ABOUT ACT

ACT is transforming college and career readiness pathways so that everyone can discover and fulfill their potential. Grounded in more than 65 years of research, ACT's learning resources, assessments, research, and work-ready credentials are trusted by students, job seekers, educators, schools, government agencies, and employers in the U.S. and around the world to help people achieve their education and career goals at every stage of life. Visit us at www.act.org.