



WISCONSIN
DEPARTMENT OF

Public Instruction

ACT[®] **Aspire**[®]

Wisconsin 2021 Technical Report

December 10, 2021

Table of Contents

Chapter 1: Introduction.....	4
1.1 Purpose and Intended Audience.....	4
1.2 Additional ACT Aspire Documentation.....	4
1.3 Proposed Uses of ACT Aspire Test Scores	8
Chapter 2: Assessment Participation	9
2.1 Spring 2021 Student Participation	9
2.2 Spring 2021 Test Administrations.....	11
2.2.1 Test mode and form accommodations	11
2.2.2 Test dates	14
Chapter 3: Achievement Summary and Trends	15
3.1 Spring 2021 Achievement Summary.....	15
3.1.1 Comparison of Mean Scores to National Norms	15
3.1.2 Comparison of ACT Readiness Benchmarks and Levels to National Norms.....	17
3.2 Achievement Trends From 2015 to 2021	21
Chapter 4: Technical Characteristics of the Tests.....	24
4.1 Test Equating.....	24
4.1.1 Equating Data Collection Design	24
4.1.2 Equating Methodology	25
4.2 Reliability and Measurement Error	25
4.2.1 Raw Score Reliability and SEM for Subject Tests	26
4.2.2 Scale Score Reliability and SEM for Subject Tests, Composite, ELA, and STEM Scores	29
4.2.3 Raw Score Reliability and SEM for Reporting Category Scores	30
4.2.4 Classification Consistency for the ACT Readiness Benchmarks and the ACT Readiness Levels	33
Chapter 5: Validity Evidence	34
5.1 Content-Oriented Evidence	34
5.2 Relationships With Conceptually Related Constructs: Correlations of ACT Aspire and 11th Grade ACT Test Scores	35

5.3 Relationships With Criteria: Prediction of High School Course Grades and Academic Rigor..... 36

5.4 Differential Item Functioning..... 38

5.5 Depth of Knowledge Analysis..... 40

Chapter 6: Growth Summary..... 42

6.1 Comparison of Mean Growth Scores to National Growth Norms 42

6.2 ACT Readiness Level Transitions..... 44

6.3 Relating ACT Aspire Interim Scores to Summative SGPs..... 45

References..... 47

Appendix..... 49

Chapter 1: Introduction

1.1 Purpose and Intended Audience

ACT Aspire[®] Summative Assessments have been administered to Wisconsin's students in 9th and 10th grade each spring since 2015 (except for spring 2020 when testing was cancelled due to the COVID-19 pandemic). In this report, we provide information documenting the technical quality of the tests, results of testing in Wisconsin, and evidence supporting intended uses of ACT Aspire test scores. The report focuses on the spring 2021 test administration, but also includes analyses of data from prior years, as well as analyses of achievement trends and student growth across years. The report is scheduled to be updated each year to reflect data from the most recent test administration and additional research pertaining to the use of ACT Aspire in Wisconsin.

The intended audience for this report is individuals seeking information about the technical quality of Wisconsin's ACT Aspire state assessment program. This could include educational professionals at all levels, state policymakers, and the general public. Some of the information presented in the report is of a technical nature geared towards individuals with training or experience in educational measurement or statistics.

1.2 Additional ACT Aspire Documentation

[The ACT Aspire Summative Technical Manual](#) (ACT, 2020a, hereafter referred to as *the technical manual*) presents information from a national point of view, and it contains detailed information about the ACT Aspire Summative Assessments, describes various content and psychometric aspects of the assessments, and documents a collection of evidence supporting interpretations of ACT Aspire test scores. The information contained in this report is intended to supplement the information in the technical manual, with a focus on Wisconsin-specific evidence. The technical manual and this report can be used together to assess the appropriateness of using ACT Aspire test scores for different purposes in Wisconsin. This report does not duplicate content from the technical manual. Topics that are covered in the technical manual but not this technical report include:

- Test development procedures
- Test specifications
- Content standards and performance level descriptors, including:
 - ACT College and Career Readiness Standards
 - ACT Aspire Grade Level Targets for English, reading, writing, and mathematics

- ACT Aspire Performance Level Descriptors
 - Scoring procedures
 - Accessibility support system and accommodations
 - Test administration procedures
 - Test and information security
 - Interpretation of scores, readiness benchmarks, and progress indicators
 - Scaling procedures

The ACT Aspire Summative Technical Manual and other forms of test documentation are available [here](#). The documentation includes FAQs, user guides, accessibility and accommodations guides, item exemplars, training resources, and guides for technology requirements and system set-up.

To help readers navigate the technical manual and this Wisconsin-specific technical report, we next describe how each chapter of this report relates to content in the technical manual. The technical manual has 14 chapters, and this technical report has six chapters. In Appendix [Table A.1](#), we provide a brief description of each chapter of the technical manual. The six chapters of this technical report relate to content in the technical manual as follows.

[Chapter 1: Introduction](#). This chapter describes the purpose, scope, and organization of the technical report and lists proposed uses of ACT Aspire test scores in Wisconsin. Chapter 1 answers questions such as:

- What is the purpose and intended audience of the technical report?
- What topics are covered in the technical report?
- What topics are covered in the ACT Aspire technical manual?
- What are the proposed uses of ACT Aspire test scores?

Proposed uses and interpretations of ACT Aspire test scores are also discussed in Chapters 1 and 12 of the technical manual.

[Chapter 2: Assessment Participation](#). This chapter documents characteristics of Wisconsin students who participated in testing and summarizes information on test administrations (including use of testing accommodations). Chapter 2 answers questions such as:

- How many students took ACT Aspire tests in spring 2021?
- How many students from each student group participated in testing?

- What types of testing accommodations were utilized?
- What were the most popular test dates?

Information on frequency of use of test forms with accommodations is also provided in Chapter 5 of the technical manual.

Chapter 3: Achievement Summary and Trends. This chapter presents summary statistics on ACT Aspire scores and readiness levels for the 2021 ACT Aspire administration in Wisconsin, with comparisons to national norms and to prior years. It answers questions such as:

- What were the average test scores in spring 2021?
- What percentage of Wisconsin students are on target for college and readiness?
- How did Wisconsin's scores compare to national averages?
- Have scores improved since the beginning of the assessment program in 2015?

Estimates of national norms for ACT Aspire are presented in Chapter 8 of the technical manual.

Chapter 4: Technical Characteristics of the Tests. This chapter presents information on procedures for equating tests. It also presents estimates of reliability, standard error of measurement, and classification consistency for the 2021 ACT Aspire administration in Wisconsin. It answers questions such as:

- What procedures were used to ensure that ACT Aspire Summative test scores are comparable across different years and test forms?
- How reliable are ACT Aspire Summative test scores?
- Are the test scores reliable for different groups of students?
- To what extent are students classified consistently with respect to being on target for college and career readiness?

Similar information for all ACT Aspire Summative users is provided in Chapters 10 and 11 of the technical manual.

Chapter 5: Validity Evidence. This chapter presents Wisconsin-specific criterion-related validity evidence, differential item functioning (DIF) analysis, and analyses related to depth of knowledge (DOK) levels. Chapter 5 answers questions such as:

- How well do ACT Aspire scores predict ACT test scores?

- What is the relationship between ACT Aspire scores and performance in high school courses?
- Do ACT Aspire test items function similarly for different groups of examinees?

Chapters 1 and 12 of the technical manual include general discussion of alignment and content-related validity evidence. Criterion-related validity evidence (not specific to Wisconsin examinees) is presented in Chapter 12 of the technical manual. A national DIF analysis (not specific to Wisconsin examinees) is presented in Chapter 13 of the technical manual.

Chapter 6: Growth Summary. This chapter summarizes Wisconsin-specific growth data, with comparisons to national norms. It answers questions such as:

- How does ACT Aspire support interpretations of academic growth?
- How much do student's scores typically increase in one year?
- How do Wisconsin's growth scores compare to national averages?
- How do ACT Aspire Interim test scores relate to growth in Summative scores?

Chapter 14 of the technical manual presents national data summarizing gain scores and student growth percentiles.

1.3 Proposed Uses of ACT Aspire Test Scores

One purpose of this report is to provide evidence supporting proposed uses of ACT Aspire test scores in Wisconsin, which include:

- To measure progress toward meeting the Wisconsin Academic Standards for high school in English Language Arts (English, reading, and writing), mathematics, and science
- To determine if Wisconsin students are on target for college and career readiness
- To assess how well Wisconsin schools and districts are preparing students for college and careers by meeting grade level standards (school and district accountability)
- To inform students' readiness for advanced high school coursework
- To understand student and group performance relative to national norms

For example, the use of ACT Aspire scores for accountability is supported by content evidence, studies examining alignment of ACT Aspire with the state's academic standards, evidence from standard setting (including development of the ACT Readiness Benchmarks), and additional evidence presented in this technical report or the technical manual. Test users may develop additional uses that are not listed here and may need to collect additional evidence to support them.

ACT Aspire includes Summative, Interim, and Classroom Assessments. While all three types of assessments have been used in Wisconsin, only the Summative Assessments for 9th and 10th grade are used statewide. This report focuses on the Summative Assessments.

Chapter 2: Assessment Participation

ACT Aspire Summative Assessments are intended for students in grades 3–10 and are designed to measure key college and career readiness constructs in a way that recognizes that skills are not isolated to specific grades, but rather should progress across grades. Assessments are available for five domains (subjects): English, mathematics, reading, science, and writing. In Wisconsin, students in both 9th and 10th grades are assessed using test forms in all five domains developed for early high school students.

In this chapter, we document characteristics of Wisconsin students who participated in testing in spring 2021 and provide information on test administrations, including mode of testing, accommodations, and dates of testing.

2.1 Spring 2021 Student Participation

Table 2.1 provides the number of students for whom scores were reported in spring 2021. Overall, 109,725 students had at least one score reported. In addition to the total number of students, counts are also provided for each subject area. Among students with at least one reported score, the overwhelming majority—95.1%—had scores reported for all five subject areas. The subject with the highest rate of missing scores was writing, with 3.4% of all students missing a writing score.

Table 2.1. Number of Students Tested in Spring 2021, by Grade Level and Subject

Grade level	Total	Subject				
		English	Mathematics	Reading	Science	Writing
9	55,722	55,097	55,191	55,125	54,820	53,835
10	54,003	53,338	53,383	53,388	53,126	52,156
Total	109,725	108,435	108,574	108,513	107,946	105,991

The groups that are included for accountability reporting include the following major racial and ethnic student groups and educationally at-risk student groups: African American, Hispanic, White, Economically Disadvantaged, Students with Disabilities, and English Learners. The Students with Disabilities group includes students with an individualized education plan (IEP), Section 504 plan, or other accommodations plan.

Students are assigned to a race/ethnicity category based on the following rules, applied sequentially:

1. If the student is of Hispanic ethnicity, they are categorized as *Hispanic*.

2. Otherwise, if a student is assigned “Yes” to two or more race indicators, they are categorized as *Two or more races*.
3. Otherwise, if a student is assigned “Yes” to one race indicator, they are categorized as that race (*African American, Asian, Native American, Native Hawaiian/Other Pacific Islander, or White*).
4. Otherwise, the student is categorized as *Missing race/ethnicity*.

Note that group membership is determined from data in the ACT Aspire Student Performance File and may differ from other state records.

In Table 2.2, we provide the number and percentage of students in each group by grade level. In addition to the major student groups identified for accountability reporting, we provide counts for additional groups including gender and smaller racial/ethnic groups.

Table 2.2. Students Tested in Spring 2021, by Group and Grade Level

Group	9th Grade		10th Grade	
	Count	%	Count	%
Gender				
Female	26,905	48.3%	26,246	48.6%
Male	28,652	51.4%	27,602	51.1%
Missing or another gender	165	0.3%	155	0.3%
Race/ethnicity				
African American	2,997	5.4%	2,660	4.9%
Asian	1,979	3.6%	1,867	3.5%
Hispanic	6,720	12.1%	5,974	11.1%
Native American	543	1.0%	516	1.0%
Native Hawaiian/OPI	49	0.1%	47	0.1%
Two or more races	2,013	3.6%	1,783	3.3%
White	40,232	72.2%	39,936	74.0%
Missing race/ethnicity	1,189	2.1%	1,220	2.3%
Econ. Disadvantaged	17,648	31.7%	15,623	28.9%
Students with Disabilities	6,251	11.2%	5,984	11.1%
English Language Learner	2,184	3.9%	1,796	3.3%

Note. Bold indicates accountability reporting group. OPI = Other Pacific Islander; Econ. = Economically.

2.2 Spring 2021 Test Administrations

2.2.1 Test mode and form accommodations

A variety of accessibility supports, tools, and options are available to ensure that the ACT Aspire assessment is administered in an accessible and standardized way. Multiple levels of accessibility are available, including default embedded tools, open-access tools, and full accommodations to allow students with disabilities to participate in testing. Tools and accommodations are available for both online and paper tests. Some, but not all, tools or accommodations require a different type of test form.

For more information about accessibility supports, tools, and options for ACT Aspire Summative testing, please see [ACT Aspire Accessibility Supports Guide](#).

Tests were primarily administered in an online mode: Across all subject areas and grade levels, only 178 tests out of 539,459 were administered using paper.

Table 2.3 (9th grade) and Table 2.4 (10th grade) report the frequency of each type of accommodation provided, by subject area. The tables also provide the number of students receiving at least one accommodation and the number of students who tested without accommodations. The most common types of accommodations included extra time (n = 41,769 tests), special seating or grouping (n = 33,552 tests), supervised breaks (n = 17,092 tests), and English text-to-speech audio (n = 16,299 tests).

Table 2.3. 9th Grade Accommodations, by Subject

Group/Accommodation Type	English	Math	Reading	Science	Writing
Total Number Tested	55,097	55,191	55,125	54,820	53,835
No Accommodation	48,679	48,569	48,690	48,302	47,703
Accommodated	6,418	6,622	6,435	6,518	6,132
Abacus	0	19	0	0	0
American Sign Language Directions Only	11	10	11	10	*
American Sign Language Full Translation	0	*	0	*	*
Audio Environment	103	105	105	103	100
Braille Contracted Unified English UEB	*	*	*	*	*
Breaks	*	*	*	*	*
Breaks Supervised Each Day	1,797	1,810	1,790	1,780	1,666
Cued Speech	0	*	0	*	*
Dictate Responses	98	98	101	98	86
Electronic Spell Checker	0	189	0	189	180
English TTS Audio	0	2,891	0	2,871	2,765
English TTS Audio Orienting Description	0	*	0	*	*
Extra Time 1.5	3,252	2,396	3,265	2,341	2,196
Extra Time 2.0	1,175	685	1,170	669	605
Extra Time 2.5	30	*	29	*	*
Extra Time 3.0	1,187	107	1,249	102	97
Extra Time 4.0	42	10	42	10	*
Home Administration	*	*	*	*	*
Human Reader English	0	73	0	68	62
Human Reader English Orienting Desc.	0	*	0	*	*
Individual Administration	151	149	150	145	134
Keyboard AAC Local Print	10	*	10	*	*
Large Print	*	*	*	*	*
Location for movement	252	250	249	244	240
Other Setting	1,199	1,204	1,203	1,199	1,130
Physical Motor Equipment	*	*	*	*	*
Respond in Test Booklet / Separate Paper	*	*	*	*	*
Signed Exact English Full Translation	0	*	0	*	*
Spanish Text Audio	0	73	0	73	35
Special Seating or Grouping	3,421	3,442	3,425	3,374	3,205
Translated Test Directions Only	95	105	96	106	75
Visual Environment	20	20	20	19	19
Word 2 Word Dictionary	0	250	0	245	214

* Indicates that count is greater than 0 and less than 10

Table 2.4. 10th Grade Accommodations, by Subject

Group/Accommodation Type	English	Math	Reading	Science	Writing
Total Number Tested	53,338	53,383	53,388	53,126	52,156
No Accommodation	47,055	46,954	47,081	46,779	46,144
Accommodated	6,283	6,429	6,307	6,347	6,012
Abacus	0	29	0	0	0
American Sign Language Directions Only	*	*	*	*	*
American Sign Language Full Translation	0	*	0	*	*
Audio Environment	109	108	109	107	104
Braille Contracted Unified English UEB	*	*	*	*	*
Breaks	*	*	*	*	*
Breaks Supervised Each Day	1,679	1,676	1,673	1,656	1,565
Custom Masking	*	*	*	*	*
Dictate Responses	73	73	75	72	70
Electronic Spell Checker	0	158	0	159	154
English TTS Audio	0	2,633	0	2,601	2,538
Extra Time 1.5	3,321	2,560	3,310	2,528	2,365
Extra Time 2.0	1,159	668	1,169	658	615
Extra Time 2.5	25	13	25	13	11
Extra Time 3.0	1,113	92	1,176	92	88
Extra Time 4.0	22	11	22	11	11
Home Administration	*	*	*	*	*
Human Reader English	0	59	0	56	49
Human Reader English Orienting Desc.	0	0	0	*	*
Individual Administration	146	144	144	143	135
Keyboard AAC Local Print	12	12	12	10	11
Location for movement	194	197	194	194	182
Other Setting	1,160	1,157	1,157	1,145	1,110
Physical Motor Equipment	*	*	*	*	*
Respond in Test Booklet / Separate Paper	11	13	11	13	13
Signed Exact English Directions Only	*	*	*	*	*
Spanish Text Audio	0	83	0	80	39
Special Seating or Grouping	3,386	3,386	3,387	3,334	3,192
Translated Test Directions Only	108	120	114	118	81
Visual Environment	13	13	13	13	13
Word 2 Word Dictionary	0	247	0	246	210

* Indicates that count is greater than 0 and less than 10

2.2.2 Test dates

Figure 2.1 provides the number of ACT Aspire tests taken by date. The planned testing window was April 5th to May 5th. Because of pandemic-related testing disruptions, the testing window was extended for two weeks to May 19. The vast majority of tests were taken prior to the week of May 17–21. Most tests were taken the week of April 19–23, followed by the week of April 12–16 and the week of April 26–30. Note that for 119 tests, the test dates extend beyond the last day of testing (May 19th). These tests were administered on paper and the test dates reflect when the tests were scored, not taken.

Among students who took all five tests, most (53.6%) took all tests on the same day, while 30.5% tested over the course of two days, 7.5% tested over three days, 4.4% tested over four days, and 4.1% tested over five days.

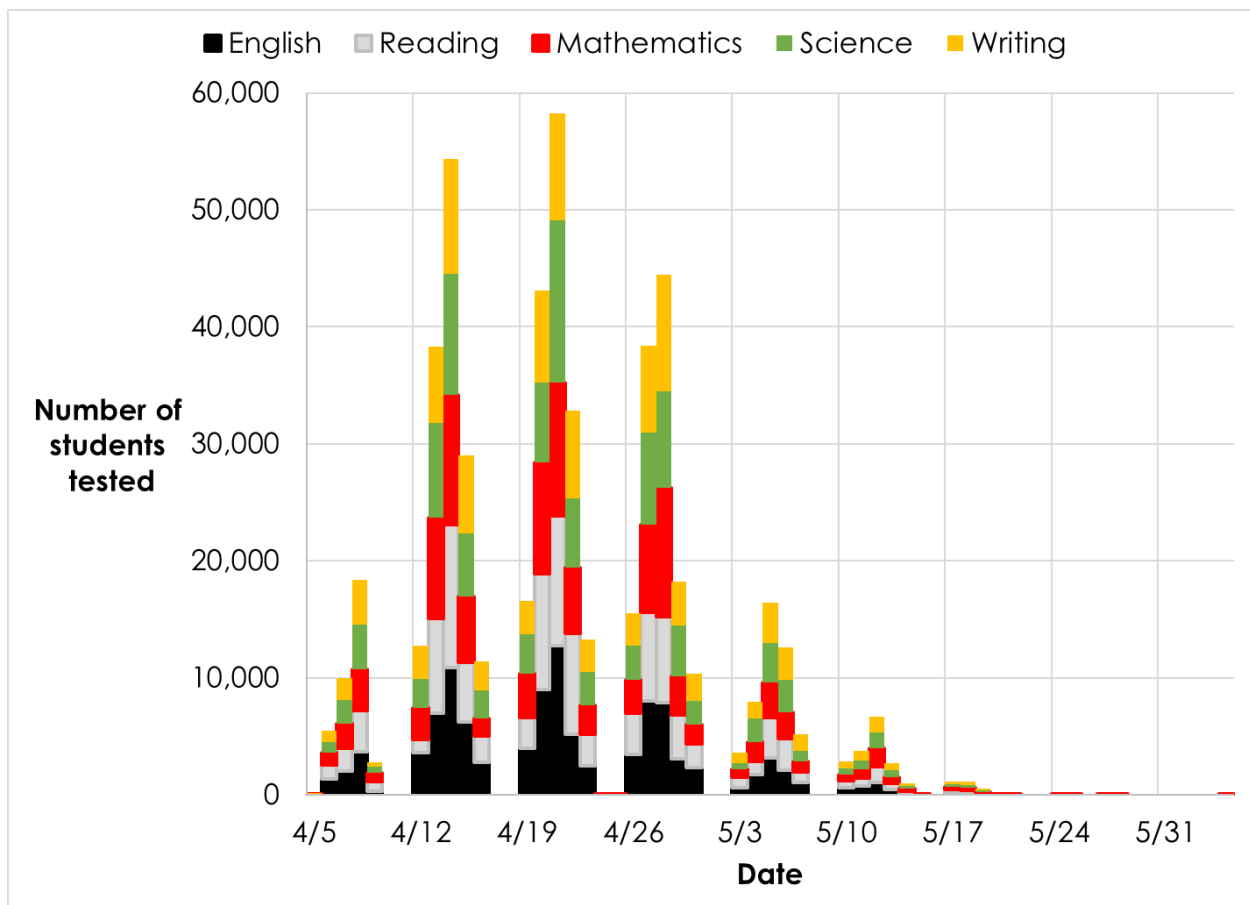


Figure 2.1. Number of students tested, by date and subject

Chapter 3: Achievement Summary and Trends

This chapter presents summary statistics on ACT Aspire scores and readiness levels for the 2021 ACT Aspire administration in Wisconsin, with comparisons to national norms and to prior years. The summary statistics and trend analyses are based on subject scores (English, math, reading, and science) and combined scores (ELA, STEM, and Composite), and the corresponding ACT Readiness Benchmarks and ACT Readiness Levels.

As discussed in [Chapter 2](#), Wisconsin administers tests for all five subject areas: English, mathematics, reading, science, and writing. Scale scores are generated for each of the five tests and are reported for four of the five tests (all but writing). In addition to scale scores for four subjects, three combined scores are reported:

- The **ELA score** is the average of the English, reading, and writing scale scores.
- The **STEM score** is the average of the mathematics and science scale scores.
- The **Composite score** is the average of the English, mathematics, reading, and science scale scores.

3.1 Spring 2021 Achievement Summary

3.1.1 Comparison of Mean Scores to National Norms

We begin by examining the mean scores for Wisconsin from spring 2021 and comparing them to the latest national norms. This analysis addresses the question: “How does the academic achievement of Wisconsin students compare to students across the nation?”

ACT normally conducts a national norming study each year to produce updated estimates of percentile ranks and mean scores for each reported scale score, by grade level. The norms used for this report are based on the 2019 norming study, which used data from spring 2017, spring 2018, and spring 2019. The norming study was not done in 2020 because ACT Aspire Summative Assessments were not administered in spring 2020 due to the COVID-19 pandemic.

The norming samples include students from both public and nonpublic schools, including those from Wisconsin. ACT Aspire-tested students are not necessarily representative of the national population of students in grades 3–10. To support interpretations of nationally representative norms, weights are assigned to the samples so that they are more representative of the national population on school affiliation

(public vs. non-public) and, among public schools, race/ethnicity and academic achievement. More details on the norming study methodology are provided in Chapter 8 of the technical manual.

For each subject and grade level, the mean and standard deviation of scale scores are provided in Table 3.1. Statistics are provided for the spring 2021 Wisconsin administration and the 2019 national norms.

The ***d*** statistic of Table 3.1 measures the difference between each Wisconsin mean score and the estimated national mean score. It is calculated as the difference between the two means, divided by the estimated national standard deviation. Positive values of ***d*** indicate that the Wisconsin mean score is larger than the estimated national mean score.

Across the 14 combinations of subject/score and grade level, the Wisconsin mean was less than the national mean in 12 cases and greater than the national mean in two cases. Relative to the national norms, Wisconsin scores were lowest for 9th grade ELA (***d*** = -0.11) and 10th grade English (***d*** = -0.10) and highest for 10th grade STEM (***d*** = 0.00) and 10th grade science (***d*** = 0.02).

In general, the mean scores for Wisconsin are similar to the national norms. When the Wisconsin mean scores are different than the national mean scores by 0.10 standard deviations and larger ($|\mathbf{d}| \geq 0.10$), we consider it a substantive difference. Using this rule, the mean scores for Wisconsin are substantively lower than the national norms for two of the 14 combinations of subject/score and grade level. For the other 12 cases, the mean scores for Wisconsin are not substantively different than the national norms.

Wisconsin's 2021 test scores were impacted by the learning disruptions caused by the COVID-19 pandemic, whereas the 2019 national norms were not. In later sections of this report, we examine Wisconsin's score trends over six years to better understand possible impacts of the pandemic.

In addition to the pandemic, there could be many other reasons for differences in Wisconsin performance across subjects and grade levels, relative to national norms. In this report, we do not attempt to explain the differences.

Table 3.1. 2021 Scale Score Summary Statistics, by Subject and Grade Level

Subject	Grade level	Wisconsin 2021		National norms 2019		<i>d</i>
		Mean	SD	Mean	SD	
English	9	428.0	9.7	428.8	10.4	-0.08
	10	429.9	10.1	431.1	10.9	-0.10
Mathematics	9	425.0	8.5	425.2	8.6	-0.02
	10	427.0	9.0	427.0	9.4	-0.01
Reading	9	421.9	7.9	422.5	7.9	-0.08
	10	423.2	8.1	423.5	8.2	-0.03
Science	9	425.6	8.8	425.7	8.8	-0.01
	10	427.4	9.2	427.2	9.5	0.02
Composite	9	425.4	7.8	425.8	8.1	-0.05
	10	427.1	8.2	427.4	8.7	-0.03
ELA	9	425.5	7.2	426.3	7.5	-0.11
	10	427.0	7.4	427.7	7.8	-0.09
STEM	9	425.6	8.2	425.8	8.2	-0.02
	10	427.5	8.7	427.5	9.0	0.00

Note. SD = standard deviation; *d* = (Wisconsin mean – National mean) / National SD.

3.1.2 Comparison of ACT Readiness Benchmarks and Levels to National Norms

The ACT College Readiness Benchmarks are the scores on the ACT® test associated with a 50% chance of earning a B or higher grade in common first-year credit-bearing college courses. ACT College Readiness Benchmarks have been developed for English (18), mathematics (22), reading (22), science (23), ELA (20), and STEM (26).

More information on the ACT College Readiness Benchmarks is [available here](#).

The ACT Readiness Benchmarks are the minimum ACT Aspire scores (grades 3–10) for which students are on target to meet or exceed the ACT College Readiness Benchmarks when they are in the 11th grade. Thus, the ACT Readiness Benchmarks can be interpreted as the minimum scores associated with being on-target for college readiness. The ACT Readiness Benchmarks are reported for each subject and grade level.

For English, mathematics, reading, science, and ELA scores, the ACT Readiness Levels are used to further classify student achievement as:

- **In Need of Support** if the score is greater than two standard errors of measurement (SEM) below the ACT Readiness Benchmark.
- **Close** if the score is below the ACT Readiness Benchmark, but within two SEMs of the Benchmark.
- **Ready** if the score is equal to the ACT Readiness Benchmark or above and within two SEMs of the Benchmark.
- **Exceeding** if the score is greater than two SEMs above the ACT Readiness Benchmark.

Similar to the analysis of mean scores, the percentage of Wisconsin students scoring at each ACT Readiness Level can be compared to the national norms (Table 3.2). For ELA and STEM, note that the table only provides the percentage of students who met the ACT Readiness Benchmark because ACT Readiness Levels have not been set for ELA and STEM scores.

Similar to the national norms, students in Wisconsin are more likely to meet the English Benchmark and least likely to meet the STEM Benchmark. The ACT STEM Benchmark was derived using college courses most commonly taken by students in STEM-related majors; the courses included Calculus, Biology, Chemistry, and Engineering. Because STEM-related coursework tends to be more difficult, the resulting Benchmark is substantially higher than the Benchmarks for other subjects.

The “Diff. Bench” column of Table 3.2 shows the difference in Benchmark attainment rates for Wisconsin students relative to the national norms, with positive values indicating that Wisconsin students outperformed the national norm. Across the 12 combinations of subjects and grade levels, Benchmark attainment for Wisconsin students was lower than that of the national norm for nine cases, the same in two cases, and greater for one case. Relative to the norms, Wisconsin Benchmark attainment was lowest for 9th and 10th grade ELA (-5% and -6%) and 9th grade reading (-5%) and highest for 9th grade mathematics (+2%).

Table 3.2. Percentage Meeting ACT Readiness Levels and Benchmarks, by Subject and Grade Level

Subject	Grade level	2021 Wisconsin					2019 National Norms					Diff. Bench.
		INS	CL	RD	EX	Met Bench.	INS	CL	RD	EX	Met Bench.	
English	9	18	22	27	34	60	18	21	23	38	61	-1
	10	19	21	26	34	60	19	19	22	41	62	-2
Mathematics	9	39	22	21	19	40	38	24	19	19	38	2
	10	43	23	21	13	34	46	20	18	16	34	0
Reading	9	34	28	20	18	38	34	23	25	18	43	-5
	10	38	27	24	10	34	40	23	26	11	37	-3
Science	9	43	22	21	15	35	42	22	19	16	35	0
	10	43	23	18	16	34	44	20	20	17	36	-2
ELA	9	38	20	20	22	42	35	17	19	28	47	-5
	10	39	20	21	20	41	37	17	20	26	46	-6
STEM	9					16					17	-1
	10					17					19	-2

Note. INS = In Need of Support; CL = Close; RD = Ready; EX = Exceeding; Bench. = Benchmark.

The percentage of students meeting the ACT Readiness Benchmarks is also presented by student group (Table 3.3). Generally, Benchmark attainment was lowest for the English Language Learner group, followed by the Students with Disabilities, African American, and Economically Disadvantaged groups.

For all groups, Benchmark attainment in mathematics and reading decreased from 9th to 10th grade. This is consistent with the national norms, where the percentage meeting the mathematics Benchmark is 38% for 9th grade and 34% for 10th grade, and the percentage meeting the reading Benchmark is 43% for 9th grade and 37% for 10th grade (ACT, 2020b).

Table 3.3. 2021 Percentage Meeting ACT Readiness Benchmark, by Group, Subject, and Grade Level

Subject	Grade level	Group							
		Female	Male	African American	Hispanic	White	Econ. Dis.	SWD	ELL
English	9	68	53	27	41	66	43	21	14
	10	68	53	27	41	65	42	21	10
Mathematics	9	40	39	8	18	46	20	9	4
	10	34	34	6	14	39	16	7	2
Reading	9	43	33	12	22	43	22	10	4
	10	39	30	10	19	38	19	9	3
Science	9	37	33	7	16	40	19	9	3
	10	36	32	8	15	39	18	9	2
ELA	9	51	34	13	23	47	24	9	4
	10	49	33	13	22	45	22	9	2
STEM	9	16	16	2	4	18	6	3	0
	10	17	18	2	6	20	6	3	0

Note. Econ. Dis. = Economically Disadvantaged; SWD = Students with Disabilities; ELL = English Language Learner.

3.2 Achievement Trends From 2015 to 2021

ACT Aspire has been used as a state assessment in Wisconsin since spring 2015. In this section, we address the question, “How has the achievement of Wisconsin students changed over the first six years of the assessment program”?

For each subject and grade level, appendix [Table A.2](#) presents summary measures of achievement for each year from 2015 to 2021 (excluding 2020 when testing was cancelled). The summary measures include the mean score and the percentage meeting the ACT Readiness Benchmark.

Across 12 subject and grade level combinations (not including Composite), mean scores increased from 2015 to 2019 in six cases and decreased in six cases. The largest improvements in mean scores were observed for 9th grade mathematics (+0.8 score points) and 10th grade mathematics (+0.7 score points). The largest decreases in mean scores were observed for 9th grade reading (-1.0 score points) and 10th grade science (-0.6 score points).

Similarly, the percentage meeting the ACT Readiness Benchmark increased from 2015 to 2019 in six cases and decreased in six cases. The largest improvements in Benchmark attainment were observed for 9th and 10th grade mathematics (+5%). The largest decrease in Benchmark attainment was observed for 10th grade reading (-5%).

From 2019 to 2021, mean scores increased in 4 cases and decreased in 8 cases, with the largest decreases observed for grade 9 mathematics (-1.2 score points) and 10th grade mathematics (-0.9 score points). From [Table A.2](#), we see that the number of students tested decreased from 2019 to 2021 (e.g., from 62,926 to 55,097 for 9th grade English), suggesting that assessment participation rates declined from 2019 to 2021. Therefore, the comparison of test scores from 2019 to 2021 is confounded by differences in participation.

For 2021, the overall (including both 9th and 10th grade) percentage meeting the Benchmark (scoring at the Ready or Exceeding levels) was 60% for English, 36% for reading, 42% for ELA, 37% for mathematics, 35% for science, and 17% for STEM. Trends in Benchmark and Readiness Level attainment from 2015 to 2021 are summarized in Figures 3.1 through 3.4 for English, math, reading, and science. The figures show the percentage of students in 9th and 10th grade performing at each ACT Readiness Level.

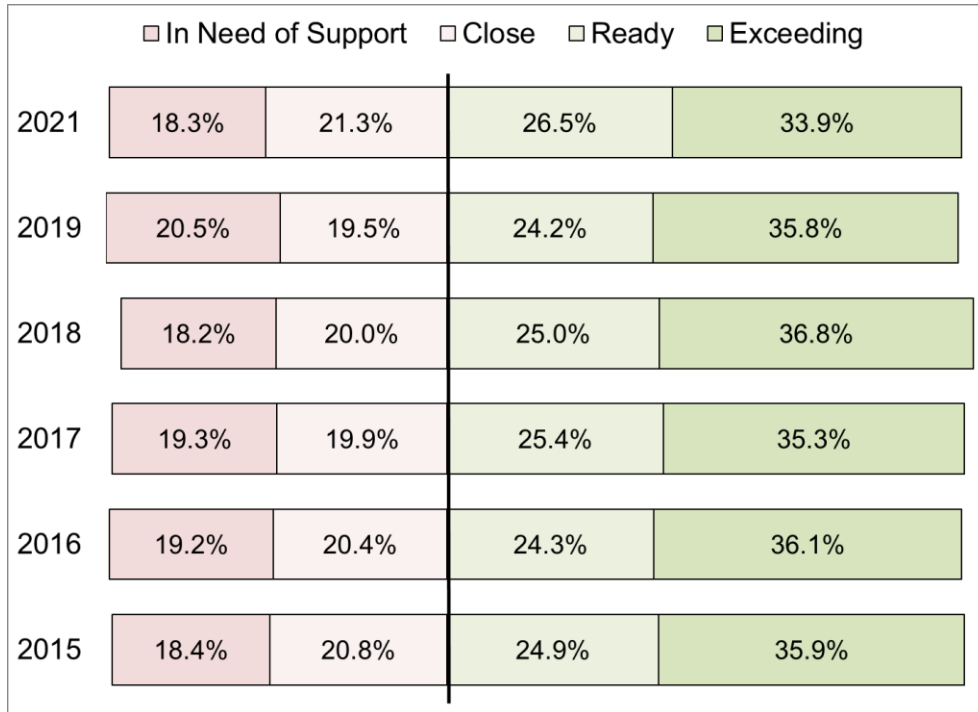


Figure 3.1. Overall percentage meeting ACT Readiness Levels for English, by year

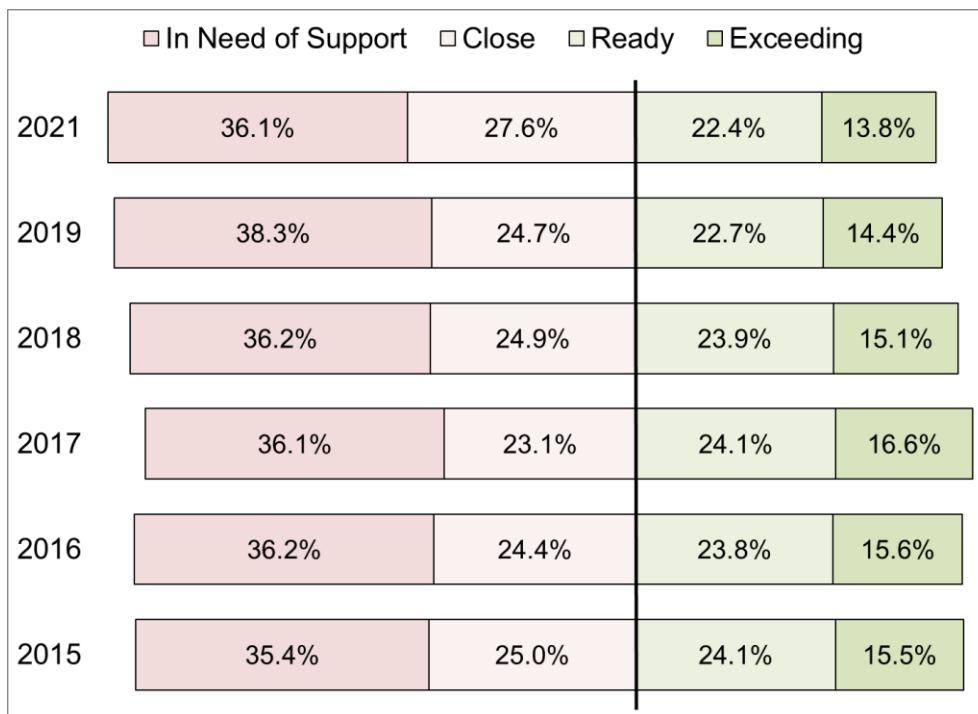


Figure 3.2. Overall percentage meeting ACT Readiness Levels for reading, by year

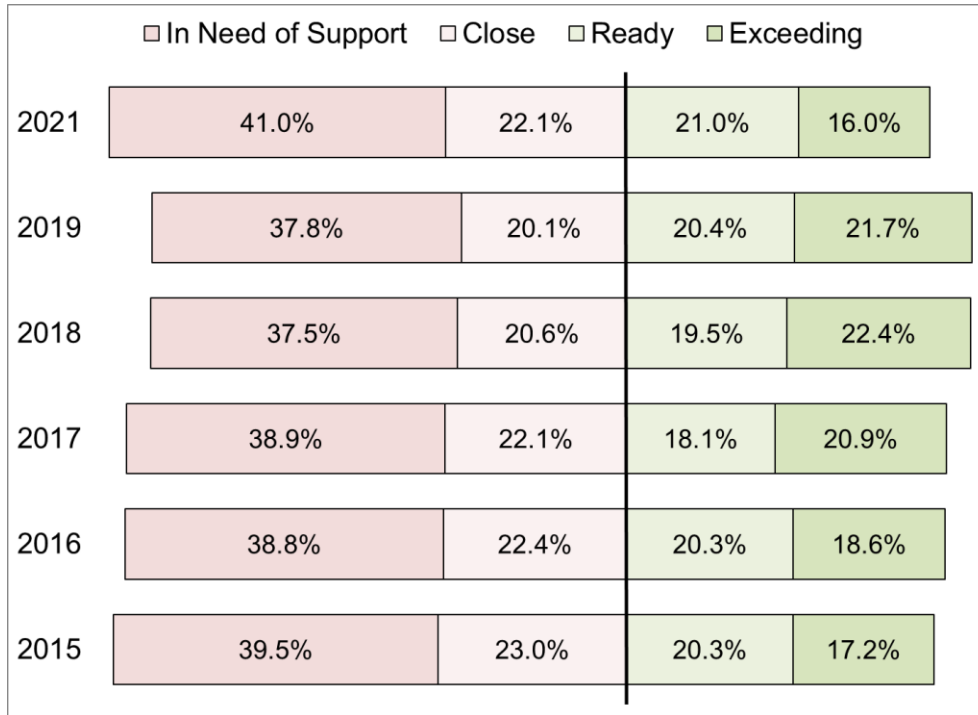


Figure 3.3. Overall percentage meeting ACT Readiness Levels for mathematics, by year

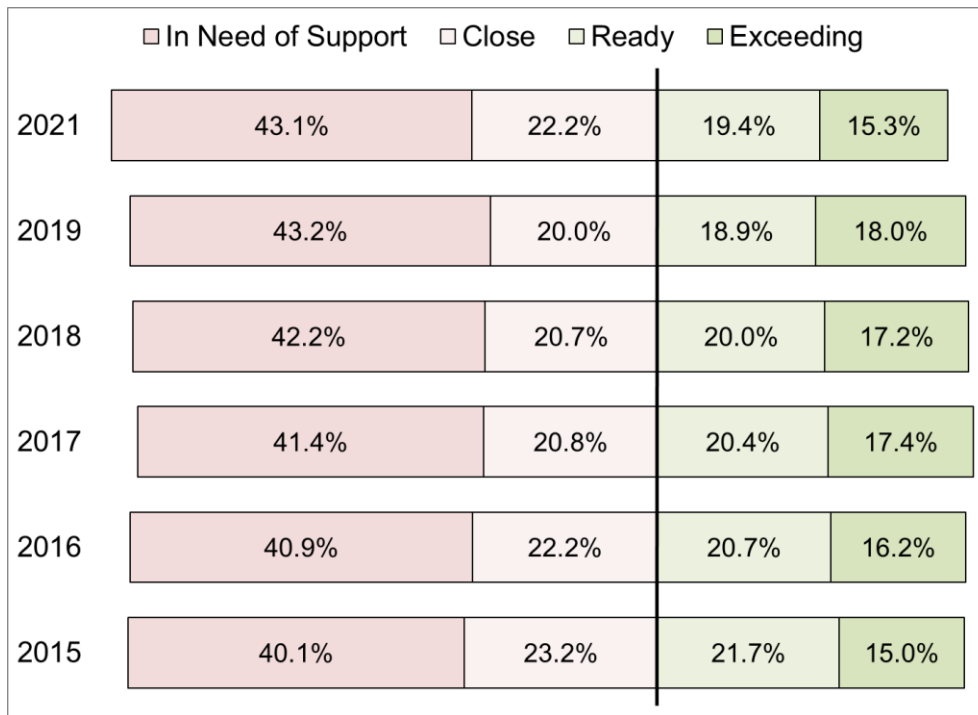


Figure 3.4. Overall percentage meeting ACT Readiness Levels for science, by year

Chapter 4: Technical Characteristics of the Tests

This chapter discusses the technical characteristics of the ACT Aspire Summative Assessments, including score equating procedures and the analysis results for reliability and measurement error using data from the spring 2021 ACT Aspire administration in Wisconsin. A description of the Wisconsin students who participated in testing in the spring of 2021 is provided in [Chapter 2](#) of this technical report.

4.1 Test Equating

The ACT Aspire Summative Assessments system, testing students in English, mathematics, reading, science, and writing, is designed to measure student achievement and key areas of college and career readiness. Multiple ACT Aspire test forms have been developed with rigorous procedures and equated based on large, representative samples. Despite being constructed to follow the same content and statistical specifications, test forms may differ slightly in difficulty. The psychometric equating procedure is used to adjust for these differences in difficulty across forms so that scale scores reported to students have the same meaning regardless of the specific form administered. Through equating, statistical adjustments are made to maintain score interchangeability across test forms (see Holland & Dorans, 2006; Kolen & Brennan, 2014).

4.1.1 Equating Data Collection Design

For ACT Aspire, equating studies are conducted separately for online and paper testing modes. To ensure ACT Aspire scale scores are comparable regardless of testing mode, the paper base form was linked to the online base form through a comparability study conducted in spring 2013 using a random groups design with equipercentile equating. Similarly, in each equating study, a random groups design is typically used. Test forms, including the anchor and new forms, are interspersed at the student level within a testing room so that the forms are distributed evenly and administered to randomly equivalent groups of students. Under a successful implementation of this design, observed differences in test performance across forms can be attributed to differences in form difficulty and equating methods can be applied to adjust for these differences. For each equating study, spiraling occurs separately for paper and online test forms. Across subjects and grades, the sample size of students taking each form ranges from 4,000 to 40,000 or more.

Under certain special situations, ACT Aspire also uses a common-item nonequivalent groups design to equate test forms. For example, in the situation where test specifications are modified to better align to content standards, anchor forms may be

revised, and a common-item equating design is implemented to collect student data across administrations. A revised anchor form is first equated to its original version using a common-item nonequivalent groups design, and then the new forms are equated to the revised anchor form using a random equivalent groups design.

Before equating is performed, collected equating data are checked for spiraling appropriateness and answer key correctness, and irregular student testing behaviors are reviewed. For each ACT Aspire equating study, depending on the equating design, the test level statistics of the anchor forms are examined and found to be similar across years. Equating data collection designs were addressed above; the following section addresses equating methodology to complete the description of an equating study for ACT Aspire.

4.1.2 Equating Methodology

For ACT Aspire, scores on parallel test forms are equated and placed on the vertical score scale using an equipercentile equating methodology (e.g., Kolen and Brennan, 2014). In equipercentile equating, scores on different test forms are considered equivalent if they have the same percentile rank in a given group of students. Equipercentile equating is applied to the raw number-of-points for each subject test separately. The equipercentile equating results are subsequently smoothed using an analytic method described in Kolen (1984) to establish a smooth curve of the equivalents which are then rounded to integers. The conversion tables that result from this process are used to transform raw scores on the new forms to scale scores.

In addition to the scale scores of subject tests, ACT Aspire reports other scores including the Composite score, the ELA score, the STEM score, and the reporting category scores. These reported scores are not equated directly. The Composite, ELA, and STEM scores are a rounded arithmetic average of the scale scores from their contributing tests. Within each grade level, they are comparable across test forms because the scores used to compute them have been equated. The reporting category scores are calculated based on the number of earned points and are not equated across forms.

4.2 Reliability and Measurement Error

For any educational assessment program, an examinee might obtain different scores when tested with parallel forms on different administrations. The variation in scores may reflect random sources of measurement error such as test anxiety, motivation, and distraction. In this context, reliability refers to the consistency of scores across replications of a measurement process. As an index of reliability and precision of

measurement, **coefficient alpha** and **scale score reliability** were computed using Wisconsin Aspire student test data. Under classical test theory (CTT), coefficient alpha is used to estimate the reliability of the test scores and indicates the internal consistency of items on a test. Similarly, scale score reliability is a concept which relates error-score variance and observed-score variance.

The standard error of measurement (SEM) is closely related to test reliability. SEM summarizes the amount of error or inconsistency in test scores. In interpreting an examinee's score, it is helpful to know the SEM of the test score. See Chapter 11 of the technical manual for more discussion on reliability and SEM.

Classification consistency indices quantify the reliability of categorizing examinees into mastery or achievement levels, with respect to specific performance standards. Several model-based approaches have been developed for estimating classification consistency for a single test administration because repeated testing data are seldom available. For ACT Aspire, the assumptions by Hanson and Brennan (1990) about distributions of measurement errors and true scores and the method by Livingston and Lewis (1995) are used and applied to student test data to obtain estimates to evaluate classification consistency. See Chapter 11 of the technical manual for detailed discussion on the approach.

For this report, classification consistency is examined with respect to the ACT Readiness Benchmarks and ACT Readiness Levels, which are introduced in [Section 3.1.2](#) of this report and described in more detail in Chapter 9 of the technical manual. For Wisconsin, after receiving the operational data from the spring 2021 Aspire administration, ACT psychometricians reviewed the data and conducted data cleaning, as needed, for reliability and measurement error analyses. Analysis results are presented as follows.

4.2.1 Raw Score Reliability and SEM for Subject Tests

For the English, mathematics, reading, and science tests, raw score reliability estimates (coefficient alpha) based on the Wisconsin student test data from the ACT Aspire standard online forms administered in 2021 are presented in Table 4.1. Coefficient alpha for the writing test was not estimated as the writing test is composed of a single prompt.

The reliability estimates were computed for the overall student population, the gender groups (female and male), and the major student groups identified for accountability

reporting (African American, Hispanic, White, Economically Disadvantaged, Students with Disabilities, and English Language Learners).

For the overall student population, the observed raw score reliability estimates were larger than 0.80 across all subjects and grades. The fact that the test length of the reading test is shorter has led to lower reliability estimates for reading, compared to other subject tests. In comparison to the national reliability analysis results for online forms, the estimates based on the Wisconsin student test data were all close to the observed range of national estimates. See Chapter 11 of the technical manual for the ranges of the estimates of raw score reliability and SEM based on the national test data. As shown in Table 4.1, the reliability estimates calculated using the data for each student group were within acceptable levels given the test length and the homogeneity of students in each under analysis.

Table 4.1. Raw Score Reliability (Coefficient Alpha) by Group, Subject, and Grade Level

Subject	Grade level	Number of items	All	Female	Male
English	9	50	0.89	0.89	0.89
	10	50	0.90	0.90	0.90
Mathematics	9	42	0.88	0.87	0.89
	10	42	0.89	0.89	0.90
Reading	9	24	0.81	0.78	0.82
	10	24	0.82	0.79	0.84
Science	9	36	0.90	0.89	0.91
	10	36	0.91	0.90	0.92

Subject	Grade level	Number of items	African American	Hispanic	White
English	9	50	0.88	0.87	0.88
	10	50	0.89	0.89	0.90
Mathematics	9	42	0.82	0.83	0.87
	10	42	0.84	0.87	0.89
Reading	9	24	0.77	0.78	0.80
	10	24	0.79	0.80	0.81
Science	9	36	0.83	0.87	0.89
	10	36	0.86	0.89	0.91

Subject	Grade level	Number of items	Econ. Dis.	SWD	ELL
English	9	50	0.88	0.85	0.76
	10	50	0.89	0.87	0.78
Mathematics	9	42	0.85	0.85	0.71
	10	42	0.87	0.87	0.76
Reading	9	24	0.78	0.78	0.65
	10	24	0.81	0.80	0.69
Science	9	36	0.88	0.87	0.70
	10	36	0.89	0.90	0.74

Note. Econ. Dis. = Economically Disadvantaged; SWD = Students with Disabilities; ELL = English Language Learner.

4.2.2 Scale Score Reliability and SEM for Subject Tests, Composite, ELA, and STEM Scores

Table 4.2 presents the scale score reliability and SEM for English, mathematics, reading, and science by grade level. The observed scale score reliability estimates based on the Wisconsin student test data were high in both 9th and 10th grade. Across subjects and grades, the estimates were larger than 0.80. Within a subject and a grade level, the magnitude of the scale score reliability was comparable to the national estimates presented in the technical manual. The values of SEM were marginally larger than those from the national results, which was expected.

Table 4.3 contains the scale score reliability and SEM for the ACT Aspire Composite, ELA, and STEM scores by grade level. The scale score reliability estimates for these combined scores were fairly high with values of 0.90 or larger and were in line with the national analysis results. See Chapter 11 of the technical manual for the national scale score reliability analysis results.

Table 4.2. Scale Score Reliability and SEM, by Subject and Grade Level

Subject	Grade level	Number of items	Reliability	SEM
English	9	50	0.87	3.52
	10	50	0.88	3.51
Mathematics	9	42	0.85	3.21
	10	42	0.87	3.14
Reading	9	24	0.81	3.39
	10	24	0.83	3.33
Science	9	36	0.88	3.09
	10	36	0.89	3.00

Table 4.3. Scale Score Reliability and SEM for the ACT Aspire Composite, ELA, and STEM Scores, by Grade Level

Grade level	Composite		ELA		STEM	
	Reliability	SEM	Reliability	SEM	Reliability	SEM
9	0.95	1.72	0.94	1.79	0.92	2.27
10	0.96	1.63	0.94	1.78	0.93	2.22

4.2.3 Raw Score Reliability and SEM for Reporting Category Scores

Within each ACT Aspire subject test, items are grouped by **reporting categories**, representing different components and topics covered by the test. The reporting categories correspond to the strands used to group the ACT College and Career Readiness Standards and ACT Aspire Performance Level Descriptors. See Chapter 3 of the technical manual for more details on the ACT Aspire reporting categories.

ACT Aspire score reports include reporting category scores, showing how students performed on different areas within each test. The reporting category scores are intended to foster a deeper understanding of strengths and weaknesses within a subject area and to make connections to descriptions of the knowledge and skills students are likely to have. For each reporting category, the percentage and number of points students earn out of the total number of points possible are calculated and reported. In general, the number of points possible for each reporting category may vary across forms. For the forms taken by the majority of students in Wisconsin this year, the number of items per reporting category is presented in Tables 4.4 through 4.7, along with the raw score reliability (coefficient alpha) and SEM for English, mathematics, reading, and science reporting category scores, respectively.

The observed raw score reliability and SEM estimates using the Wisconsin student test data were comparable to those obtained from the national results, and within a subject and a grade, the reliability and SEM estimated from the Wisconsin student test data were quite similar to those estimated from the national data. See Chapter 11 of the technical manual for the raw score reliability and SEM for reporting category scores based on the national data.

For both the Wisconsin and national results, the reliability of reporting category scores is low for some of the reporting categories with very few items. It is important for test users to understand that reporting category scores are not intended for high-stakes decisions. Instead, the reporting category scores can help indicate which areas of the test students found most difficult and can be used as one source of evidence for identifying students' relative strengths and weaknesses.

Table 4.4. Number of Items and Raw Score Reliability and SEM for English Reporting Category Scores, by Grade Level

Grade level	Reporting category	Number of items	Reliability	SEM
9	POW	13	0.72	1.52
	KLA	7	0.68	1.01
	COE	30	0.81	2.39
10	POW	13	0.74	1.49
	KLA	7	0.69	0.97
	COE	30	0.83	2.35

Note. POW = Production of Writing; KLA = Knowledge of Language; COE = Conventions of Standard English.

Table 4.5. Number of Items and Raw Score Reliability and SEM for Mathematics Reporting Category Scores, by Grade Level

Grade level	Reporting category	Number of items	Reliability	SEM
9	GLP	28	0.81	2.30
	N	3	0.39	0.69
	A	7	0.53	1.07
	F	6	0.47	1.07
	G	7	0.42	1.19
	S	4	0.39	0.87
	IES	14	0.75	1.80
	JE	3	0.64	1.10
	MODELING	21	0.80	1.94
10	GLP	28	0.84	2.28
	N	3	0.41	0.69
	A	7	0.62	1.06
	F	6	0.57	1.06
	G	7	0.46	1.17
	S	4	0.41	0.85
	IES	14	0.76	1.80
	JE	3	0.65	1.12
	MODELING	21	0.82	1.91

Note. GLP = Grade Level Progress; N = Number & Quantity; A = Algebra; F = Functions; G = Geometry; S = Statistics & Probability; IES = Integrating Essential Skills; JE = Justification & Explanation.

Table 4.6. Number of Items and Raw Score Reliability and SEM for Reading Reporting Category Scores, by Grade Level

Grade level	Reporting category	Number of items	Reliability	SEM
9	KID	14	0.72	1.94
	CAS	7	0.40	1.19
	IOK	3	0.39	1.38
	TC	12	0.68	1.87
10	KID	14	0.74	1.92
	CAS	7	0.44	1.18
	IOK	3	0.42	1.40
	TC	12	0.70	1.87

Note. KID = Key Ideas and Details; CAS = Craft and Structure; IOK = Integration of Knowledge and Ideas; TC = Text Complexity.

Table 4.7. Number of Items and Raw Score Reliability and SEM for Science Reporting Category Scores, by Grade Level

Grade level	Reporting Category	Number of items	Reliability	SEM
9	IOD	15	0.80	1.78
	SIN	9	0.65	1.39
	EMI	12	0.77	1.53
10	IOD	15	0.82	1.76
	SIN	9	0.69	1.38
	EMI	12	0.79	1.51

Note. IOD = Interpretation of Data; SIN = Scientific Investigation; EMI = Evaluation of Models, Inferences, and Experimental Results.

4.2.4 Classification Consistency for the ACT Readiness Benchmarks and the ACT Readiness Levels

For the ACT Aspire English, mathematics, reading, and science tests, classification consistency for both the ACT Readiness Benchmarks and the ACT Readiness Levels are evaluated. Table 4.8 presents the classification consistency rates by subject and grade level. Both the ACT Readiness Benchmarks and the ACT Readiness Levels are grade-based indicators and thus, the classification consistency indices for 9th and 10th grade are separately estimated. The observed classification consistency rates based on the Wisconsin student test data were either close to the lower bound of the range or within the range of classification consistency rates estimated from the national test data.

For ACT Aspire ELA and STEM scores, the classification consistency rates are computed for the ACT Readiness Benchmarks. The results of classification consistency analyses for ACT Aspire ELA and STEM scores using Wisconsin data are presented in Table 4.9. The observed classification consistency rates for ELA and STEM were fairly high and close to those obtained from the national test results. See Chapter 11 of the technical manual for the national classification consistency analysis results.

Table 4.8. Classification Consistency Rates for ACT Readiness Benchmarks and ACT Readiness Levels, by Subject and Grade Level

Subject	Grade level	Number of items	ACT Readiness Benchmark	ACT Readiness Levels
English	9	50	0.85	0.62
	10	50	0.85	0.63
Mathematics	9	42	0.84	0.62
	10	42	0.86	0.65
Reading	9	24	0.84	0.59
	10	24	0.84	0.62
Science	9	36	0.86	0.65
	10	36	0.87	0.66

Table 4.9. Classification Consistency Rates for ACT Readiness Benchmarks for ACT Aspire ELA and STEM Scores, by Grade Level

Grade level	ELA	STEM
9	0.89	0.93
10	0.89	0.93

Chapter 5: Validity Evidence

According to the *Standards for Educational and Psychological Testing* (American Educational Research Association [AERA], American Psychological Association, & National Council on Measurement in Education, 2014), “validity refers to the degree to which evidence and theory support the interpretations of test scores for proposed uses of tests” (p. 11). Validation is the process of justifying particular interpretations and uses, and it may involve logical, empirical, or theoretical components.

In this chapter, evidence of the validity of ACT Aspire scores for the proposed uses (described in [Section 1.3](#)) is presented. Validity evidence is often organized into the following six areas, as described by the *Standards* (AERA et al., 2014):

1. content
2. cognitive processes
3. internal structure
4. relationships with conceptually related constructs
5. relationships with criteria
6. consequences of testing

This chapter includes evidence related to content, internal structure, relationships with conceptually related constructs, and relationships with criteria.

5.1 Content-Oriented Evidence

ACT Aspire scores are intended to provide inferences about students’ knowledge and skills in English, mathematics, reading, science, and writing. Therefore, one aspect of validation for ACT Aspire is gathering content evidence for the foundational interpretation that ACT Aspire scores are indicative of academic achievement in English, mathematics, reading, science, and writing. Content evidence is important for all uses of ACT Aspire scores and is arguably the most important class of evidence for supporting the use of ACT Aspire scores for measuring progress toward meeting the Wisconsin Academic Standards. Chapter 12 of the technical manual includes a summary of content-oriented validity evidence.

5.2 Relationships With Conceptually Related Constructs: Correlations of ACT Aspire and 11th Grade ACT Test Scores

Often the intended interpretations of test scores imply that the scores should be correlated with conceptually related constructs (AERA et al., 2014). This section provides correlations of ACT Aspire test scores with ACT test scores.

ACT Aspire and the ACT both intend to measure the knowledge and skills most important for success in college and careers (ACT, 2020b). ACT Aspire is intended for earlier grades but is aligned to the same college and career readiness standards as the ACT and tests the same subjects as the ACT. If ACT Aspire and the ACT measure related constructs, high correlations would be expected between the two sets of test scores. Because the ACT is a commonly used measure of college readiness, high correlations of ACT Aspire scores and 11th grade ACT scores directly support the use of ACT Aspire scores for determining if Wisconsin students are on target for college and career readiness.

We examined correlations of ACT Aspire and 11th grade ACT scores collected through spring 2021 for students from Wisconsin. For both ACT Aspire and the ACT test, only tests administered in the spring are included.

Table 5.1 presents the sample sizes, test score means and standard deviations, and correlations of ACT Aspire and ACT scores for students from Wisconsin. Note that the 9th and 10th grade samples each include five cohorts of students: The 9th grade sample includes ACT-tested cohorts of 2017 through 2021, while the 10th grade sample includes ACT-tested cohorts of 2016 through 2020. For the 10th grade sample, the ACT-tested cohort of 2021 is not included because ACT Aspire was not administered in spring 2020. As a result, the sample sizes are similar for the two grade levels.

The correlations are generally similar for the two grade levels. For example, the correlation of Composite scores was 0.88 for both 9th and 10th grade. The correlations ranged from 0.72 for reading to 0.88 for Composite. The correlations suggest that ACT Aspire scores are strong predictors of ACT scores.

In addition to simple correlations, we also estimated disattenuated correlations, adjusted for measurement error. The disattenuated correlations of 10th grade ACT Aspire scores and 11th grade ACT scores ranged from 0.84 for reading to 0.93 for mathematics. The fact that the correlation coefficients are high indicates that ACT Aspire and the ACT measure similar constructs.

Table 5.1. Wisconsin Correlations of ACT Aspire Scores With 11th Grade ACT Scores

Subject	Grade level	N	ACT Aspire		ACT		<i>r</i>	<i>r_{dis}</i>
			Mean	SD	Mean	SD		
English	9	264,485	428.5	9.8	18.6	6.4	0.81	0.90
	10	259,729	430.5	10.4	18.9	6.4	0.82	0.91
Mathematics	9	266,811	426.4	8.6	19.6	5.3	0.82	0.92
	10	262,001	427.8	9.3	19.9	5.3	0.83	0.93
Reading	9	265,304	422.6	7.8	20.0	6.4	0.72	0.85
	10	259,861	423.5	8.1	20.2	6.4	0.72	0.84
Science	9	266,054	426.2	8.8	20.2	5.5	0.75	0.86
	10	261,227	427.7	9.4	20.4	5.4	0.76	0.87
Composite	9	262,154	426.1	7.9	19.7	5.4	0.88	0.91
	10	256,781	427.6	8.4	20.0	5.3	0.88	0.91
ELA	9	271,779	426.4	7.0	18.4	5.5	0.83	0.90
	10	267,393	427.7	7.3	18.6	5.5	0.84	0.90
STEM	9	265,006	426.6	8.2	20.1	5.1	0.85	0.91
	10	260,037	428.1	8.9	20.4	5.1	0.86	0.92

Note. *r* = Pearson correlation; *r_{dis}* = disattenuated Pearson correlation

5.3 Relationships With Criteria: Prediction of High School Course Grades and Academic Rigor

Intended uses of test scores imply that the scores should be predictive of criterion measures that are hypothetically related to the construct measured by the test. In this section, we examine how well ACT Aspire scores predict performance in high school courses.

High school courses help students meet academic standards and prepare for college and careers. Students who are struggling in high school courses are candidates for extra academic support. Thus, by measuring academic standards important for college and career readiness, ACT Aspire test scores should predict high school course grades. And, if predictive of performance in high school courses, ACT Aspire test scores can help to earlier identify students in need of support.

Ninth grade ACT Aspire scores are linked to high school coursework and grades data reported by students when they registered for the 11th grade ACT test in 2021. For 30 different high school courses, students are asked to report the grade they earned in each course already taken, with five options (A, B, C, D, or F). For courses not yet taken, students are asked if they plan to take the course later in high school.

High school GPA (HSGPA) was calculated by averaging the grades reported by students. Only students who reported course grades in each core subject area (English, math, social studies, and natural science) were included in the analysis. On average, students reported grades in 13.6 of the 30 courses. When students register for the ACT test, they are also asked whether they have taken advanced placement, accelerated, or honors courses in English, mathematics, social studies, natural sciences, or foreign languages.

An index of course rigor was obtained using an item response theory (IRT) model known as the graded response model (Samejima, 1969). The model treats the 30 different courses as different items on a test and treats grades as item scores. The model also treats the indicators for advanced coursework and student plans for taking upper-level STEM courses (chemistry, physics, advanced math, and calculus) as “items.” Under the IRT framework, the resulting estimates of student performance are calibrated across students with different (but sometimes overlapping) courses. The graded response model assumes that the course grade probability distribution is determined by course-specific discrimination and difficulty parameters as well as a latent trait distribution that is assumed to have a mean of 0 and standard deviation of 1. (The latent trait to represent student ability is often denoted with the theta symbol θ , and we refer to this trait as the “rigor index.”) Prior research has found that, relative to HSGPA, the rigor index has less skewness and higher correlations with college degree attainment (Allen & Mattern, 2019).

Table 5.2 presents the sample sizes, summary statistics for ACT Aspire test scores and HSGPA, and correlations of ACT Aspire test scores with HSGPA and the rigor index. Results are provided for each ACT Aspire subject. Note that HSGPA and the rigor index include all courses (not subject specific).

Table 5.2. Wisconsin Correlations of 9th Grade ACT Aspire Scores With High School GPA and Academic Rigor

Subject	N	ACT Aspire		HSGPA		r HSGPA	r Rigor
		Mean	SD	Mean	SD		
English	35,456	430.5	9.4	3.19	0.74	0.56	0.62
Math	35,589	428.8	8.6	3.19	0.74	0.60	0.67
Reading	35,497	423.8	7.6	3.19	0.74	0.55	0.60
Science	35,671	428.2	8.7	3.19	0.74	0.58	0.64
Composite	35,205	428.0	7.7	3.19	0.74	0.64	0.70
ELA	36,068	427.9	6.7	3.22	0.73	0.61	0.67
STEM	35,503	428.8	8.2	3.19	0.74	0.63	0.69

Note. r = Pearson correlation

Correlations of ACT Aspire test scores with HSGPA ranged from 0.55 for reading to 0.64 for Composite. Correlations of ACT Aspire test scores with the rigor index are even higher, ranging from 0.60 for reading to 0.70 for Composite. The correlations suggest that ACT Aspire test scores from 9th grade are strong predictors of academic rigor and overall performance in high school courses.

5.4 Differential Item Functioning

According to the *Standards for Educational and Psychological Testing* (AERA et al., 2014), “analyses of the internal structure of a test can indicate the degree to which the relationships among test items and test components conform to the construct on which the proposed test score interpretations are based” (p. 16). **Differential item functioning** (DIF) analyses examine whether items on a test may function differently for identifiable groups of examinees. Results of DIF analyses can be used as one form of validity evidence based on the internal structure of the test.

An item is flagged for DIF when examinees from one group have a higher probability of responding correctly than examinees from another group with the same ability. The procedures used for the analysis of the 2021 Wisconsin student test data include the Mantel-Haenszel common odds-ratio (MH; Holland & Thayer, 1988) procedure and the standardized difference in proportion-correct (STD; Dorans & Schmitt, 1991; Zwick, Donoghue, & Grima, 1993) procedure. Established guidelines were used to classify DIF for each item into one of three levels: A (insignificant DIF), B (slight to moderate DIF), or C (moderate to large DIF). Chapter 13 of the technical manual provides more information on the DIF classification rules.

To obtain stable estimates for the DIF statistics, we analyzed data when, in a comparison, both focal and reference groups had more than 500 students with item-level responses. The DIF analysis results presented in Tables 5.3 and 5.4 include the percentage of items that met DIF A, DIF B, and DIF C flagging criteria for the gender (female vs. male) and race/ethnicity (African American vs. White and Hispanic vs. White) comparisons, respectively. Reviewing the Wisconsin student test data, we found that a vast majority of the items were flagged at the A level, regardless of the comparisons. For gender comparisons, few items were flagged as DIF B level items in mathematics and science (Table 5.3). The percentages of DIF C level items were considered small for all the comparisons. For the gender comparison, 8.4% of the reading test (equivalent to two items) were classified as DIF C level items in both 9th and 10th grades; for the African American vs. White comparison, one item was flagged. In general, the DIF analysis results based on the Wisconsin student test data were

comparable to those of the national analyses. Note that flagging an item does not mean the item is necessarily biased. Some items that are flagged and appear to favor one group over another might just be due to random fluctuations in samples. For ACT Aspire, items that are statistically flagged are further reviewed by content and measurement specialists to eliminate the potential concern of item bias. See Chapter 13 of the technical manual for the national DIF analysis results.

Table 5.3. Summary of Gender DIF Analysis, by Subject and Grade Level

Subject	Grade level	DIF classification (%) for Male vs. Female		
		A	B	C
English	9	100.0	0	0
	10	100.0	0	0
Mathematics	9	92.9	7.2	0
	10	91.7	8.4	0
Reading	9	91.7	0	8.4
	10	91.7	0	8.4
Science	9	98.6	1.4	0
	10	93.1	7.0	0

Note. Row percentages may not add up to 100% because of rounding.

Table 5.4. Summary of Ethnicity DIF Analysis, by Subject and Grade Level

Subject	Grade level	DIF classification (%) for African American vs. White			DIF classification (%) for Hispanic vs. White		
		A	B	C	A	B	C
English	9	98.0	2.0	0	98.0	2.0	0
	10	100.0	0	0	100.0	0	0
Mathematics	9	100.0	0	0	100.0	0	0
	10	95.2	4.8	0	97.6	2.4	0
Reading	9	79.2	16.7	4.2	95.8	4.2	0
	10	79.2	16.7	4.2	91.7	8.4	0
Science	9	100.0	0	0	97.2	2.8	0
	10	100.0	0	0	97.2	2.8	0

Note. Row percentages for a comparison may not add up to 100% because of rounding.

5.5 Depth of Knowledge Analysis

The cognitive complexity level of written passages and the cognitive demands of test items are important characteristics to consider when measuring a student's academic achievement. ACT Aspire assessments reflect the skills that students are expected to have to think, reason, and analyze at high levels of cognitive complexity. ACT Aspire items and tasks target different levels of cognitive complexity with most items targeted at upper levels.

Webb's Depth-of-Knowledge (DOK) system (2002) is widely used across the nation and in many educational contexts for understanding a test item's cognitive complexity. For ACT Aspire, ACT incorporates substantial training, discussion, and multiple inputs to achieve consistent implementation of cognitive complexity levels based on Webb's DOK language. The DOK levels are assigned to reflect the complexity of the cognitive process required, not the psychometric "difficulty" of the item. Unlike other DOK interpretations, ACT only assigns a DOK level 4 value to describe multiday, potentially collaborative classroom activities and assessments designed for learning purposes. By this definition, DOK assignments on any summative assessment including ACT Aspire are limited to values of 1 to 3.

ACT's DOK level 1 corresponds to Webb's level 1 where students are primarily actively using knowledge and skills with limited extended processing. ACT's DOK level 2 extends beyond level 1 and involves applying these cognitive processes to many situations, including real-world scenarios. Therefore, ACT's DOK level 2 aligns with Webb's DOK level 2 and some of Webb's DOK level 3. ACT's DOK level 3 involves situations where the student must apply high-level, strategic thinking skills to short- and long-term situations. Some of these situations are novel, and some require generating something such as a graph, but all involve higher-level thinking skills. Given this interpretation, ACT's DOK level 3 aligns with Webb's DOK levels 3 and 4.

Based on the spring 2021 Wisconsin data, Table 5.5 contains the average percent correct by DOK level for each subject and grade. For mathematics and science at both grade levels, the observed average percent correct decreased as the DOK level increased. Different patterns were observed for English and reading. As discussed previously, items with higher DOK are not necessarily more difficult than items with lower DOK. Generally, the relationship between item difficulty and DOK level, based on the Wisconsin student test data, is strongest for mathematics and science, particularly for upper grade levels.

Table 5.5. Average Percent Correct, by DOK Level, Subject, and Grade Level

Subject	Grade level	Depth of Knowledge (DOK) Level		
		1	2	3
English	9	55.6	56.8	60.7
	10	59.2	60.5	63.8
Mathematics	9	55.8	49.2	30.0
	10	62.2	53.8	32.2
Reading	9	50.3	55.1	46.6
	10	54.6	58.2	50.2
Science	9	69.3	51.7	34.5
	10	72.5	56.3	39.3

Chapter 6: Growth Summary

ACT Aspire Summative Assessments are designed to support interpretations of student growth through the following:

- vertical scaling of test scores across 3rd through 10th grades
- reporting of longitudinal progress charts with a student's current and prior years' scores in English, mathematics, reading, and science
- classification of a student's scores into ACT Readiness Levels, showing how a student scored relative to the ACT Readiness Benchmarks
- predicted paths, which predict a range of a student's ACT Aspire test scores over the next two years
- predicted 10th grade PreACT score range and predicted 11th grade ACT score range
- classification of student growth as low, average, or high on the basis of student growth percentiles (SGPs)

These features most directly support the use of ACT Aspire for determining if students are on target for college and career readiness and for assessing how well Wisconsin schools and districts are preparing students for college and career (accountability). Chapter 14 of the technical manual provides more information on how ACT Aspire supports interpretations of student growth.

In this chapter, we summarize Wisconsin-specific growth data, with comparisons to national norms. We also present an analysis showing how ACT Aspire Interim scores relate to the Summative scores and measures of student growth.

6.1 Comparison of Mean Growth Scores to National Growth Norms

We compared the mean growth scores for Wisconsin to the latest national norms. Results are provided for Wisconsin students who were in 9th grade in spring 2018 and 10th grade in spring 2019. Because tests were cancelled in spring 2020, this is the last cohort of students to have tested in both 9th and 10th grades.

We focused on two types of growth scores: gain scores and SGPs. Gain scores support interpretations of absolute growth, and SGPs support comparisons of growth to norms established from a reference group.

Supported by the vertical scales developed for ACT Aspire, gain scores can be calculated as the arithmetic difference in scores from one year to the next. Positive mean gain scores are anticipated because students are expected to

increase their knowledge and skills each year. SGPs represent the rank of a student's test score compared to the scores of students with the same prior year scores. ACT Aspire SGPs, ranging from 1 to 100, are available for students who test in consecutive years approximately one year apart. The primary use of ACT Aspire SGPs is to support growth comparisons across schools, subject areas, grade levels, instructional programs, and student groups. Such comparisons could lead to insights into what conditions are having greater effects on student learning.

SGPs are a normative measure of growth that must be interpreted with respect to a reference group of students. ACT periodically updates the reference groups used to estimate the SGPs, using data from the most recent year of testing. Reference group samples are created for each subject and pair of adjacent grade levels, and each sample is designed to be representative of the ACT-tested population on race/ethnicity, school affiliation (public or nonpublic), and school percentage eligible for free or reduced lunch. The national growth norms used for this report are based on the 2019 reference group samples, which used ACT Aspire tests administered through spring 2019. The national reference samples include students from both public and nonpublic schools, including those from Wisconsin.

For each grade level and subject, Table 6.1 provides the mean prior year (2018) score, mean current year (2019) score, mean gain score, and mean SGP for students from Wisconsin. The mean gain score and mean SGP are also provided for the national reference samples.

Consistent with national norms, there is considerable variation across grades and subjects in mean gain scores for Wisconsin. As expected, the mean gain scores are positive, showing that students typically increased their knowledge and skills after one year of schooling. For both Wisconsin and the national norms, mean gain scores were smallest in reading and largest in English and mathematics: For Wisconsin, the mean gain score was 0.3 for reading and 1.4 for English and mathematics.

Table 6.1. 9th-Grade (2018) to 10th-Grade (2019) Gain Score and SGP Means, by Subject

Subject	Wisconsin mean					National mean	
	N	9th grade score	10th grade score	Gain	SGP	Gain	SGP
English	57,454	428.9	430.3	1.4	46.1	2.1	49.3
Mathematics	57,470	426.9	428.3	1.4	48.3	1.6	49.2
Reading	57,378	422.8	423.1	0.3	46.9	0.9	49.3
Science	57,216	426.3	427.5	1.2	49.2	1.5	49.5
Composite	56,312	426.5	427.6	1.1	45.0	1.6	49.0
ELA	55,114	426.9	427.4	0.5	43.1	1.4	49.2
STEM	56,825	426.9	428.2	1.3	47.7	1.6	49.4

The mean gains for Wisconsin were less than the mean gains for the national reference sample for all seven reported scores. It is important to keep in mind that the national reference samples are designed to be representative of the ACT-tested population, not the general US population. The mean gain scores for Wisconsin may compare more favorably to the general population.

By definition, the mean SGP for the national reference samples is close to 50 for all subject areas and grade levels (Table 6.1). For Wisconsin, the mean SGP ranged from 43.1 for ELA to 49.2 for science. Similar to the gain score comparison, the mean SGPs for Wisconsin were less than the mean SGPs for the national reference sample, with science showing a very small difference (49.2 for Wisconsin and 49.5 for the national reference sample).

6.2 ACT Readiness Level Transitions

As described in [Section 3.1.2](#), the ACT Readiness Levels for ACT Aspire include four levels: **In Need of Support**, **Close**, **Ready**, and **Exceeding**. To better understand the percentage of students transitioning across readiness levels, we provide the relative frequency of each 10th grade readiness level, conditional on 9th grade readiness level, in Appendix [Table A.3](#). The percentages in Table A.3 are based on the same data used for the gain score and SGP analyses (Wisconsin students with scores from spring 2018 and spring 2019).

For an example of how to interpret the percentages in Table A.3, consider students in 9th grade who were at the In Need of Support level in reading (see cells of Table A.3 in

the red box). Most of the students (82%) remained at the In Need of Support level in 10th grade, 15% improved to the Close level, and 3% improved to the *Ready* level.

6.3 Relating ACT Aspire Interim Scores to Summative SGPs

ACT Aspire Interim Assessments can be administered at any time during the academic year to provide information about students' progress toward end-of-year learning goals (ACT, 2020c). The Interim tests can be thought of as shorter versions of the ACT Aspire Summative tests and cover the same subjects (English, math, reading, and science) and grades (3rd–10th) as the ACT Aspire Summative.

More information on the ACT Aspire Interim Assessments is provided in the ACT Aspire Periodic Technical Manual, available [here](#).

Analyses were conducted to examine how performance on the Interim tests throughout the 2018–2019 academic year relates to performance on the spring 2018 and 2019 Summative tests and to growth from spring 2018 to 2019.

The analysis only includes students who took the Summative tests in spring 2018 and spring 2019 and took at least one Interim test during the 2018–2019 academic year. Because many students took an Interim test more than once, their average Interim test score was used for the analysis. Note that the 2018–2019 school year was used because it is the last year in which students were assessed in both 9th and 10th grades.

Results are provided in Table 6.2 by subject. As expected, Interim test scores earned during the 2018–2019 school year were highly correlated with Summative test scores from both spring 2018 and spring 2019. Correlations ranged from 0.76 for reading (both 9th and 10th grades) to 0.84 for English (10th grade).

The correlations of Interim score and growth from spring 2018 to spring 2019 ranged from 0.20 for mathematics to 0.32 for reading. The positive correlations of Interim test scores with spring-to-spring growth is expected because students who perform well during the academic year should show more growth on Summative tests.

The positive correlations of Interim performance and Summative SGPs suggest that students who perform better on the ACT Aspire Interim tests are more likely to demonstrate more growth on the ACT Aspire Summative tests. The evidence supports using Interim test scores and SGPs to indicate a student's progress over an academic year.

Table 6.2. Relationships of Interim Scores and Summative Scores

Subject	N	Mean Summative Score			Correlation of Interim and Summative Scores		
		9th grade	10th grade	Gain	9th grade	10th grade	SGP
English	6,029	429.1	430.5	1.4	0.83	0.84	0.27
Mathematics	6,492	427.0	428.2	1.2	0.81	0.80	0.20
Reading	6,180	422.7	423.0	0.2	0.76	0.76	0.32
Science	6,055	426.4	427.4	1.1	0.79	0.80	0.27

References

- ACT. (2020a). *ACT Aspire Summative technical manual* [August 2020 Version 2020.1]. Iowa City, IA: ACT.
- ACT. (2020b). *ACT technical manual*. Iowa City, IA: ACT.
- ACT. (2020c). *ACT Aspire Periodic technical manual* [August 2020 Version 2020.1]. Iowa City, IA: ACT.
- 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.
- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.
- Dorans, N. J., & Schmitt, A. P. (1991). *Constructed response and differential item functioning: A pragmatic approach* (ETS RR-91-47). Princeton, NJ: ETS.
- Hanson, B. A., & Brennan, R. L. (1990). An investigation of classification consistency indexes estimated under alternative strong true score models. *Journal of Educational Measurement*, 27(4), 345–359.
- Holland, P. W., & Dorans, N. J. (2006). Linking and equating. In R. L. Brennan (Ed.), *Educational measurement* (4th ed., pp. 187–220). Westport, CT: American Council on Education/Praeger Publishers.
- Holland, P. W., & Thayer, D. T. (1988). Differential item performance and the Mantel-Haenszel procedure. In H. Wainer & H. I. Braun (Eds.), *Test validity* (pp. 129–45). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
- Kolen, M. J. (1984). Effectiveness of analytic smoothing in equipercentile equating. *Journal of Educational and Behavioral Statistics*, 9(1), 25–44.
- Kolen, M. J., & Brennan, R. L. (2014). *Test equating, scaling, and linking: Methods and practices* (3rd ed.). New York, NY: Springer-Verlag.
- 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.

Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometrika Monograph Supplement*, 34, 100–114.

Webb, N. L. (2002, March 28). Depth-of-knowledge levels for four content areas. [Unpublished manuscript]. Retrieved from <https://apps.web.maine.gov/doe/sites/main.gov.doe/files/inline-files/dok.pdf>

Zwick, R., Donoghue, J. R., & Grima, A. (1993). Assessment of differential item functioning for performance tasks. *Journal of Educational Measurement*, 30, 233–251.

Appendix

Table A.1. ACT Aspire Summative Technical Manual Chapters

Chapter Number and Title	Content
1. General description of ACT Aspire assessments and standards	Primary uses and claims of ACT Aspire, content frameworks, and standards alignment
2. Test development	Assessment design and test development processes
3. Assessment specifications	Each subject test's framework, reporting categories, scoring rubrics (for writing), item types/tasks, and test blueprints
4. Item and task scoring	Procedures for scoring multiple-choice, technology-enhanced, and constructed-response items; performance scoring quality control; and automated scoring procedures
5. Accessibility	Description of the ACT Aspire accessibility support system, test administration and accessibility levels of support, and accommodations and tools
6. Test administration	An overview of the processes used to administer ACT Aspire (Users are referred to test administration manuals for more information.)
7. Test security	An overview of test security and the information security framework governing ACT Aspire
8. Scores, indicators, and norms	The meaning of scale scores, combined scores, reporting category and writing domain scores, ACT Readiness Levels, progress indicators, development of norms, and norms for scale scores and combined scores
9. ACT Readiness Benchmarks and Progress Toward Career Readiness	The development and interpretation of college and career readiness indicators, including ACT Readiness Benchmarks and Levels and Progress Toward Career Readiness
10. Scaling and equating	Construction of the vertical score scales and equating procedures (Note that scaling and mode comparability studies are also documented in the appendix of the technical manual.)
11. Reliability and measurement error	Estimates of reliability and standard error of measurement for subject test scores, combined scores, and reporting category scores; rater consistency for writing scores; and classification consistency
12. Validity evidence	Evidence supporting the validity of proposed interpretations and uses of ACT Aspire scores
13. Fairness	Four aspects of fairness, with an emphasis on evidence from differential item functioning (DIF) analyses
14. Growth interpretations	Methodology supporting ACT Aspire's predictions and student growth percentiles; summary data on gain scores and student growth percentiles

Table A.2. Mean Scores and Percent Meeting Benchmark, by Subject, Grade Level, and Year

Subject	Grade level	Year	N	Mean score	Benchmark attainment (%)
English	9	2015	63,743	428.2	60%
		2016	63,281	427.7	58%
		2017	61,889	427.9	60%
		2018	62,036	428.3	61%
		2019	62,926	428.0	60%
		2021	55,097	428.0	60%
	10	2015	60,796	430.4	62%
		2016	61,526	430.7	63%
		2017	60,623	430.3	62%
		2018	59,789	430.5	62%
		2019	60,275	430.0	60%
		2021	53,338	429.9	60%
Mathematics	9	2015	64,054	425.5	41%
		2016	63,304	425.6	42%
		2017	62,135	426.0	42%
		2018	62,112	426.3	44%
		2019	62,984	426.3	46%
		2021	55,191	425.0	40%
	10	2015	60,893	427.2	34%
		2016	61,541	427.6	36%
		2017	60,836	427.6	36%
		2018	59,859	428.2	40%
		2019	60,236	427.9	38%
		2021	53,383	427.0	34%
Reading	9	2015	64,072	422.2	40%
		2016	63,269	422.2	40%
		2017	61,963	422.7	43%
		2018	62,002	422.3	42%
		2019	62,829	421.9	40%
		2021	55,125	421.9	38%
	10	2015	60,757	423.8	39%
		2016	61,467	423.5	38%
		2017	60,858	423.6	38%
		2018	59,805	423.4	36%
		2019	60,168	422.8	34%
		2021	53,388	423.2	34%

Table A.2. Mean Scores and Percent Meeting Benchmark, by Subject, Grade Level, and Year (continued)

Subject	Grade level	Year	N	Mean score	Benchmark attainment (%)
Science	9	2015	63,700	425.8	35%
		2016	63,159	425.7	35%
		2017	61,517	425.9	37%
		2018	61,906	425.7	37%
		2019	62,715	425.9	37%
		2021	54,820	425.6	35%
	10	2015	60,709	427.9	38%
		2016	61,409	427.8	39%
		2017	60,245	427.5	39%
		2018	59,644	427.4	38%
		2019	60,032	427.2	37%
		2021	53,126	427.4	34%
Composite	9	2015	62,924	425.6	
		2016	62,440	425.5	
		2017	60,635	425.9	
		2018	61,108	426.0	
		2019	61,908	425.8	
		2021	54,122	425.4	
	10	2015	59,902	427.5	
		2016	60,624	427.7	
		2017	59,528	427.5	
		2018	58,954	427.7	
		2019	59,278	427.3	
		2021	52,391	427.1	
ELA	9	2015	62,287	425.7	44%
		2016	61,995	425.6	43%
		2017	60,224	425.9	46%
		2018	60,252	426.3	48%
		2019	60,983	425.8	45%
		2021	53,367	425.5	42%
	10	2015	59,482	427.5	44%
		2016	60,052	427.2	43%
		2017	59,038	427.3	44%
		2018	58,128	427.6	46%
		2019	58,250	427.0	43%
		2021	51,664	427.0	41%

Table A.2. Mean Scores and Percent Meeting Benchmark, by Subject, Grade Level, and Year (continued)

Subject	Grade level	Year	N	Mean score	Benchmark attainment (%)
STEM	9	2015	63,393	425.9	17%
		2016	62,836	425.9	17%
		2017	61,187	426.3	19%
		2018	61,570	426.3	20%
		2019	62,375	426.4	21%
		2021	54,511	425.6	16%
	10	2015	60,381	427.8	18%
		2016	61,068	428.0	21%
		2017	59,926	427.9	21%
		2018	59,329	428.2	22%
		2019	59,673	427.9	21%
		2021	52,776	427.5	17%

Table A.3. 10th Grade Readiness Level Transition Percentages, by Subject

Subject	9th Grade = In Need of Support (INS) (%)				9th Grade = Close (CL) (%)				9th Grade = Ready (RD) (%)				9th Grade = Exceeding (EX) (%)			
	INS	CL	RD	EX	INS	CL	RD	EX	INS	CL	RD	EX	INS	CL	RD	EX
English	71	22	6	1	31	40	24	4	8	23	45	25	1	3	17	79
Mathematics	85	13	2	0	42	40	16	2	11	34	40	15	1	7	28	64
Reading	82	15	3	0	40	40	18	2	14	35	40	11	3	15	44	37
Science	82	14	4	1	35	38	22	5	10	25	40	25	2	6	23	70
ELA	88	10	2	0	43	38	17	2	11	30	43	16	1	5	25	68

Note. INS = In need of support; CL = Close; RD = Ready; EX = Exceeding