MAP® Growth™ Technical Report

March 2019

Suggested citation: NWEA. (2019). *MAP® Growth™ technical report*. Portland, OR: Author.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

Below is a list of abbreviations that appear in this technical report.

ALT .... Achievement Level Test (paper-pencil precursor to MAP Growth)
AOR .... Aspects of Rigor
ASG .... Achievement Status and Growth
CCSS .... Common Core State Standards
CCSSO .... Council of Chief State School Officers
CGI................... conditional growth index
CGP .... conditional growth percentile
DIF .... differential item functioning
DOK .... Depth of Knowledge
ELA .... English Language Arts
ELL................... English language learner
ETS .... Educational Testing Service
GRD .... Growth Research Database
HLM .... hierarchal linear model
IEP .... Individualized Education Program
IRT .... item response theory
MAP .... Measures of Academic Progress® (now MAP Growth)
MH .... Mantel-Haenszel
MLE.................. maximum likelihood estimation
MoM .... Model of Man
MPG................. MAP for Primary Grades (now MAP Growth K–2)
MSE .... mean square error
NCRTI .... National Center on Response to Intervention
NGSS............... Next Generation Science Standards
PARCC............. Partnership for Assessment of Readiness for College and Careers
RIT .... Rasch Unit
RMSE............... root mean square error
RTI .... response to intervention
SBAC .... Smarter Balanced Assessment Consortium
SCI .... School Challenge Index
SD.................... standard deviation
SEM .... standard error of measurement
TEI .... technology-enhanced item
TTS .... text-to-speech
UDL.................. Universal Design for Learning

Acknowledgements

It is with great appreciation that we recognize the many people at NWEA who contributed to this technical report. It was a collaborative effort involving people from numerous departments in the organization. We give special thanks to those who conducted the analyses and wrote and edited the document, including Emily Bo, Jing Chen, Laurence Dupray, Garron Gianopulos, Kelly Larson, Sylvia Li, Patrick Meyer, Mary Resanovich, Adam Withycombe, and countless others whose expertise and knowledge about MAP Growth was crucial.

Executive Summary

This technical report is written for measurement professionals and administrators to help evaluate the quality of the MAP® Growth™ assessments. Principal information presented in each chapter is summarized below. This report is not intended to be an administration guide for the tests or a technical description of the hardware and software needed for use of the system. For additional information not covered in this technical report, please contact your local NWEA® representative or consult the NWEA website at www.nwea.org.

Chapter 1: Introduction

This chapter summarizes MAP Growth and describes the background and rationale behind the development of the assessments. MAP Growth assessments are interim adaptive tests that measure a student's academic achievement and growth. Scores are reported on the Rasch Unit (RIT) scale and can be used to track growth and predict performance on state summative assessments. The rationale behind the MAP Growth development has two primary aspects: the need for accurate measurement for all students and the need to provide schools with tests that align to their academic standards. As of February 2018, NWEA has partnered with more than 9,700 education organizations worldwide and has reached approximately 11 million students.

Chapter 2: Test Design

This chapter summarizes the different types of MAP Growth assessments and the rationale behind their designs. The assessments are structured by content area, instructional area, and sub-area. Items are carefully aligned to the standards and assigned learning statements. When new tests are constructed or updated, they are first validated to ensure that each newly aligned MAP Growth item pool performs as intended and that the assessments can withstand multiple administrations per year. Tests are classified as pass, pass with qualifiers, or fail. Most tests pass or receive a qualified pass.

Chapter 3: Item Development

This chapter describes the MAP Growth item types and the item development and review processes, including the MAP Growth Reading passage development process. MAP Growth assessments draw from an item bank containing more than 42,000 items that are carefully aligned to standards and assigned learning statements. All newly developed items are field tested, and items that meet psychometric quality criteria are added to the item bank. Item development and field testing for MAP Growth assessments occurs continually to enhance and deepen the item pool.

Chapter 4: Test Administration and Security

This chapter describes the test administration and test security processes. MAP Growth assessments are untimed and can be administered up to four times a year (fall, winter, and spring, with a fourth optional administration in summer). Access to the MAP Growth system is based on differentiated roles such as system administrator and proctor. Administration training is provided as part of the NWEA professional learning services, and practice tests are available that provide the same access and functionality as the real MAP Growth tests. MAP Growth assessments have several features to improve test fairness and provide more precise and valid measurement, including universal features such as a calculator and highlighter, designated features such as text-to-speech (TTS), and accommodations such as assistive technology. Test security is maintained in a variety of ways, including with large item pools, adaptive testing advantages, a lockdown browser, data encryption, and role-based access.

Chapter 5: Test Scoring and Item Calibration

This chapter describes the development of the RIT scale, the calculation of RIT scores, item calibration, evaluation of field test items, and item parameter drift. It also provides RIT score descriptive statistics, including the mean, standard deviation, and the minimum and maximum RIT scores. The RIT scale is a vertical scale based on the Rasch item response theory (IRT) model. During testing, each item is selected to yield maximum information about the student's ability. Individual tests are constructed based on the student's performance while responding to items constrained in content to a set of standards. A student's final ability estimate indicates the student's location on the RIT scale and is reported as a RIT score from 100 to 350. Each content area has its own unique scale. Scores also include percentile ranks based on the 2015 MAP Growth norms (Thum & Hauser, 2015) to compare students' achievement status and growth to their peers. Field test items are administered in fixed positions during an operational test. Responses are continuously collected on field test items until the items successfully pass calibration and can be administered operationally. Good item parameter estimates are critical to the validity of a test based on IRT, so field test items are checked for model fit via item fit statistics, the Model of Man (MoM) procedure, and human reviews. Finally, periodic reviews of item performance are conducted based on item parameter drift to ensure scale stability across time and student subgroups. Thus far, results have shown that a large majority of MAP Growth items are stable over time and have little to no drift.

Chapter 6: Reporting

This chapter summarizes the MAP Growth reports that are available at the student, class, and district levels. Report types include the Student Profile, Student Progress, Achievement Status and Growth (ASG), Class Breakdown by RIT, District Summary, and Skills Checklists and Screening reports. The learning continuum shows the content a student can encounter throughout the test by instructional area, standards, and RIT bands. This report can be used to show what students performing at a given RIT level on MAP Growth assessments have achieved and what they are typically ready to learn. It has two views: the class view and test view. The reporting software undergoes routine quality assurance processes.

Chapter 7: Reliability

This chapter summarizes the reliability evidence provided for MAP Growth. Reliability refers to the consistency of achievement estimates obtained from the assessment. The reliability of the MAP Growth assessments was examined via test-retest reliability, marginal reliability (internal consistency), and score precision based on the standard error of measurement (SEM). Test-retest results indicate that students' MAP Growth scores are highly consistent for students at different grade levels and from different states. The overall marginal reliabilities for all grades and content areas are in the .90s, which suggests that MAP Growth tests have high internal consistency. Regarding score precision, the MAP Growth adaptive test algorithm selects the best items for each student, producing a significantly lower SEM than fixed-form tests.

Chapter 8: Validity

Validity is defined as the "the degree to which evidence and theory support the interpretations of test scores for proposed uses. Validity is, therefore, the most fundamental consideration in developing tests and evaluating tests" (AERA, APA, & NCME, 2014, p. 11). This chapter summarizes evidence based on test content, internal structure, and relations to other variables.

Chapter 1: Introduction

This technical report documents the processes and procedures employed by NWEA® to build and support the MAP® Growth™ and MAP Growth K–2 assessments for use with the Common Core State Standards (CCSS; National Governors Association Center for Best Practices & Council of Chief State School Officers [CCSSO], 2010)[1] and Next Generation Science Standards (NGSS; NGSS Lead States, 2013)[2].

1.1. MAP Growth Overview

MAP Growth assessments are interim adaptive tests that measure a student's academic achievement and growth in Reading, Language Usage, Mathematics, and Science, as shown in Table 1.1. The assessments are untimed and can be administered up to four times a year in the fall, winter, and spring, with a fourth optional administration in summer. It generally takes students about one hour to complete each MAP Growth test.

Table 1.1. MAP Growth Assessed Grades by Content Area

Content Area	Assessed Grades												
	K	1	2	3	4	5	6	7	8	9	10	11	12
Reading	X	X	X	X	X	X	X	X	X	X	X	X	X
Mathematics	X	X	X	X	X	X	X	X	X	X	X	X	X
Language Usage			X	X	X	X	X	X	X	X	X	X	X
Science*			X	X	X	X	X	X	X	X	X	X	X

*MAP Growth Science assessments in Grades 9–12 were published for the first time in July 2018. MAP Growth Science 3–5 can be administered to students in Grades 2–5. The MAP Growth Science 6+ assessments can be administered to students in Grades 6–12.

MAP Growth assessments have many benefits, including the following:

- Dynamic adjustment to each student's achievement level, providing an accurate indication of their performance and instructional level
- Performance and growth summaries of an individual student and group of students at the grade, classroom, school, and district levels relative to a reference group of examinees
- Frequent administrations throughout the year, allowing teachers to make timely instructional adjustments
- Grade-independent scaling that allows educators to monitor a student's academic achievement and growth regardless of the student's current grade level
- Score reports that include status and growth scores for describing a student's learning from different perspectives
- Untimed test administrations to best measure what students know rather than what they can read and complete in a fixed period of time

---

[1] © Copyright 2010 National Governors Association Center for Best Practices and Council of Chief State School Officers. All rights reserved.
[2] Next Generation Science Standards is a registered trademark of Achieve. Neither Achieve nor the lead states and partners that developed the Next Generation Science Standards were involved in the production of this product, and do not endorse it.

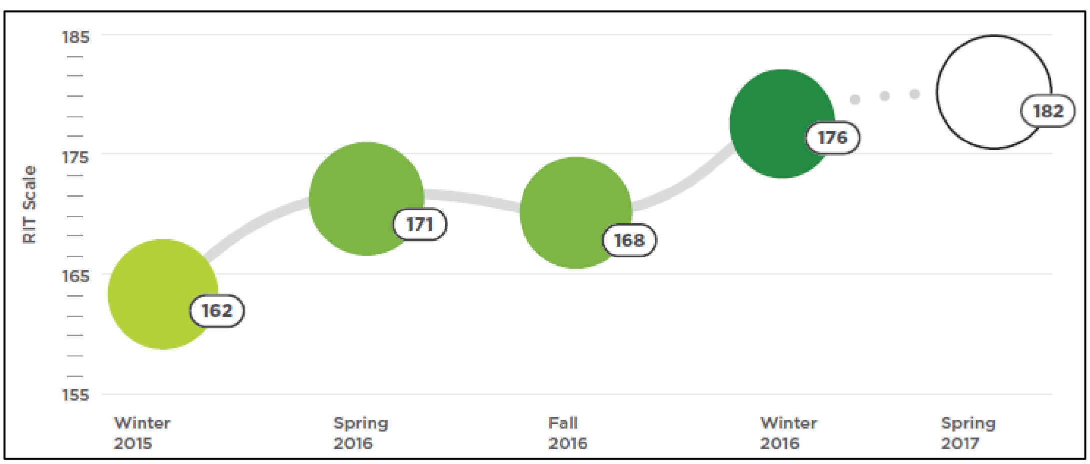
MAP Growth has an item bank containing more than 42,000 items aligned to various content standards. Many states use the CCSS and NGSS, but NWEA also creates a unique set of item pools and assessments for states that have their own state-specific content standards. For each version of the MAP Growth assessment, NWEA content specialists review the standards, select items from the MAP Growth item bank that directly align to the standard statements, and write new items to ensure coverage of the standards. MAP Growth items are dichotomously scored multiple-choice items or technology-enhanced items (TEIs). Each MAP Growth adaptive assessment selects items balanced across the breadth of student learning expectations, ensuring that students see a variety of content across the standards.

MAP Growth assessments are designed to provide accurate measurement of student performance by featuring content across grades and adjusting the assessment outside of grade level. For example, a Grade 3 student would see items aligned to the Grade 3 standards but could also see items aligned to higher and lower grade levels depending on their test performance. Because MAP Growth is administered adaptively, individual students' learning levels, not simply grade-specific achievement levels, are identified. This means that off-grade alignment may be appropriate for an individual student.

Each MAP Growth assessment produces a score in the overall content area, as well as instructional area subscores that can be used to tailor instructional practices and identify specific content a student is most ready to learn. MAP Growth scores are reported on the NWEA Rasch Unit (RIT) scale, an equal-interval vertical scale that is continuous across grades and unique to each content area. Tests of the same content area share a common RIT scale. Score reports also include achievement and growth norms used by teachers to set learning goals for students and provide context for interpreting changes in RIT scores related to the age and grade of students. NWEA conducts MAP Growth norming studies every three to five years. The 2015 MAP Growth norms (Thum & Hauser, 2015) are the most recent.

Changes in students' test scores over time may be interpreted as growth in academic achievement. MAP Growth reveals how much growth has occurred between testing events and, when combined with the NWEA norms, shows how growth compares to a reference group of students. Educators can track growth through the school year and over multiple years, as shown in Figure 1.1.

Figure 1.1. Tracking Growth



1.2. Background

NWEA began in 1973 by a group of school districts looking for practical answers to the following questions. To this day, these questions remain central to the mission of NWEA and, more broadly, to educational assessment and research.

- How can student achievement be efficiently and accurately measured?
- How can assessment results be leveraged to inform instruction?
- How can the rate of learning be accelerated using assessment information?

In 1977, NWEA became an incorporated not-for-profit and began to work with individual school districts in Oregon and Washington (with Portland providing the largest sample of students) to write and field test items that covered the spectrum of student performance in Grades 3–8 in Reading and Mathematics. This work allowed NWEA to create the Achievement Level Tests (ALTs) to improve measurement for students who were progressing normally, falling behind their peers, or excelling beyond their peers. These tests used a multi-stage test design and were administered in paper-pencil form (Ingebo, 1997). The multiple levels made ALTs more precise than a fixed-form test but also logistically complex to administer. These tests were constructed from the NWEA item banks to fit the content standards of each school district.

In 1985, NWEA began to work with districts in Oregon and Washington to create adaptive tests administered on personal computers to make the assessment even more efficient and precise. By this time, NWEA had expanded its testing capabilities to include high school grades and had added content in Language Usage and Science. These tests used the full range of adaptive testing capabilities developed in universities to improve measurement (Weiss & Vale, 1987; Kingsbury & Weiss, 1980). These adaptive tests provided excellent measurement accuracy for a variety of students. However, due to the limitations on computers available in the schools, limitations on networking, and limitations on the client-server software available at that time, most districts continued to use the ALTs and used the NWEA adaptive tests only for special-purpose testing.

In 2000, NWEA released Measures of Academic Progress® (MAP®) using improvements in educational technology. These tests used expanded item pools and took advantage of technological advancements to allow schools to replace their ALTs with adaptive tests for all but a few students with special needs. Since almost every state had a set of content standards in place at the time of the release of MAP, specific items were selected from the item banks to match the content standards in each state.

In 2006, NWEA responded to the growing need for better assessment of younger students by introducing MAP for Primary Grades (MPG). These assessments include audio support to enable students who are beginning readers to access the content and demonstrate their achievement. They include adaptive tests and a set of specific fixed-form pre-tests designed to measure precursor skills that are common to kindergarten curriculum.

Starting in 2017, MAP and MPG are now known as MAP Growth and MAP Growth K–2, respectively. The client-server version of MAP Growth was also retired in 2017 and replaced by the web-based version. As of February 2018, NWEA has partnered with more than 9,700 education organizations worldwide and has reached approximately 11 million students.

1.3. Rationale

The rationale behind the development of MAP Growth has two primary aspects:

1. The quest for accurate measurement for all students
2. A need to provide schools with tests that match their academic content standards

1.3.1. Accurate Measurement

Fixed-form tests tend to lack information for certain segments of the student population. For example, if a fixed-form test is designed to measure well for the middle of the distribution of students, most of the items will be concentrated near the middle of the distribution. These items will be too difficult for students who are struggling and too easy for students who are excelling. This means that the result of the test will provide less information for students at the extreme ends of the distribution than it provides for the students near the middle. Giving the teacher less information about students at the low or high end of the distribution makes it more difficult to target instruction for those students. This is an equity issue for these students, and it certainly reduces the efficiency of teaching them.

The early NWEA researchers realized the equity problem and understood that the tests available at the time failed to give equally precise information for all students. In searching for answers to this problem, these researchers discovered two useful tools:

1. The Rasch item response theory (IRT) model (Rasch, 1960/1980) that allows the development of item banks in which the items have known characteristics. This means that the item characteristics, once estimated, can be applied to new groups of students in the population of interest. This, in turn, makes it possible to create and administer different tests to different students while having all the test scores associated to a common measurement scale.
2. Adaptive testing (Weiss, 1974) that draws items from an item pool according to the performance of each student. As the student answers items correctly, the system chooses more difficult items to administer. If the student answers items incorrectly, the next item will be easier. This type of test allows the test developer to provide a test that has scores with similar precision for every student tested, provided the item pool is large enough and the adaptive testing design is adequate.

The NWEA researchers employed both these tools to create large item banks calibrated to known measurement scales. They then used these item banks to create adaptive tests that measure the students in their schools well by presenting items that, given the purpose of the test, are well matched to a student's experience, characteristics, or behavior. This is known as item targeting, which is a critical influence on test quality.

A fixed-form test might be carefully aligned to a set of specific content standards. If all students in a class were taught according to those content standards, it might be concluded that the items were targeted indirectly to the students through the content. This would be considered a low level of item targeting because it is directed exclusively at the student's experience and ignores other student characteristics and behaviors. A test administered adaptively, on the other hand, presents a higher level of targeting. Items presented may be selected from a core grade-level content pool and from pools that extend both above and below the core pool. Items are selected using a specified content structure. An algorithm is used to estimate the student's achievement level after the student's response to each item and randomly selects the next item from all available items having difficulty values that match the estimate of the student's achievement. Such a test engages the student by presenting items that are neither too easy (leading to boredom) nor too hard (leading to frustration).

When a student remains sufficiently engaged in such a test, the measurement error associated with the test score will be much smaller than a fixed-form test of the same length or even somewhat longer. Therefore, an adaptive test makes efficient use of the time that the student spends in the testing environment by maximizing the level of information that each item contributes to the total test score. The result is total test scores with higher information values, for virtually all students, than would be expected from a fixed-form test of the same length administered to the same group of students.

1.3.2. Content Standards Match

Creation of the adaptive tests depends on the match of the item pools to the content standards of the state. Another difficulty that struck NWEA researchers early on was that assessments taken off the shelf rarely matched the content being taught in the schools. Further, since content standards differed from state to state (and from district to district at that time), no one test could capture the nuances associated with the way a content area was taught in schools from one district or state to the next. It was clear that to establish consistent measurement across locations, the assessment content had to be matched to the content standards of each agency (i.e., a district or state).

The NWEA item banks are large and include content that goes beyond the bounds of any one curriculum structure. Therefore, when developing MAP Growth assessments for an agency, only a portion of the items in the item banks are included in the item pools for the assessments. Content specialists isolate the items in the banks that match the respective content standards, and only those items are included in the assessments. This allows the assessments to be appropriate for the content standards of the agency. When this feature is combined with the capabilities of adaptive testing using IRT, it provides an assessment that uses appropriate content to measure all students in a school with a consistent level of accuracy.

1.4. Intended Uses of Test Scores

MAP Growth assessment data can be used in numerous ways to support student growth and achievement. NWEA supports the use of MAP Growth scores to:

- Monitor student achievement and growth over time, from kindergarten to high school
- Plan instruction for individual students and groups of students at the classroom, grade, school, and district levels
- Compare student performances within normed groups
- Make universal screening and placement decisions within a response to intervention (RTI) framework or for talented and gifted programs
- Predict student performance on external measures of academic achievement, such as the ACT®, SAT®, and on statewide summative achievement tests
- Evaluate programs and conduct school improvement planning
- Summarize scores for district- or school-level resource allocation
- Combine RIT scores with other information (e.g., homework, classroom tests, state assessments) to make educational decisions

Chapter 2: Test Design

The design of each MAP Growth test starts with an analysis of the content standards to be assessed. Items that align to standards are included in a pool and grouped into instructional areas and sub-areas. Although each item pool is tailored to specific standards, all MAP Growth assessments follow the same design principles and content rationale. These principles and rationales are described in this chapter, along with procedures for aligning items to the standards and constructing and validating the assessments.

2.1. Design Principles

This section describes the design principles that provide the foundation for the MAP Growth assessments, including six guiding principles and universal design.

2.1.1. Six Guiding Principles

The MAP Growth system was designed according to guiding principles that reflect educators' needs and help NWEA design assessments for a specific educational purpose. Given its intended purpose, the test should:

1. Be challenging for a student across all items. It should not be frustrating or boring. The goal is to minimize disengagement that can affect a student's results. The adaptivity of MAP Growth ensures that students are presented with content that is neither too far above nor too far below their achievement level.
2. Be economical in its use of student time. It should provide as much information as possible for the time it takes to administer. The adaptivity of MAP Growth helps decrease the amount of testing time required for accurate results.
3. Provide a reflection of a student's achievement that is as accurate and reliable as needed for the decisions to be made based on its results. This is demonstrated by score precision as measured by the standard error of measurement (SEM). The adaptivity of MAP Growth helps lower the SEM, which indicates greater precision in the scores.
4. Consist of content the student should have had an opportunity to learn. The alignment of test items to partner standards ensures that students encounter expected content.
5. Provide information about a student's change in achievement level from one test occasion to another, as well as the student's current achievement level. A single test result is only a snapshot of student achievement. Multiple snapshots are needed to gauge a student's growth over time.
6. Provide results to educators and other stakeholders as quickly as possible while maintaining a high level of integrity in the reported results.

2.1.2. Universal Design

Test development incorporates Universal Design for Learning (UDL) principles to address the needs of diverse populations of students taking the MAP Growth assessments. The NWEA content team applies the UDL principles summarized in Table 2.1 (Thompson, Johnstone, & Thurlow, 2002) and the UDL guidelines (Center for Applied Special Technology [CAST], 2018) when creating test items. These principles improve tests and test fairness by removing characteristics of tests that are unrelated to the measured construct but may inadvertently affect test scores. The result is a more accurate score for the student and a clearer picture of what the student knows and can do. It also provides a framework for incorporating flexibility in the ways the content is presented and how students respond or show their knowledge. It also allows multiple ways for students to be engaged.

Table 2.1. Universal Design Principles

UDL Principle	Description
Inclusive assessment population	Field tests should include students with a wide range of abilities, students with limited English proficiency, and students across racial, ethnic, and socioeconomic lines.
Precisely defined constructs	The test design is clear on the construct(s) to be measured and the purpose for which scores will be used and inferences that will be made from the scores. Universally designed assessments do this by removing barriers, which is referred to as construct-irrelevant variance.
Accessible, non-biased items	To ensure the quality of items, a differential item functioning (DIF) analysis can investigate whether certain items perform differently for various subpopulations. Additionally, using a bias, sensitivity and fairness panel can help eliminate bias before the item is seen by students.
Amenable to accommodations	Accommodations are used to increase access to assessments and to the items within the assessments. Accommodations change the environment on how the test is presented or responded to and is typically used by students with disabilities and by English language learners (ELLs).
Simple, clear, and intuitive instructions and procedures	Assessments should be easy to understand regardless of a student's knowledge and experience. The instructions and procedures of the test and the items should not create barriers for students. The student must be able to access the test as intended.
Maximum readability and comprehensibility	Ensuring readability and comprehensibility is important for clarity and access purposes. It is vital that the construct to be measured is presented clearly with plain language and at the appropriate reading level.
Maximum legibility	This refers to the capability of being deciphered with ease.

2.2. Types of MAP Growth Assessments

There are several types of MAP Growth assessments, as shown in Table 2.2. MAP Growth assessments are offered for different grade bands (K–2, 2–5, and 6+) and account for the developmental needs of students at different age levels.

Table 2.2. MAP Growth Assessments

Test Type	Description	Testing Frequency	Content Areas
MAP Growth K–2	Adaptive test with a cross-grade vertical scale that assesses achievement according to standards-aligned content. Scores from repeated administrations are used to measure growth over time.	Four times per year (three times per school year, plus an optional summer administration)	• Reading • Mathematics
MAP Growth 2–12	Adaptive test with a cross-grade vertical scale that assesses achievement according to standards-aligned content. Scores from repeated administrations are used to measure growth over time.	Four times per year (three times per school year, plus an optional summer administration)	• Reading • Language Usage • Mathematics • Science
Course-Specific High School Mathematics	Adaptive test designed to measure specific content a student may understand in one specialty of Mathematics. It can be used to measure growth over one academic year, fall to spring. Resulting scores provide one indicator of whether a student is ready to move to the next Mathematics course.	Two to three times per year	• Algebra I, II • Geometry • Integrated Mathematics I, II, III

Test Type	Description	Testing Frequency	Content Areas
High School Discipline-Specific MAP Growth Science	Adaptive test designed to measure specific content a student may understand in Life Science. It can be used to measure growth over one academic year, fall to spring. Resulting scores provide one indicator of growth for high school Life Science.	Two to three times per year	• 9–12 Life Science

2.2.1. MAP Growth K–2

MAP Growth K–2 assessments in Reading and Mathematics are designed for students in the primary grades of kindergarten through Grade 2. MAP Growth K–2 includes an adaptive Growth test (formerly known as Survey with Goals), Screening tests, and Skills Checklist tests.[3]

- Screening tests are designed to get baseline information for a new student who is in the earliest stages of learning. They are administered once at the end of pre-K or when a student enters kindergarten. These tests are designed to assess the most foundational skills of literacy and numeracy and are helpful in gathering information about students for whom a teacher may have no previous data.
- Skills Checklists are diagnostic tests that assess knowledge of a specific skill before or after teaching it, or after seeing screening or growth results. Skills Checklists cover a subset of the early reading and early numeracy skills taught in Grades K–2. Each skill area has its own individual assessment. These tests are not adaptive and give students the same items every time they take the same Skills Checklist test. These items are not part of the MAP Growth vertical RIT scale. Skills Checklist tests can be administered as many times as necessary during the school year between Growth assessments to assess skills identified as needing work or currently being instructed in the classroom.

Early identification of each student’s achievement level provides a strong foundation for educators to use in establishing an environment for academic success. The MAP Growth K–2 assessments are designed to:

- Provide student achievement and growth information to aid instructional decisions during the early stages of a student's academic career
- Identify the needs of a variety of primary grade students, from struggling to advanced learners
- Use engaging items, interactive elements, and audio to encourage student participation for more accurate results and to help beginning readers understand the items

All MAP Growth K–2 items include some audio. The amount of audio in each item depends on the skill being assessed, but the stem (i.e., the question in the item) is always read aloud. In other words, every K–2 item has audio, but some items only have audio on the stem while other items are completely presented in audio. For example, number answers in Mathematics items are not typically read, and some standards ask students to identify the number words, so no audio is provided. When the item loads, at least some audio is played automatically. The student can replay any part that has audio. Some graphics also have audio that identifies the graphic (e.g., a graphic of a peach pit may have the audio “pit” associated with it).

[3] Screening tests and Skills Checklist tests are not included in the psychometric analyses described in this technical report.

Most of the content in the MAP Growth Mathematics K–2 assessments has audio. For MAP Growth Reading K–2, audio is provided on items where decoding is not the skill being assessed. For example, items use audio in Reading Foundational Skills to allow students to hear words and associated sounds. Audio support for K–2 students in Reading is essential for assessing foundational content such as phonological awareness and phonics. Since students in Grades K–2 are learning to read rather than reading to learn, providing audio ensures that they will be measured based on what they know and can do, rather than solely on their current reading ability. For assessing comprehension, the assessment includes items that:

- Assess listening comprehension
- Provide audio support with text
- Have audio to be used at the discretion of the student
- Include no audio at all, other than the directions and stem

Professional voiceover artists are used so that items sound as natural and fluent as possible. These professionals are chosen for their voice timbre and crispness of enunciation. The voiceover artists are directed to read the content the way they would to a child with natural pacing and appropriate enunciation.

2.2.2. MAP Growth 2–12

MAP Growth 2–12 assessments measure what students know and inform what they are ready to learn in Reading, Language Usage, Mathematics, and Science. They include an adaptive Growth test and Screening tests. The Screening tests for Grades 2–12 are 20-item adaptive tests that yield an overall score and are administered only once to a student for intake or placement purposes. MAP Growth Mathematics tests are also available for high school students in Algebra 1, Algebra 2, Geometry, and Integrated Mathematics 1, 2, and 3. MAP Growth Science tests are also available for high school students in Life Science (Biology). MAP Growth 2–12 tests are content area specific and built to adhere to the content of agency-specific standards. Test content is organized into large categories called instructional areas and sub-areas. The number of instructional areas ranges from three to seven per test depending on the content area. MAP Growth assessments provide instructional area scores in each content area that supplement an overall score.

2.3. Content Design Rationale

2.3.1. Reading and Language Usage

MAP Growth assesses English Language Arts (ELA) on two scales: Reading and Language Usage. For MAP Growth assessments from Grades 2–12, tests on the Reading scale address reading comprehension, understanding of genres and text, and vocabulary. Assessments on the Language Usage scale cover grammar, mechanics, and the elements of writing. MAP Growth Reading K–2 tests are also on the Reading scale but cover some elements of Language Usage as well as Reading. The MAP Growth Reading K–2 and MAP Growth Reading and Language Usage 2–12 literature reviews (Jiban, 2017) establish a rationale for why Reading and Language Usage are combined on the Reading K–2 test but have separate scales for 2+.

MAP Growth Reading is broken into K–2, 2–5, and 6+ tests. The K–2 test provides targeted audio support and addresses skills appropriate for students who are learning to read, including Reading Foundational Skills and Language and Writing standards. In contrast, students who take the 2–5 and 6+ tests tend to have better reading skills than primary students. The split between the 2–5 and 6+ test helps ensure that students see content appropriate to their age and achievement level. For example, when taking the 6+ test, middle school students reading below grade level will see texts that allow them to demonstrate their reading skills without including overly juvenile references that may be perceived as demeaning. Similarly, advanced elementary readers will be challenged with increasingly complex texts without encountering excerpts from Shakespeare or college course catalogs for which they have no frame of reference.

MAP Growth Language Usage is designed for Grades 2–12 and provides an in-depth, focused exploration of grammar, mechanics, and the elements of writing. Students see increasingly challenging items as their writing abilities grow and flourish, building on the early foundations to add nuance and complexity.

2.3.2. Mathematics

MAP Growth Mathematics is broken into K–2, 2–5, 6+, and high school tests. The decision to have separate K–2 tests was influenced by the unique learning needs of young students and the types of skills assessed at this level, such as counting and cardinality. Audio is provided for K–2 students who are still learning to read and thus require audio support to fairly assess their Mathematics skills. MAP Growth Mathematics tests are built for grade bands 2–5 and 6+ because new content is often introduced at the Grade 6 level as students move into middle school mathematics courses. There is overlap of content across the 2–5 and 6+ tests to support students performing both above and below grade expectations. High school Mathematics tests were created to meet the specific structure of course-based mathematics at the high school level.

2.3.3. Science

MAP Growth Science is broken into grade band tests according to the structure of the standards and breadth of the MAP Growth item bank. Some Science tests are offered with grade bands 3–5, 6–8, and 9–12, while some are offered as 3–5 and 6+. The decision to separate the tests into grade bands was influenced by content appropriateness and standard coverage. This ensures that only well-aligned, appropriate content is part of each test.

2.4. MAP Growth Transition

MAP Growth assessments in each content area and grade band have some overlap in grades and content covered, which is essential given the adaptive nature of the assessments. Determining which assessment is most appropriate for each student depends on the purposes of the assessments, the intentions and uses of the results, and each assessment's measurement characteristics. There may be times when comparisons are desirable across students, classes, schools, or even districts, or required by state policy where it is important to have data from the same MAP Growth assessments for a given grade (e.g., all Grade 2 students taking MAP Growth 2–5).

Grade 2 content is represented in the MAP Growth K–2 tests and the Reading 2–5, Language 2–12, and Mathematics 2–5 tests. MAP Growth K–2 and 2–5 transition decisions should consider students' reading readiness and exposure to content. NWEA recommends students take the same test within a school year, meaning students should not switch tests mid-year because of the need to make strong growth comparisons from fall to spring.

2.5. Instructional Areas and Sub-areas

Each MAP Growth test is defined by a content area such as Mathematics and a grade band such as 2–5. Within each test, the content is further defined by instructional areas such as Geometry, Number Sense, and Measurement that are derived from the structure of the content standards and provide information about how the content area is represented in the test. The instructional areas act as reporting categories. As another layer of defining the test content, each instructional area is further divided into sub-areas. The instructional areas and sub-areas from each MAP Growth test are posted online for partner viewing and use at https://cdn.nwea.org/state-information/index.html. As examples, Table 2.3 – Table 2.9 present the instructional area charts for MAP Growth tests for use with the CCSS and NGSS.

Once NWEA content specialists have created instructional areas and sub-areas for a test, they align standard statements to these areas to establish the test structure and content. This combination of instructional areas, sub-areas, and standard statements is called a test blueprint. Once the blueprints are created, the MAP Growth item bank is reviewed, and appropriate items are aligned to the standards. During test administration, the blueprint helps drive item selection to ensure that items presented to a student cover all instructional areas at a difficultly level appropriate to that student's performance, both overall and within each instructional area. Item selection is not restricted to items within a student's grade, allowing MAP Growth to better target students who are performing above or below the grade level mean for an instructional area.

Table 2.3. Instructional Area Chart for use with CCSS—Reading K–2

CCSS Reading Strands	Instructional Areas & Sub-Areas
MAP Growth Reading K–2	
Reading: Foundational Skills • Print Concepts • Phonological Awareness • Phonics and Word Recognition	**Foundational Skills** • Phonics and Word Recognition • Phonological Awareness • Print Concepts
Writing • Text Types and Purposes • Production and Distribution of Writing • Research to Build and Present Knowledge **Language** • Conventions of Standard English • Knowledge of Language	**Language and Writing** • Capitalize, Spell, Punctuate, • Language: Grammar, Usage • Writing: Purposes: Plan, Develop, Edit
Reading: Literature • Key Ideas and Details • Craft and Structure • Integration of Knowledge and Ideas **Reading: Informational Text** • Key Ideas and Details • Craft and Structure • Integration of Knowledge and Ideas **Speaking and Listening** • Comprehension and Collaboration (SL.2)	**Literature and Informational Text** • Literature: Key Ideas, Craft, Structure • Informational Text: Key Ideas, Details, Craft, Structure
Language • Vocabulary Acquisition and Use **Speaking and Listening** • Presentation of Knowledge and Ideas (SL.4)	**Vocabulary Use and Functions** • Language: Context Clues and References • Vocabulary Acquisition and Use

Table 2.4. Instructional Area Chart for use with CCSS—Reading 2–5 and 6+

CCSS Reading Strands*	Instructional Areas & Sub-Areas
MAP Growth Reading 2–5 and 6+	
Reading: Literature • Key Ideas and Details • Integration of Knowledge and Ideas (RL.9)	**Literary Text: Key Ideas and Details** • Draw Conclusions, Infer, Predict • Summarize; Analyze Themes, Characters, and Events
Reading: Literature • Craft and Structure • Integration of Knowledge and Ideas (RL.7) **Language** • Vocabulary Acquisition and Use (L.5)	**Literary Text: Language, Craft and Structure** • Figurative, Connotative Meanings; Tone • Point of View, Purpose, Perspective • Text Structures, Text Features
Reading: Informational Text • Key Ideas and Details • Integration of Knowledge and Ideas (RI.9)	**Informational Text: Key Ideas and Details** • Draw Conclusions, Infer, Predict • Summarize; Analyze Central Ideas, Concepts and Events
Reading: Informational Text • Craft and Structure • Integration of Knowledge and Ideas (RI.7, RI.8) **Language** • Vocabulary Acquisition and Use (L.5)	**Informational Text: Language, Craft and Structure** • Point of View, Purpose, Perspective, Figurative and Rhetorical Language • Text Structures, Text Features
Reading: Informational Text • Craft and Structure (RI.4) **Language** • Vocabulary Acquisition and Use (L.4, L.5, L.6)	**Vocabulary: Acquisition and Use** • Context Clues and Multiple-Meaning words • Word Relationships and Nuance • Word Parts, Reference, and Academic Vocabulary

*Where strands are mapped among multiple goals, specific standards are indicated for each goal.

Table 2.5. Instructional Area Chart for use with CCSS—Language Usage 2–12

CCSS Reading Strands*	Instructional Areas & Sub-Areas
MAP Growth Language Usage 2–12	
Writing • Text Types and Purposes • Production and Distribution of Writing • Research to Build and Present Knowledge **Language** • Knowledge of Language	**Writing: Write, Revise Texts for Purpose and Audience** • Plan and Organize; Create Cohesion, Use Transitions • Provide Support; Develop Topics; Conduct Research • Establish and Maintain Style; Use Precise Language
Language • Conventions of Standard English (L.1)	**Language: Understand, Edit for Grammar, Usage** • Parts of Speech • Phrases, Clauses, Agreement, Sentences
Language • Conventions of Standard English (L.2)	**Language: Understand, Edit for Mechanics** • Capitalization • Punctuation • Spelling

Table 2.6. Instructional Area Chart for use with CCSS—Mathematics K–2 and 2–5

<table>
<tr><th>CCSS Mathematics Domains</th><th>Instructional Areas & Sub-Areas</th></tr>
<tr><td rowspan="4">Counting & Cardinality
Operations & Algebraic Thinking
Number & Operations in Base Ten
Number & Operations – Fractions
Measurement & Data
Geometry</td><td>MAP Growth Mathematics K–2</td></tr>
<tr><td>Operations and Algebraic Thinking
Represent and Solve Problems
Properties of Operations
Number and Operations
Understand Place Value, Counting, and Cardinality
Number and Operations: Base Ten and Fractions
Measurement and Data
Solve Problems Involving Measurement
Represent and Interpret Data
Geometry
Reason with Shapes and Their Attributes</td></tr>
<tr><td>MAP Growth Mathematics 2–5</td></tr>
<tr><td>Operations and Algebraic Thinking
Represent and Solve Problems
Analyze Patterns and Relationships
Number and Operations
Understand Place Value, Counting, and Cardinality
Number and Operations in Base Ten
Number and Operations – Fractions
Measurement and Data
Geometric Measurement and Problem Solving
Represent and Interpret Data
Geometry
Reason with Shapes, Attributes, & Coordinate Plane</td></tr>
</table>

Table 2.7. Instructional Area Chart for use with CCSS—Mathematics 6+

<table>
<tr><th>CCSS Mathematics Domains</th><th>Instructional Areas & Sub-Areas</th></tr>
<tr><td colspan="2">MAP Growth Mathematics 6+</td></tr>
<tr><td>Ratios & Proportional Relationships
The Number System
Expressions & Equations
Functions
Geometry
Statistics & Probability</td><td>Operations and Algebraic Thinking
Expressions and Equations
Use Functions to Model Relationships
The Real and Complex Number Systems
Ratios and Proportional Relationships
Perform Operations
Extend and Use Properties
Geometry
Geometric Measurement and Relationships
Congruence, Similarity, Right Triangles, & Trigonometry
Statistics and Probability
Interpreting Categorical and Quantitative Data
Using Sampling and Probability to Make Decisions</td></tr>
</table>

Table 2.8. Instructional Area Chart for use with CCSS—High School Mathematics

<table>
<tr><th>CCSS Mathematics Courses/ Domains</th><th>Instructional Areas & Sub-Areas</th></tr>
<tr><td rowspan="2">High School: Number and Quantity
The Real Number System
Quantities
The Complex Number System
Vector & Matrix Quantities</td><td>MAP Growth Mathematics Algebra 1</td></tr>
<tr><td>Equations and Inequalities
Reason Quantitatively and Use Units
Creating Equations and Inequalities
Reasoning with Equations and Inequalities</td></tr>
</table>

<table>
<tr><th>CCSS Mathematics Courses/ Domains</th><th>Instructional Areas & Sub-Areas</th></tr>
<tr><td rowspan="9">High School: Algebra
• Seeing Structure in Expressions
• Arithmetic with Polynomials & Rational Expressions
• Creating Equations
• Reasoning with Equations & Inequalities
High School: Functions
• Interpreting Functions
• Building Functions
• Linear, Quadratic, & Exponential Models
• Trigonometric Functions
High School: Geometry
• Congruence
• Similarity, Right Triangles, & Trigonometry
• Circles
• Expressing Geometric Properties with Equations
• Geometric Measurement & Dimension
• Modeling with Geometry
High School: Statistics & Probability
• Interpreting Categorical & Quantitative Data
• Making Inferences & Justifying Conclusions
• Conditional Probability & the Rules of Probability
• Using Probability to Make Decisions</td><td>Numerical and Algebraic Expressions
• The Real Number System
• Seeing Structure in Expressions
• Arithmetic with Polynomials
Functions
• Interpreting Functions
• Building Functions
• Linear and Exponential Models
Descriptive Statistics
• Interpreting Categorical and Quantitative Data</td></tr>
<tr><td>MAP Growth Mathematics Algebra 2</td></tr>
<tr><td>Equations and Inequalities
• Creating Equations and Inequalities
• Reasoning with Equations and Inequalities
Numerical and Algebraic Expressions
• The Complex Number System
• Seeing Structure in Expressions
• Arithmetic with Polynomials and Rational Functions
Functions
• Interpreting Functions
• Building Functions
• Linear, Exponential, and Trigonometric Functions
Descriptive Statistics
• Descriptive Statistics</td></tr>
<tr><td>MAP Growth Mathematics Geometry</td></tr>
<tr><td>Congruence, Similarity, Right Triangles, & Trig
• Congruence
• Similarity, Right Triangles, and Trigonometry
Geometric Properties with Equations and Circles
• Expressing Geometric Properties with Equations
• Understand and Apply Theorems About Circles
Geometric Measurement and Modeling
• Geometric Measurement and Dimension
• Modeling with Geometry
Applications of Probability
• Applications of Probability</td></tr>
<tr><td>MAP Growth Mathematics Integrated Mathematics 1</td></tr>
<tr><td>Algebra and Quantities
• Reason Quantitatively and Use Units
• Creating Equations and Inequalities
• Reasoning with Equations and Inequalities
• Seeing Structure in Expressions
Functions
• Interpreting Functions
• Building Functions
• Linear and Exponential Models
Geometry
• Congruence
• Expressing Geometric Properties with Equations
Descriptive Statistics
• Interpreting Categorical and Quantitative Data</td></tr>
<tr><td>MAP Growth Mathematics Integrated Mathematics 2</td></tr>
<tr><td>Algebra and Number
• The Real Number System
• The Complex Number System
• Creating Equations and Inequalities</td></tr>
</table>

CCSS Mathematics Courses/ Domains	Instructional Areas & Sub-Areas
	• Reasoning with Equations and Inequalities • Seeing Structure in Expressions • Arithmetic with Polynomials **Functions** • Interpreting Functions • Building Functions • Linear, Exponential, and Trigonometric Functions **Geometry** • Congruence • Similarity, Right Triangles, and Trigonometry • Circles • Expressing Geometric Properties with Equations • Geometric Measurement and Dimension **Applications of Probability** • Applications of Probability
	MAP Growth Mathematics Integrated Mathematics 3
	Algebra and Number • The Complex Number System • Seeing Structure in Expressions • Arithmetic with Polynomials and Rational Expressions • Creating Equations and Inequalities • Reasoning with Equations and Inequalities **Functions** • Interpreting Functions • Building Functions • Linear, Exponential, and Trigonometric Functions **Geometry** • Geometry **Descriptive Statistics** • Descriptive Statistics

Table 2.9. Instructional Area Chart for use with NGSS—Science 2–12

NGSS Science Domains*	Instructional Areas & Sub-Areas
MAP Growth Science 2–12	
Life Science • From Molecules to Organisms: Structures and Processes • Ecosystems: Interactions, Energy, and Dynamics • Heredity: Inheritance and Variations of Traits • Biological Evolution: Unity and Diversity	**Life Science** • From Molecules to Organisms: Structures and Processes • Ecosystems: Interactions, Energy, and Dynamics • Heredity: Inheritance and Variations of Traits; Biological Evolution: Unity and Diversity
Physical Science • Matter and Its Interactions • Motion and Stability: Forces & Interactions • Energy • Waves and Their Applications in Technologies for Information Transfer	**Physical Science** • Matter and Its Interactions • Motion and Stability: Forces and Interactions • Energy; Waves and Their Applications in Technologies for Information Transfer
Earth and Space Science • Earth's Place in the Universe • Earth's Systems • Earth and Human Activities	**Earth and Space Science** • Earth's Place in the Universe • Earth's Systems • Earth and Human Activities
Engineering Design*	**N/A**

*Items aligned to Engineering Design standards are embedded in each instructional area.

2.6. Learning Statements

Every item in the NWEA item bank is associated with a learning statement, which is a simple statement that describes the content the item is assessing. Learning statements are authored and assigned to items by NWEA content specialists. A content specialist will review an item—its intent, target, and existing standard alignments—and select or write a learning statement that captures the content of the item (without describing the item in detail). Learning statements allow NWEA to describe the contents of a MAP Growth assessment without exposing the items themselves. Because learning statements are assigned to items, they have indirect relationships to standard statements, RIT values, and other data points via the items. These relationships among learning statements, standards, and RIT values form the basis of the learning continuum (for more information on the learning continuum, please see Section 6.1.4. of this technical report).

2.7. Item Alignment to Standards

MAP Growth items are aligned to many unique standard sets. When a new standard set is released by a state or other agency, NWEA content specialists review the standard set and align the MAP Growth item bank to the standard statements. This is done for every standard set that is the basis for a MAP Growth assessment. To perform alignment, NWEA content specialists craft alignment guidelines tailored to the structure of the standards that are based on a review of supporting documents (e.g., progressions documents, tools for the Common Core, Illustrative Mathematics items). An item is considered aligned when the item targets either the whole standard or an integral part of a standard in a way that is both grade-appropriate and at a level of cognitive complexity addressed by the standard.

2.7.1. Alignment Studies

As part of the ongoing commitment to improve the alignment of items, NWEA content specialists conduct internal alignment analyses to assess how well MAP Growth items align to standards. Regular reviews of alignment are valuable, as changes in standards, academic and pedagogical thinking, and industry expectations necessitate consideration and adjustments to alignment practices. This work examines and rates each item in the item bank against a content-specific rubric. It not only checks alignment to standards, but also helps to inform future item development.

NWEA also engages with third parties to conduct external alignment studies. For example, EdMetric completed an external alignment study for MAP Growth CCSS assessments (Egan & Davidson, 2017). NWEA randomly sampled 20% of the MAP Growth and MAP Growth K–2 CCSS item pools for use in the study. Overall, EdMetric's results show that MAP Growth assessments have very good alignment in terms of categorical concurrence, cognitive complexity, and range and balance of knowledge.

2.7.2. Alignment Guidelines

Table 2.10 presents the alignment guidelines for all MAP Growth content areas and standard sets.

Table 2.10. Alignment Guidelines for MAP Growth

<table>
<tr><th>Approach to:</th><th>ELA</th><th>Mathematics</th><th>Science</th></tr>
<tr><td>Definition of an aligned item</td><td colspan="3">A student needs to demonstrate the knowledge and/or skill expressed* in the standard to respond correctly to the item. The student cannot or most likely cannot answer correctly without that knowledge and/or skill. The item may address the whole standard or a part of the standard in order to best focus on a single skill, a single portion of significant content, and/or a single cognitive level within the standard.</td></tr>
<tr><td rowspan="2">Assessable and non-assessable standards</td><td colspan="3">NWEA only aligns to standards that have been defined as assessable. Assessable standards are the most granular standards for each MAP Growth product on each scale. Exceptions to granularity are noted further below. Standards are only marked as assessable if they are appropriate for interim/formative assessment; NWEA has the functionality to assess them; and they are intended to be used on current blueprints.</td></tr>
<tr><td>Skills that are impractical for NWEA products (e.g., lengthy multi-part tasks that require longer than a normal class period) are not marked assessable. However, some standards (such as in writing, oral responses) are considered assessable via an approximation (for now).
For all CCSS-like ELA tests, including K–2, parent standards are marked as non-assessable. Exception: parents used to assess progressive standards (Progressives are L.1 at grades 4+, L.2 at grades 6+, and L.3 at grades 4+.)
MAP Growth K–2:
The inclusion of audio in MAP Growth K–2 allows for assessment of standards in Reading: Foundations and some listening standards from the Speaking and Listening strand.
Standards requiring students to produce oral responses are assessed in a manner befitting a computer-adaptive assessment because these items still provide valuable information to teachers about students' knowledge of specific skills.</td><td>Skills that are impractical for NWEA products (e.g., lengthy multi-part tasks that require longer than a normal class period, or evidence cannot be provided that they are preforming the standard) are not marked assessable. If some part of the standard CAN be assessed, mark assessable.</td><td>Assessability is based only on content, not skills, since most science standard sets recommend a “mix-and-match” approach to content and skills.</td></tr>
<tr><td>Prerequisite skills, related content, and implied content</td><td colspan="3">Items assessing prerequisite skills and/or content are not aligned.
Implied content is often open for interpretation. Therefore, content teams must make decisions and document those decisions for specific standards that are open to interpretation. Decisions must be based on deep consideration of the standard, standard set, and available resources from experts.
The term “e.g.” indicates examples of the type of content/skills that could fulfill the standard, but it is not an exhaustive list and the listed examples are not required to be assessed. The term “i.e.” indicates a rewording of the standard and therefore defines the limits of the content/skills that are included as an integral part of the standard.
If a standard says including, it means the content must be included when assessing that entire standard (it does not all have to be included in a single MAP Growth item, though); when such as is used, it has a similar meaning as e.g.</td></tr>
</table>

<table>
<tr><th>Approach to:</th><th>ELA</th><th>Mathematics</th><th>Science</th></tr>
<tr><td>Cognitive verbs/ cognitive expectation in a standard</td><td>The cognitive verbs are closely considered as the primary indication of the cognitive expectation associated with a given standard. Items that do not meet that cognitive expectation should not be aligned. However, some standards, most notably writing, are assessed via an approximation that does not meet the expectation or exact action encompassed by the cognitive verb. Decisions should be clearly documented. This can be more difficult to achieve with non-CCSS standard sets.</td><td>Consider the intended cognitive demand (including rigor) of the standard. As the Mathematics team continues to define their approach to rigor, this will be addressed more in the alignment to multiple dimensions section.
Exceptions: product/tech limits may reduce the ability to assess at the intended level.</td><td>Not used for alignment (in lieu of aligning items that combine the content with a range of cognitive demand and science/engineering practices, which is more in keeping with current practices in science education)</td></tr>
<tr><td rowspan="2">Granularity of alignment (e.g. parent/child, anchors, clusters)</td><td colspan="3">Align to most granular portion of standard except in cases noted below.</td></tr>
<tr><td>MAP Growth Reading and MAP Growth K–2 do not align items to CCSS parent standards, and Language Usage does so only in a limited circumstance. NWEA tries to apply this approach to non-CCSS standard sets as well, but sometimes doing so would not match the apparent intent of the standard creators (to have the granular standards be the definition of what is assessed by that parent standard) and so the approach is adapted.
For ELA, NWEA recognizes the special assessability concerns around the standards CCSS designates as Language Progressive skills. NWEA has items targeting these progressive skills not only when they are first introduced but also at subsequent grades in accordance with the CCSS grade recommendation. Because CCSS has no codes or ways to directly note that alignment at the higher grades, NWEA uses the overarching/parent standards (L.1, L.2, and L.3) to align items assessing these progressive skills at higher grades.
Many CCSS-based standard sets do not adopt this aspect of the CCSS.</td><td>Items designed to assess the standard level must match the language of both the cluster and the standard but are aligned at the standard level.
Criterion for aligning to the cluster level: The item assesses a single skill not specifically spelled out in granular standards, but either covers multiple standards in the cluster OR matches the intent of the grade.</td><td></td></tr>
<tr><td>Alignment to the whole standard or portions of a standard</td><td colspan="3">If possible, alignment would be to the entire standard. However, when standards are broad or complex, single items can target portions of a standard.</td></tr>
</table>

<table>
<tr><th>Approach to:</th><th>ELA</th><th>Mathematics</th><th>Science</th></tr>
<tr><td rowspan="2">Grade-level considerations</td><td colspan="3">Items with distractors that have content that is above grade level should be aligned to a higher grade-level standard, if at all.</td></tr>
<tr><td>A holistic determination of grade level must be made that considers vocabulary, context, complexity of the task, readability of the text, and the content included in distractors.
The text in an item must be sufficiently complex for the grade level for it to fully align to that grade's standard. Consequently, for items in common stimulus passage sets, the text complexity of the passage is always considered.**
The Reading passage asset adheres to quantitative (Lexile® & Flesh-Kincaid) text complexity and qualitative (conceptual appropriateness) measures as appropriate for the grade/grade band indicated in the item specifications.</td><td colspan="2">All parts of a Mathematics or Science item should be at a reading level of at least two grades below the standard grade. Language should be as simple as possible to avoid assessing reading ability instead of mathematics/science ability. Construct-specific vocabulary can be used if necessary to appropriately assess the standard. An item should not align if it uses content vocabulary that is more advanced than the target standard.</td></tr>
<tr><td>Alignment to multiple dimensions</td><td>n/a</td><td>Math practices and Aspects of Rigor (AOR) are not currently being used for alignment.
Math Practices: LS's have been tagged with these but are hard to determine without a student explaining their thought process.
Aspects of Rigor: Upcoming project will involve tagging bank with AOR, which will play a role in alignment in the future.</td><td>Only the content dimension is used to determine alignment to a standard, but items aligned to multidimensional standard sets must include at least one additional dimension (does not have to be the same dimension as in the standard). This is due to the recommended "mix-and-match" nature of the science education community's current approach to integrating science/engineering practices, concepts, and content.</td></tr>
<tr><td>Basis for alignment decisions</td><td colspan="3">Alignment decisions are based on information and resources obtained from the CCSS website (Mathematics and ELA) and the NGSS website (Science). For all content areas, this includes the appendices and other materials available at the sites. Additional resources provided by organizations closely involved with developing the CCSS or NGSS, sample items from the consortia, and other vetted sources are also consulted.</td></tr>
</table>

*Content/skills should be directly stated or strongly implied. If implied, the acceptable content/skills should be documented by the content team, with decisions based on discussion and resources from expert sources.
**Alignment philosophy for ELA common stimulus items.

2.8. Test Construction

MAP Growth tests are constructed by combining a blueprint containing instructional areas and sub-areas, standards aligned to these areas, a standard-aligned item bank, and an appropriate test design. These components form the eligible item pool for the test, along with the reporting structure and how all the eligible items fit into this structure. Additional constraints may be added to a test that may further limit the eligible item pool, including item selection requirements during test administration as required by the test type and item filters based on specific item metadata. These constraints are based on the target student population and may include item attributes such as item language or item accessibility for different student populations.

The test behavior during testing is also defined in terms of the test length and item selection criteria for each section of the test as determined by the test content area and purpose. Once these elements are combined, the test is published to the testing platform as a defined set of behaviors and test metadata elements. Each item is also published to the testing platform, along with item metadata and information that determines to which tests the items belong. Tests go through a series of checks, including test content validation that simulate test runs of students at different ability levels, to ensure that the test item pools provide sufficient depth to cover the achievement continuum within each instructional area. Tests are then made available to specific partners based on their licensing agreements with NWEA.

2.9. Test Content Validation

Test content validation is performed as part of the broader process of aligning MAP Growth to different content standards and publishing new tests. The purpose of content validation is to ensure that each newly aligned MAP Growth item pool performs as intended. It takes the form of test simulations with the operational item pool to determine the accuracy of student ability estimation and content coverage of an adaptive test. Tests are classified as pass, pass with qualifiers, or fail. Most tests pass or receive a qualified pass.

An NWEA psychometrician conducts the simulation studies by following the steps below:

1. Set each simulated student's RIT score to a known value. This known student ability or "true RIT score" represents the extreme ends of the distribution (10th and 90th percentiles according to the 2015 norms). Once the estimated RIT score is obtained from the simulation, it is compared to the known value to determine the accuracy of estimation resulting from the adaptive testing process.
2. Simulate a MAP Growth adaptive test based on the operational item pool.
3. Simulate student growth over a two-year timeframe, typically six to eight administrations.
4. Apply longitudinal constraints that prevent a student from seeing the same item more than once in a set timeframe, typically 14 months (e.g., a student is not supposed to see the same items within 14 months).

The simulation produces information about estimation accuracy, content balancing, item selection, and item-pool depth. To determine if a test passes the validation, the psychometrician evaluates the following:

- Ability estimation based on statistics including bias, mean square error (MSE), root mean square error (RMSE), and SEM. The better the estimation, the smaller these statistics will be.

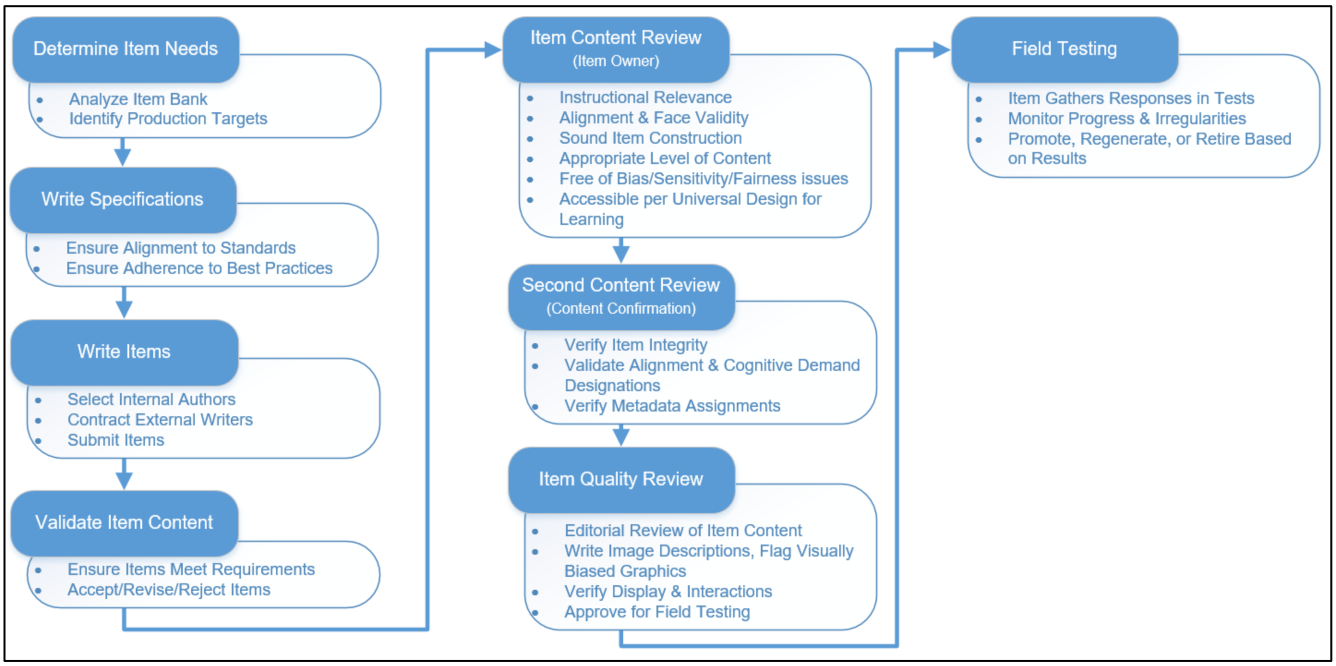
- Content balancing based on how well the adaptive algorithm produces a test that meets the blueprints. A quality adaptive test should administer items distributed equally among the instructional areas in the blueprint.
- The efficiency of the adaptive algorithm based on the discrepancy between the interim ability estimate and item difficulty. The sooner the algorithm settles on the simulated student's true ability value, the sooner the SEM criteria are satisfied.
- Item pool depth based on item RIT distribution at the overall test and instructional area levels. At each level, the pool should ideally span the full range of RIT values and have an adequate number of items at each RIT value to avoid running out of items.

Chapter 3: Item Development

MAP Growth assessments draw from an item bank containing more than 42,000 items. Item pools are subsets of the entire bank that are aligned to specific content standards such as the CCSS. The pools cover all instructional areas and difficulty levels across the full range of the RIT scale and are large enough to support multiple administrations annually without a student seeing the same item twice. The quality and depth of the MAP Growth item pools ensure precise measurement while meeting the test requirements.

Items are continuously added to the pools using a rigorous item writing, review, and field testing process. Figure 3.1 illustrates the MAP Growth item development steps. Item development processes occur year-round and are efficient, allowing items to be ordered, reviewed, and in front of students for field testing quickly. New MAP Growth items are constantly being developed and added to the item pool; 15,000+ items have been published over the last three years across all content areas.

Figure 3.1. Item Development Flowchart



In addition to new items, the MAP Growth item bank is reviewed regularly for quality, examining elements that may include alignment, content accuracy, relevance, bias and sensitivity, style standards, and display. Items may be removed from the bank because of these reviews, public exposure, or issues reported by partners through the in-test interface.

3.1. Item Types

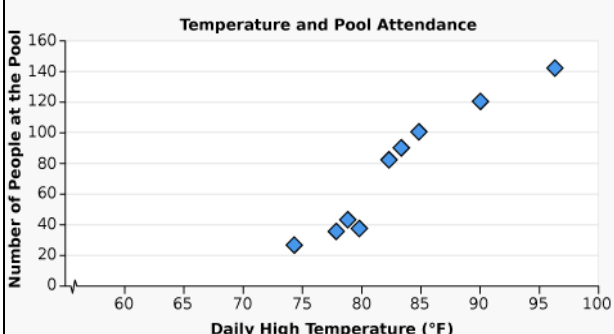
NWEA provides students with multiple ways to respond to questions within the MAP Growth assessments, as shown in Table 3.1. Students either select responses or construct and generate their responses. Figure 3.2 – Figure 3.12 present sample items.

Table 3.1. Item Types

Item Type	Description
Selection (student selects answer option(s))	
Multiple-Choice (Choice)	Students select one response from multiple options.
Multiple Select/Multiselect (Choice Multiple)	Students select two or more responses from multiple options.
Selectable Text (Hot Text)	Students select a response from within a piece of text or a table of information (e.g., word, section of a passage, number, symbol, or equation).
Construction (student constructs the response using provided options)	
Drag-and-Drop	Students select an option or options in an area called the toolbar and move or "drag" these options (e.g., words, phrases, symbols, numbers, or graphic elements) to designated containers on the screen.
Click-and-Pop	Students move options (e.g., words, phrases, symbols, numbers, or graphic elements) from the area called the toolbar to designated container(s) on the screen by selecting an option; the option then "pops" into the container on screen.
Generation (student generates the response with no answer options available)	
Text Entry (short constructed-response)	Students use the keyboard to type their response directly onto the screen in response to a question or prompt.
Item Delivery Mechanism (ways items are presented in addition to standalone)	
Item Set	Students are presented with a set of items that all focus on a single passage or a narrowly defined topic. (Currently used only in MAP Growth Reading and Science. Not used in K–2.)
Composite Items	Students interact with multiple interaction types included within a single item.

Figure 3.2. Sample Item—Multiple-Choice (Mathematics)

The pool manager recorded data on the number of people who went to the pool each day and the high temperature for that day. The data is shown in the scatter plot.



The forecast temperature for next Wednesday is 88 °F.

Which is the best estimate of the number of people who will go to the pool on Wednesday?

A. 160
C. 110
B. 135
D. 85

Figure 3.3. Sample Item—Multiple Select/Multiselect (Reading)

Choose two things Daniel would most likely do at 7:05 A.M.

- [ ] A. go to bed
- [ ] B. eat breakfast
- [ ] C. walk his dog before school
- [ ] D. finish his homework after dinner
- [ ] E. come home from soccer practice

Figure 3.4. Sample Item—Selectable Text (Language Usage)

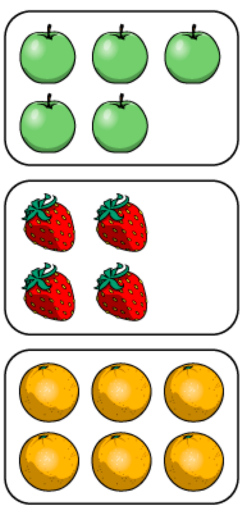
Read the draft of the story. Then, choose the word from each pair that provides the most descriptive detail.

Each Saturday morning, my sister Olivia and I awaited the verdict on our weekly chores. Olivia dreaded getting assigned the job of scrubbing the bathtub. Not only did she find the task **[tedious / ordinary]**, but she somehow always ended up getting totally **[damp / drenched]** when she turned on the shower to rinse the tub.

However, last Saturday was different. Although Olivia got stuck with tub duty again, our brother Max had gotten up early to tackle another chore—cleaning our fish aquarium. When Olivia pulled back the shower curtain to get started, the bathtub was full of tropical fish **[moving / gliding]** around in the temporary home Max had found for them.

Figure 3.5. Sample Item—Selectable Text (Mathematics)

Choose whether the number of objects in each set is odd or even.



Odd Even

Odd Even

Odd Even

Figure 3.6. Sample Item—Drag-and-Drop (Language Usage)

Read the paragraph.

Determine which words are <u>closest</u> in meaning to the words in parentheses and move them to the blanks.

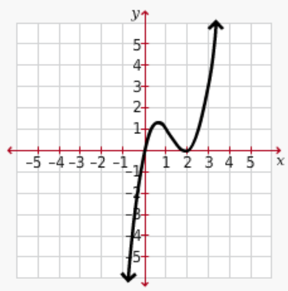
Some of the words will not be used.

The Grand Teton National Park contains mountain views that are so (pretty) ____________ they can take your breath away. Wildlife is everywhere; animals like moose, bison, bears, and even wolves are (common) ____________ sights. The ridges and peaks are (rough) ____________ rather than smooth, showing that the Tetons are relatively new mountains. They are still very high, though; people who climb to the summit will find the air to be very thin.

additional   dignified   essential   frequent   jagged   majestic

Figure 3.7. Sample Item—Click-and-Pop (Mathematics)

Use the graph to complete the task.



Choose <u>all</u> the factors of the polynomial shown in the graph.

x x^2 x^3 (x^2+4) (x^2-4) $(x+2)^2$ $(x-2)^2$

Figure 3.8. Sample Item—Text Entry (Mathematics)

Write 1 hundred $+3$ tens $+2$ ones as a number. Enter the answer in the box.

1 hundred $+3$ tens $+2$ ones $=$ ______

Figure 3.9. Sample Item—Item Set, Multiple-Choice (Reading)

Read the passage. There are several questions about this passage.

Beautiful Invader

1 Imagine yourself taking a walk on a summer day—somewhere in a lazy meadow, near a stream. All along the stream banks and up through the grasses in the meadow, a flowering plant grows from three to ten feet tall. You admire the tiny flowers and their stunning rosy-purple color. You whip out your cell phone and are about to capture a photo when you hear a scolding voice in your head ask: "Why are you about to take a picture of purple loosestrife? It's not something to celebrate. It's an invasive species!"

Purple loosestrife (*Lythrum salicaria*)

2 Purple loosestrife isn't native to North America. It is originally from Europe and Asia. In North America, purple loosestrife grows so thickly and spreads so rapidly that it crowds out native grasses and other flowering plants. Furthermore, wildlife that depends on native plants for food and shelter suffer when purple loosestrife moves in. Because purple loosestrife can destroy the natural balance of an environment, some people believe that we should eliminate this flowering invader.

Which sentence states a central idea in the passage?

1. "Because purple loosestrife can destroy the natural balance of an environment, some people believe that we should eliminate this flowering invader." (Paragraph 2)
2. "Purple loosestrife plants first arrived in the northeastern United States and Canada in the 1800s from Europe." (Paragraph 3)
3. "In some states, it is illegal to buy, sell, plant, or transport the species." (Paragraph 4)
4. "From every new root stem, new plant stalks emerge—each of which produces new flowers and thousands more seeds." (Paragraph 5)

Figure 3.10. Sample Item—Item Set, Multiple Select/Multiselect (Reading)

Read the passage. There are several questions about this passage.

Beautiful Invader

1 Imagine yourself taking a walk on a summer day—somewhere in a lazy meadow, near a stream. All along the stream banks and up through the grasses in the meadow, a flowering plant grows from three to ten feet tall. You admire the tiny flowers and their stunning rosy-purple color. You whip out your cell phone and are about to capture a photo when you hear a scolding voice in your head ask: "Why are you about to take a picture of purple loosestrife? It's not something to celebrate. It's an invasive species!"

Purple loosestrife (*Lythrum salicaria*)

2 Purple loosestrife isn't native to North America. It is originally from Europe and Asia. In North America, purple loosestrife grows so thickly and spreads so rapidly that it crowds out native grasses and other flowering plants. Furthermore, wildlife that depends on native plants for food and shelter suffer when purple loosestrife moves in. Because purple loosestrife can destroy the natural balance of an environment, some people believe that we should eliminate this flowering invader.

The author presents the argument that purple loosestrife is harmful.

Which two details support this argument?

1. "All along the stream banks and up through the grasses in the meadow, a flowering plant grows from three to ten feet tall." (Paragraph 1)
2. "Furthermore, wildlife that depends on native plants for food and shelter suffer when purple loosestrife moves in." (Paragraph 2)
3. "Today, purple loosestrife grows in almost every U.S. state." (Paragraph 4)
4. "Its seeds are small and lightweight." (Paragraph 5)
5. "A breeze or the gentle current of a stream is enough to carry purple loosestrife seeds to new territory where it can vanquish native vegetation." (Paragraph 5)

Figure 3.11. Sample Item—Composite Item (Reading)

Read the passage and answer both questions.

1 When Marco entered the room, he thought everyone would be looking at him. After all, he was the new kid at school. His name was even written on the board at the front of the room: "Welcome, Marco!"

2 He looked around quickly, hoping to spot a friendly face. Instead, no one was looking at all. The other students were busy doing their classwork, and nobody noticed him standing there. The teacher must have stepped out of the room for a minute. Marco hesitated, then sat down at an empty desk next to a boy wearing a blue shirt.

3 The boy stopped writing and looked up at Marco. He smiled. "Hi," the boy said, "I'm Sam."

4 Marco felt relieved. "Hi," Marco answered.

5 Marco's new teacher returned and told him she would get his books for him after lunch. She seemed unsure of what to do with him in the meantime. Sam glanced around the room at the other students. Then Sam grinned and said to Marco, "I can share my book with you for now, if you want."

6 "Great," the teacher said.

7 Marco looked at his teacher and Sam and realized he had found friendly faces after all.

Which word best describes the way Sam, the boy in the blue shirt, acts?

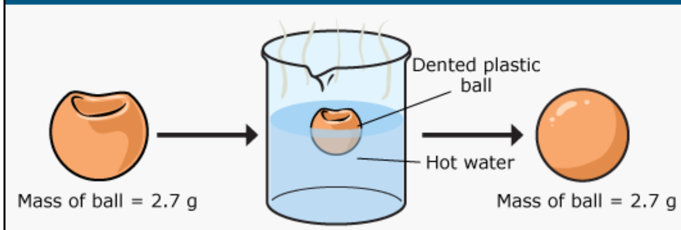
- ○ 1. busy
- ○ 2. careful
- ○ 3. quiet
- ○ 4. thoughtful

Which detail from the passage best supports your answer?

- ○ 1. "Marco hesitated, then sat down at an empty desk next to a boy wearing a blue shirt." (Paragraph 2)
- ○ 2. "The boy stopped writing and looked up at Marco." (Paragraph 3)
- ○ 3. "Sam glanced around the room at the other students." (Paragraph 5)
- ○ 4. "Then Sam grinned and said to Marco, 'I can share my book with you for now, if you want.'" (Paragraph 5)

Figure 3.12. Sample Item—Composite Item (Science)

A student wants to remove a dent from a hollow plastic ball used for table tennis. He reads that table tennis balls are filled with oxygen gas. He decides to put the dented ball into hot water to see what happens. The diagram shows the results.



Which statement explains the results of the investigation? Choose one explanation.

- ○ A. Oxygen molecules inside the ball move farther apart and push out the dent.
- ○ B. Oxygen molecules inside the ball fill with heat, grow larger, and push out the dent.
- ○ C. Hot air molecules enter the ball. The increased number of molecules pushes out the dent.
- ○ D. Hot water molecules enter the ball. The increased number of molecules pushes out the dent.

Which information is evidence that supports this explanation? Choose all the supporting evidence.

- ☐ A. Ball loses its dent.
- ☐ B. Volume of the ball increases.
- ☐ C. Mass of the ball stays the same.
- ☐ D. Ball floats on the surface of the water.

3.2. Item Development Resources

Item development resources include item specifications and cognitive expectation frameworks that provide guidance regarding the content, context, cognitive complexity, and form of items. Content developers are also directed to an external documentation site with access to documents that provide guidance and requirements for the following:

- Item formatting and style
- Item type guidelines for when and how to construct a certain type of item
- Content-area-specific item writing guidelines
- UDL guidelines, including those for bias, sensitivity, fairness, and accessibility
- How to request media for items
- Copyright and permissions guidelines
- Equation descriptions for screen readers

3.2.1. Item Specifications

Item specifications are written to help content developers create items that are aligned to and assess an intended topic or skill. NWEA item specifications include the following elements of guidance for item writers:

- Describe a direct and demonstrable relationship to areas of need
- Unpack an objective into discrete statements when the objective has numerous aspects
- Focus on one topic/skill and indicate a grade or grade range
- Ensure that no relevant skills are overlooked when unpacking an objective
- Match the cognitive complexity of the learning indicator
- Match the content to the item type based on best practices
- Provide guidance around passage/item resource/context when applicable
- Provide parameters, examples, definitions, and resources when applicable
- Provide suggestions on the types of answer choice options (e.g., the options for this item could be charts or graphs) when applicable

Content specialists review each specification for clarity, completeness, and alignment to ensure that content developers will understand the types of items expected. The specifications are reviewed and updated on an ongoing basis.

3.2.2. Cognitive Complexity

Webb's Depth of Knowledge (DOK) and Bloom's revised taxonomy are two different ways of classifying cognitive expectations and are the most commonly used cognitive expectation classifications in education. To ensure that the MAP Growth assessments include a pool of items that span the full range of cognitive levels and skills, content specialists have created cognitive expectation frameworks that define the target DOK for every standard. The cognitive levels are based on three of Webb's DOK categories (1997):

1. Recall and Reproduction
2. Skill/Concept
3. Strategic Thinking and Reasoning

Each item in the pool is evaluated and tagged with a DOK level and one of Bloom's cognitive process dimensions (e.g., remembering, understanding, applying, analyzing) (Anderson &

Krathwohl, 2001, pp. 67–68). Additionally, Mathematics items have been tagged according to Student Achievement Partners' Aspects of Rigor (AOR) model (Achieve, 2018). NWEA content specialists were trained by Student Achievement Partners in January 2019 on how to assign aspects of rigor to test items and have tagged Mathematics items aligned to the CCSS for rigor.

3.3. Item Writing

NWEA is committed to creating items that assess what they are intended to assess, adhere to best practices, and are fair and free from bias. NWEA content specialists fulfill the item writing internally or contract out to freelance content developers, although most items are written by freelance content developers. To begin the process, the NWEA content team creates an item acquisition plan based on an item pool analysis and identified areas of need. Once item assignments are given to the content developers, the developers are provided ongoing guidance and feedback throughout the development process by NWEA content specialists until items are approved. The NWEA content management system enables content developers to submit items directly into the content review work queues. Writers are provided with guides such as item specifications and the item writing guide, as well as ongoing feedback specific to their item-writing assignments.

3.3.1. Freelance Recruitment and Selection

NWEA selects freelance content developers by following a strict vetting process that requires candidates to demonstrate expertise in their content area. NWEA requires that prospective content developers submit sample items in support of evidence in their resumes that they have the relevant content area knowledge, classroom teaching experience, and/or professional assessment writing experience. When there is a need for higher volumes of items, NWEA contracts with established content development vendors whose item samples are rigorously evaluated by NWEA content specialists and copyright and permissions specialists.

3.3.2. Media

If an item needs graphics or audio, the request is sent to the media developers who maintain a set of asset creation guidelines to ensure the clarity and consistency of all media assets and adherence to the following rules:

- The content of the photo or illustration is essential in assessing the context in the item.
- UDL principles are followed.
- Asset requests are fulfilled within the parameters of approved guidelines.
- All media are legible and readable.
- All media adhere to legal usage guidelines.

3.3.3. Metadata

During item construction, metadata fields such as those listed below are added to each item and reviewed. Item metadata define attributes of the item and provide information for systems to include and exclude items from pools as necessary. Metadata are entered and confirmed by content specialists during each stage of item review.

- Scale
- Grade
- Blooms cognitive level
- DOK
- Provisional RIT
- Language
- Legal ownership
- Unit of measure

- Item type
- Scored
- Allowable tools
- Calculator
- Product use
- Excluded market & reason
- Included market & reason
- Test grade start
- Test grade end
- Stimulus code
- Item size exception
- Content area

The metadata inform whether each item is included in an item pool. For example, the “scale” field ensures that systems select only Reading items for Reading tests. For items on the Mathematics and Science tests, metadata fields for allowable tools (e.g., ruler, protractor) and calculator (e.g., basic, scientific) determine which item tools are available during testing. Other metadata such as grade, DOK, and item type are used to inform item development needs and other types of internal analysis.

When passage or graphic assets are associated with an item, content specialists add or confirm element metadata used primarily for internal tracking and analysis purposes. For passages, the element metadata include readability, word count, author, and genre. Additional element data is added by permissions, including disposition, rights status, copyright information, publisher information, and source documentation. For graphic assets, the asset type, file ID, element location, date, and fulfiller identification information is stored for each graphic asset.

3.4. Item Review

Each item in the MAP Growth item pool undergoes the review process summarized below. A minimum of three separate professionals (i.e., two content specialists and a copy edit/quality control specialist) thoroughly review each item. All items (except Mathematics items that only include calculation with no additional context or graphics) undergo a copyright and permissions review. An item can be sent back to a previous stage or rejected if it does not meet the strict standards of NWEA at any point during these reviews.

1. A copyright and permissions specialist ensures that public domain content is from authoritative, authentic sources; that copyrighted texts are approved by the copyright holders; and that content is free of plagiarism.
2. Content specialists ensure that the content is valid and meets the NWEA quality content and alignment standards. Content specialists also validate factual material, ensure that current topics are used, review for bias and sensitivity, and ensure instructional relevance. They also validate the grade appropriateness of the item and assign a DOK level and Bloom’s classification.
3. A content specialist assigns a preliminary difficulty level (i.e., a provisional RIT) to the item for field test purposes.
4. The media developers create any graphics or audio required for an item.
5. A copy editor reviews items for grammar, usage, and mechanics errors and ensures that the items adhere to style guidelines. The item is reviewed for visual bias, and image descriptions (“alt text”) are added to graphics for use by screen readers. Image descriptions may allow students who use refreshable braille and/or screen readers to answer items that would otherwise be inaccessible. They also ensure that items display correctly in all supported browsers.

3.4.1. Copyright and Permissions Review

The copyright and permissions specialist performs the first review once an item or asset has been written and submitted. Subsequent copyright and permissions reviews are performed as needed throughout the item development process when significant revision or new authorship is introduced. The NWEA content management system supports this process by maintaining a historical version of an item each time it is edited and saved. The copyright and permissions specialist ensures the following:

- Item and asset content (i.e., anything added to an item beyond the stem and answer options such as a passage, photograph, illustration, graph, or chart) is free of plagiarism.
- Public domain texts and visual assets (i.e., item or passage art) are selected from authoritative, authentic sources.
- Uses of copyrighted texts and visual assets are approved by the copyright holders.
- All trademark and Right of Publicity requirements are researched and correctly documented.

Plagiarism review is conducted largely through an internet search engine. Phrases, strings of words, and images are searched to ensure that items and item assets are free from plagiarism. Source materials provided by content developers are also reviewed regarding item content. When items or passages are factually based, writers must provide proof of their factual content. For example, Science writers provide URLs to the sources they used. For ELA passages, writers attach documents and/or provide URLs showing where they obtained the information. The permissions team reviews these to make sure the sources have not been plagiarized.

Public domain texts and visual assets are compared to authentic sources found online to ensure accuracy. The permissions and copyright specialist documents sources and proof of public domain status and provides proper citation for the work. Copyrighted texts and assets must be authorized by the copyright holders. For a copyrighted passage text, the copyright and permissions specialist facilitates and negotiates a contractual agreement between NWEA and the copyright holder or an authorized agent, which is then approved by the legal team. The copyright and permissions specialist ensures that NWEA complies with contractually agreed upon publishing requirements and tracks expirations and renewals.

Some copyrighted assets employ licenses that do not require direct contact with copyright holders, such as Creative Commons licensing. In these cases, the copyright and permissions specialist documents the material and legal requirements and ensures that the assets are properly cited and published. The copyright and permissions specialist conducts research to be certain that the party licensing the work is the author or an authorized agent. Materials licensed by users with no apparent connection to the author are not permitted.

Trademark databases, such as USPTO.gov or WIPO.int, are used to ensure that items or assets do not improperly use trademarks or service marks, which can be in the form of words, phrases, symbols, or designs. State laws and other legal resources are consulted to ensure that items do not violate the Right of Publicity (i.e., the legal right for an individual, living or deceased, to control commercial use of their name, likeness, or image). This review only applies to content where people are mentioned or shown.

3.4.2. Content Validation

Concurrently with the copyright and permissions review, items undergo a content validation review performed by a content specialist who determines whether the item content meets the requirements outlined in the item specifications and other item development resources. The NWEA content specialist reviews items for the following:

- Content validity
- Instructional relevance
- Currency
- Alignment to the standard
- Item construction
- Bias, sensitivity, and fairness
- Confirmation that the item passed the copyright and permissions review

The main purpose of content validation is to determine whether a newly submitted item meets basic quality requirements. If the item does not meet the requirements, a content specialist will send the item back to the item writer with a request for revision. At this stage, any revisions made to the item are done by the item writer. Items that meet content validation requirements are approved for payment and moved to the item owner review.

3.4.3. Item Owner Review

During the item owner review, a content specialist performs a thorough in-depth review of the item and makes any further revisions. The content specialist who performs this review is considered the item's "owner" and is contacted if there are any questions about the item as it moves through the rest of the item review process. During this review, items are revised as needed based on a detailed set of criteria developed by NWEA content specialists to confirm that the item is:

- Instructionally relevant and a valid measure of the target concept
- Aligned with clear face validity
- Free of bias, sensitivity, and fairness issues
- Sound in terms of item construction
- At an appropriate reading level so that reading difficulty does not interfere with the concept being assessed
- Accessible for all students according to UDL principles

This determination is also recorded for system use. Content specialists use content area-specific versions of a checklist like Table 3.2 during item owner and content confirmation reviews. Any item with graphical content is also evaluated for visual bias/appropriateness to include on accessible MAP Growth tests. Items are formatted according to the NWEA Formatting and Style Guide, a compilation of style and formatting guidelines. Additional resources used during item owner review to maintain consistency in items are the *Merriam-Webster's Online Dictionary, Chicago Manual of Style,* and *Scientific Style and Format: The CSE Manual for Authors, Editors, and Publishers*, among others. In addition to content-specific reviews, NWEA content specialists also confirm that the functionality of a given item type is used appropriately for an item.

Table 3.2. Item Review Checklist

Content	Edits are made to ensure factual accuracy.
NWEA Style	Edits are made to ensure that the item adheres to the NWEA style guide.
Components	Edits are made to ensure that all required components are included in the item.
Copyediting	Edits are made to ensure correct grammar, spelling, punctuation, capitalization, language usage, and syntax.
Bias/ Sensitivity/ Fairness	Edits are made to ensure that the item meets the following bias, sensitivity, and fairness criteria: • Content is accessible to all students without a need for prior knowledge. • Item avoids bias (e.g., cultural, linguistic, socioeconomic, religious, colorblind, gender, geographical). • Item avoids common issues for ELL students (e.g., idioms, unnecessary phrases, convoluted sentence structure). • Item avoids stereotypes. • Item avoids sensitive topics (e.g., smoking, death, crime, violence, profanity, sex, religion, body/weight issues).
Item Purpose	Edits are made to ensure that an item meets the following criteria: • Item aligns to the standard. • Item is instructionally relevant. • Item is not a trick question. • Concept in item is accurately reflected in item resource (passage/graphic). • Item context is appropriate.
Readability	Edits are made to ensure that the readability of an item, passage, or asset meets the following criteria: • Item uses an appropriate level of vocabulary and readability for the skill level. • Item includes directions and/or introductory text that is clear, appropriate, and useful.
Passage	Edits are made to ensure that passages meet the following criteria: • Passage is relevant, essential, and engaging. • Passage length is within established guidelines for the intended grade. • Passage citation is correct. • Passage has appropriate permissions for use.
Graphics	Edits are made to ensure that graphics meet the following criteria: • Graphics are accurate, relevant, and clear. • Citation is correct. • Graphics include appropriate labels and titles.
Stem	Edits are made to ensure that a stem meets the following criteria: • Stem is focused, concise, and precise. • Stem uses appropriate terminology, vocabulary, wording, and formatting. • Stem is consistent with answer options.
Answer Options	Edits are made to ensure that distractors and/or the key meet the following criteria: • There is only one key (for single-select items) or only one correct set of keys (for multiselect items). • Key is correctly marked for scoring purposes. • Options are independent (e.g., not overlapping, not logical opposites). • Terminology, vocabulary, wording, and formatting are appropriate. • Options are balanced in length, complexity, and grammatical form. • Distractors are plausible. • Key is not cued. • Options are consistent with what the stem is asking.
Functionality	Edits are made to ensure that the functionality meets the following criteria: • Functionality works as intended. • Number of objects allowed in a container is correct. • Size and type of container are correct. • Items scores correctly and as intended.
Overall Appearance	Edits are made to ensure that the overall finished appearance of the item includes UDL considerations such as clear layout and appropriate use of color.

Once the content and formatting review is complete, the content specialist validates the grade appropriateness of the item and assigns a cognitive demand to the item by designating both a DOK level and a Bloom's classification. Additional metadata values are added at this time. The content specialist also writes or confirms the equation description for content written in MathML (an application of XML for describing mathematical notations) so that it can be read by a screen reader for Mathematics and Science items intended for Grades 2–12. Finally, the content specialist assigns the item a preliminary difficulty level (i.e., provisional calibration or provisional RIT) needed for field test purposes. The preliminary difficulty level is based on the observed difficulty of similar items and the content specialist's professional expertise, and it allows items to be chosen for presentation that closely match the student's estimated achievement level. This helps to optimize the use of the student's testing time by presenting items that are neither too difficult nor too easy.

3.4.4. Content Confirmation Review

A second content review is performed by a different content specialist from the same content area. This second reviewer attends to the overall editorial and pedagogical integrity of the item and validates the alignment and cognitive demand designations. The content specialist also verifies that the fields have been set appropriately in the NWEA content management system to ensure that the item is ready for field testing, which includes confirming the equation descriptions for MathML images as needed.

3.4.5. Item Quality Review

During the item quality review, a copy editor reviews each item for syntax, grammar, usage, spelling, and punctuation. The item is reviewed for visual bias, and image descriptions are added to graphics for use by screen readers.[4] Image descriptions may allow students who use refreshable braille and/or screen readers to answer items that otherwise would be inaccessible. They also ensure that items will display correctly in all supported browsers. Finally, an editor validates that the item display and interactions are performing as expected and approves the item for field testing. If at any point changes are required that may impact the content of the item, a content specialist is consulted during this stage of review.

3.4.6. Bias, Sensitivity, and Fairness

NWEA takes seriously the task of creating items that are fair to all students and free from bias and sensitivity issues. All MAP Growth items are reviewed for bias, sensitivity, and fairness. Items are revised to eliminate these issues, or they are rejected when an issue cannot be remedied through the revision process. NWEA defines these three overlapping areas as follows:

- **Bias:** Item content, unrelated to the concept or skill being assessed, that may unfairly influence a student's performance, or an item construct that does not have equivalent meaning for all students.

- **Sensitivity:** The experience of taking a test differs from the classroom experience in that students do not have the opportunity to discuss the material with a teacher or their peers. Without teacher facilitation, sensitive content risks drawing students out of the testing experience by provoking negative emotional responses. A sensitive assessment avoids content that distracts students in this way.

[4] Image descriptions follow the NWEA Image Description Guidelines for Assessments: https://www-cms.nwea.org/content/uploads/2017/06/Image-Description-Guidelines-for-Assessments-2017.pdf

- **Fairness:** Equitable treatment of all test takers during the assessment process, regardless of testing purpose. Fairness should be considered to ensure measurement quality, measurement bias, and access to the construct being assessed. To make a test fair, test developers must work to eliminate any barriers to content for all students. Barriers are factors outside of the knowledge, skill, or ability being assessed that prevent students from understanding and interacting with item content in a manner that accurately demonstrates what they know or are able to do.

The job of an item is to activate a student's thought process and help them focus on the task. A successful item is free of bias and sensitivity issues and is accessible to all students. An item should NOT:

- Distract, potentially upset, or confuse in any way
- Contain inappropriate or offensive topics
- Require construct-irrelevant knowledge or specialized knowledge
- Favor students from certain language communities
- Favor students from certain cultural backgrounds
- Favor students based on gender
- Favor students based on socioeconomic issues
- Employ idiomatic or regional phrases and expressions
- Stereotype certain groups of students or behaviors
- Favor students from certain geographic regions
- Favor students who have no visual impairments
- Use height, weight, test scores, or homework scores as content or data in an item

There is not a rigid list of material that is potentially distracting or upsetting, but some topics are seldom appropriate for K–12 assessments, such as sexuality, illegal substances, illegal activities, excessive violence, discriminatory descriptions, death, grieving, catastrophes, animal neglect or abuse, and loss of a family member.

3.5. Reading Passage Development

Text excerpts are used with MAP Growth Reading items. Some are short passages attached to standalone items, whereas others are extended texts that can support multiple items (i.e., common stimulus passages). To assess students' ability to analyze reading passages in a way that fully integrates the depth and breadth of academic reading standards, students need to engage in close reading of high-quality complex text of various genres and types. Therefore, common stimulus passages are included to address concepts and state standards that require complex texts. Currently, the MAP Growth Reading 2–12 item bank includes approximately 255 common stimulus passages. Of these passages, 45% are commissioned from external content developers, 46% are copyrighted works, and 9% come from the public domain.[5] The MAP Growth Reading K–2 assessment includes very short assets in standalone items and does not have common stimulus passages.

[5] As of April 2018. These numbers are approximate and will change as passages are retired or developed.

A common stimulus passage is presented with a set of several text-based items that require close reading of an extended text. These passages undergo internal and external review by NWEA content specialists, subject matter experts, and members of the permissions, media, and copyediting teams. Because MAP Growth is an adaptive test, the pool of common stimulus reading passages must accommodate a variety of student ability levels. The length of a common stimulus passage varies depending on the targeted grade band. Table 3.3 presents the common stimulus passage word count guidelines by grade. These guidelines apply to prose only. Content specialists use professional judgement when considering appropriate length for poetry and drama. These are guidelines only, and actual passage lengths may be slightly over or under these counts.

Table 3.3. Common Stimulus Passage Word Count Guidelines

Grade	Minimum	Maximum
2	200	450
3	200	650
4	450	750
5	450	750
6	650	950
7	650	950
8	650	950
9	650	1,100
10	650	1,100
11	800	1,100
12	800	1,100

MAP Growth Reading includes both literary and informational texts. Literary texts include a diverse range of fiction and poetry by authors of various cultures and life experiences. Informational texts include literary nonfiction works and works by published authors with expertise in the disciplines of science and humanities. Also included are canonical public domain works of historical and literary significance, as well as technical, functional, and procedural documents.

Alignment criteria for passages are as follows:

- Each common stimulus passage is assigned to a grade based on a careful qualitative and quantitative analysis of text complexity and appropriateness. These grade assignments are recorded in the passage database. Most of the items within a set will align to the grade assigned for the passage. On occasion, an item may instead be aligned to an adjacent grade (off-grade alignment) to ensure a tight standard alignment.
- The following rules are observed:
    - Items connected to highly complex passages may be aligned +1 grade to ensure tight alignment.
    - Items connected to moderately complex passages may be aligned +1 or -1 grade to ensure tight alignment.
    - Items connected to minimally complex passages may be aligned -1 grade to ensure tight alignment.
- Secondary alignments are not used with common stimulus items.

3.5.1. Passage Writer Recruitment and Selection

Some common stimulus passages are commissioned works. Freelance content developers must meet strict qualification requirements and are typically current or retired educators or educational consultants who make their living through freelance opportunities in item or passage writing, curriculum design, and development. All candidates for freelance passage writing undergo a selection process that includes submission of their resume or curriculum vitae and a review of sample passages written to set specifications.

3.5.2. Passage Acquisition and Review Process

Passage acquisition and review for MAP Growth Reading occurs on a continuous basis and follows the process outlined below:

1. Content specialists write passage specifications to garner literary, informational, and persuasive passages, as well as technical, domain-specific, and historical documents. Specifications detail the desired readability, text complexity, word count, and genre.
2. External content developers fulfill passage specifications when submitting commissioned works. NWEA content specialists also conduct focused searches for copyright and public domain diverse literary passages, informational and technical texts, and seminal/historical documents.
3. For commissioned works, content developers send a synopsis of the passage topic to NWEA for preapproval. Before preapproving a topic, content specialists ensure that the topic is age- and grade-appropriate, does not overlap with topics of other passages, and is unlikely to present bias, sensitivity, or fairness concerns. Passage writers/finders submit passage files and relevant source documentation to NWEA.
4. All passages undergo a series of reviews conducted by NWEA copyright and permissions specialists; content specialists; members of an external bias, sensitivity, and fairness panel; and content production specialists. Reviews include the following tasks:

    i. Copyright and permissions specialist verifies that the passage is free of plagiarism (if commissioned) and documents its permissions status (public domain or copyrighted).
    ii. Copyright and permissions specialist ensures that the passage does not have copyright, trademark, or rights of publicity issues.
    iii. Content specialist ensures that the passage meets the specifications and quality requirements and verifies that it meets the text complexity requirements for the grade level and is free of bias, sensitivity, and fairness issues. The content specialist also fact-checks commissioned informational passages.
    iv. Content specialist reviews and revises commissioned passages to ensure accuracy and overall structural and mechanical quality and applies readability analysis to help gauge grade-appropriateness and quantitative text complexity.
    v. All passages are reviewed for bias, sensitivity, and fairness internally and by an external panel of six reviewers from across the U.S. that is trained to implement internal NWEA bias, sensitivity, and fairness guidelines. Panelists complete a checklist for each passage to record their recommendations and meet online when needed.
    vii. Content production specialists perform a final copyedit of commissioned passages to ensure that the passages conform to both NWEA-specific and publishing industry styles.

When evaluating texts, content specialists apply the following criteria:

- Expert and credible authorship: Does the author write with authority about the topic? What are the author's journalistic and academic credentials? Does the author have an authentic connection to the culture depicted in the work?
- Text worthy of study: Is the work well crafted? Does it lend itself to close reading and analysis? Does it contain a clear central idea, relevant evidence, opportunities for reasoning, concrete details, an effective structure, and rich and varied language?
- Text not widely taught: Is the text one that students are unlikely to have encountered in the classroom?
- Free of bias and sensitivity concerns: Does the text present people fairly, respectfully, and without stereotype?
- Engaging and appropriate for target readers: Is the topic and tone of the writing likely to appeal to students?
- Ideal for assessment: Does the text yield a variety of challenging, standards-aligned items?

3.6. Text Readability

The expected readability of text in items is specific to the item scale. In Mathematics and Science, item readability is kept to two grade levels below the grade of the content being assessed to avoid inadvertently assessing a student's reading skills rather than their mathematical or science skills.

NWEA content specialists evaluate the readability of passages and scenarios in Science item sets using both quantitative and qualitative measures. Passages within a grade level are assigned a range of complexity: minimally complex, moderately complex, and highly complex. Table 3.4 presents the quantitative and qualitative analyses conducted for passages.

Table 3.4. Quantitative and Qualitative Analyses

Quantitative Analysis	• Research-based recommendations highlight the use of two or more quantitative text analyzers/readability measures. • NWEA captures several quantitative readability scores (e.g., Lexile, Flesch-Kincaid, and Coh-Metrix) for each passage. • While variation exists among text analyzers, no single measure is interpreted to outperform the others.
Qualitative Analysis	• Qualitative dimensions of a work are evaluated for developmental appropriateness, cognitive difficulty, and intended audience. • NWEA has developed an internal rubric used to evaluate passages on such criteria as Levels of Meaning, Structure, Language Convention and Clarity, and Knowledge Demand. • Qualitative analysis includes how information and ideas are communicated implicitly, such as through literary techniques like allusion or analogy. Also evaluated are reader's purpose, type of reading (surface level or deep analysis), and intended outcome (knowledge, solution, engagement, assessment).

3.7. Field Testing

Field testing is required to maintain the item bank as existing items are retired or removed due to changes in standards or item parameter drift. All newly developed items are field tested by embedding them in an operational testing environment instead of as standalone field tests to reduce the amount of testing time and encourage students to respond to field test items with as much effort as they would operational items. Field test item responses are not included in a student's final score. The purpose of field testing is to use the item response data to analyze the quality of the field test items and incorporate them into the RIT scales. Field test results presented within a set of calibrated items are used to analyze and calibrate the difficulty estimate for each new item to the existing scale. Successfully calibrated field test items are added to the item banks as operational items. Once this empirical information is collected, the provisional difficulty estimate is retired. Only information from student samples is used from that point on. Items that fail to meet quality standards are reviewed and either revised and returned to field testing or rejected altogether.

Each item is administered to a sample of at least 1,000 students, although Ingebo (1997) has shown that a sample size of 300 is adequate for accurate item calibrations. Finally, the environment for data collection should be free from the influence of other confounding variables such as cheating or fatigue. Since the field test data are collected within the normal operational test administration process designed to equalize or minimize the impact of outside influences, the environment is optimal for data collection. The items are administered to sizable samples of students, and the field test data are collected in a manner that motivates the students to work seriously in an environment free from external influences on the data.

3.8. Statistical Summary of the Item Pools

Table 3.5 presents the content structure of the MAP Growth item pools available for use with the CCSS and NGSS, including the number of items in the item pools and the average difficulty and standard deviation (SD) of the items by sub-area. These large MAP Growth item pools allow the assessments to provide accurate achievement estimates for students in each content area across all grade levels.

Table 3.5. MAP Growth Content Structure for use with CCSS and NGSS

Instructional Area	Sub-Area	N	RIT Mean	RIT SD
Reading 2–5				
Informational Text: Key Ideas and Details	Draw Conclusions, Infer, Predict	457	196.9	16.8
	Summarize; Analyze Central Ideas, Concepts and Events	255	204.7	13.8
	Overall	712	199.7	16.2
Informational Text: Language, Craft, Structure	Point of View, Purpose, Perspective, Figurative and Rhetorical Language	217	207.1	13.6
	Text Structures, Text Features	214	201.9	16.5
	Overall	431	204.5	15.3
Literary Text: Key Ideas and Details	Draw Conclusions, Infer, Predict	474	191.1	16.2
	Summarize; Analyze Themes, Characters, Events	403	201.3	15.6
	Overall	877	195.8	16.7
Literary Text: Language, Craft, Structure	Figurative, Connotative Meanings; Tone	223	199.7	15.1
	Point of View, Purpose, Perspective	77	207.6	10.4
	Text Structures, Text Features	85	206.2	15.2
	Overall	385	202.7	14.7
Vocabulary: Acquisition and Use	Context Clues	403	199.5	13.7
	Reference and Word Parts; Academic Vocabulary	538	194.4	18.5
	Word Relationships and Nuance	165	194.6	21.1
	Overall	1,106	196.3	17.5

Instructional Area	Sub-Area	N	RIT Mean	RIT SD
Reading 6+				
Informational Text: Key Ideas and Details	Draw Conclusions, Infer, Predict	515	205.1	16.1
	Summarize; Analyze Central Ideas, Concepts and Events	381	213.6	14.7
	Overall	896	208.7	16.1
Informational Text: Language, Craft, Structure	Point of View, Purpose, Perspective, Figurative and Rhetorical Language	365	215.8	14.8
	Text Structures, Text Features	275	209.2	16.6
	Overall	640	213.0	15.9
Literary Text: Key Ideas and Details	Draw Conclusions, Infer, Predict	467	199.3	17.2
	Summarize; Analyze Themes, Characters, Events	526	210.5	16.5
	Overall	993	205.2	17.7
Literary Text: Language, Craft, Structure	Figurative, Connotative Meanings; Tone	339	210.3	17.6
	Point of View, Purpose, Perspective	124	215.8	12.8
	Text Structures, Text Features	123	217.7	13.2
	Overall	586	213.0	16.1
Vocabulary: Acquisition and Use	Context Clues	476	204.9	15.8
	Reference and Word Parts; Academic Vocabulary	516	202.0	16.9
	Word Relationships and Nuance	170	202.7	21.5
	Overall	1,162	203.3	17.2
Reading K–2				
Foundational Skills	Phonics and Word Recognition	736	149.6	14.2
	Phonological Awareness	318	154.9	10.5
	Print Concepts	238	138.5	8.1
	Overall	1,292	148.9	13.5
Language and Writing	Capitalize, Spell, Punctuate	217	163.9	14.8
	Language: Grammar, Usage	264	164.9	15.5
	Writing Purposes: Plan, Develop, Edit	51	175.5	13.8
	Overall	532	165.5	15.4
Literature and Informational	Informational Text: Key Ideas, Details, Craft, Structure	241	172.3	17.9
	Literature: Key Ideas, Craft, Structure	389	163.6	17.4
	Overall	630	166.9	18.1
Vocabulary Use and Functions	Language: Context Clues and References	171	167.5	13.6
	Vocabulary Acquisition and Use	273	152.2	21.9
	Overall	444	158.1	20.6
Language Usage 2–12				
Language: Understand, Edit for Grammar, Usage	Parts of Speech	720	191.6	19.7
	Phrases, Clauses, Agreement, Sentences	467	197.5	18.6
	Overall	1,187	193.9	19.5
Language: Understand, Edit for Mechanics	Capitalization	243	190.5	15.6
	Punctuation	673	199.8	17.7
	Spelling	303	193.8	18.0
	Overall	1,219	196.4	17.8
Writing: Write, Revise Texts for Purpose and Audience	Establish and Maintain Style: Use Precise Language	316	212.1	13.9
	Plan, Organize; Create Cohesion, Use Transitions	588	208.1	14.1
	Provide Support; Develop Topics; Conduct Research	388	211.3	15.2
	Overall	1,292	210.0	14.5

Instructional Area	Sub-Area	N	RIT Mean	RIT SD
Mathematics 2–5				
Geometry	Reason with Shapes, Attributes, & Coordinate Plane	384	190.9	24.8
	Overall	384	190.9	24.8
Measurement and Data	Geometric Measurement and Problem Solving	860	207.3	22.6
	Represent and Interpret Data	289	187.9	23.3
	Overall	1,149	202.4	24.3
Number and Operations	Number and Operations - Fractions	558	219.1	18.7
	Number and Operations in Base Ten	494	204.9	19.6
	Understand Place Value, Counting, and Cardinality	592	190.6	23.6
	Overall	1,644	204.6	24.0
Operations and Algebraic Thinking	Analyze Patterns and Relationships	231	220.8	15.5
	Represent and Solve Problems	898	196.8	21.5
	Overall	1,129	201.7	22.6
Mathematics 6+				
Geometry	Congruence, Similarity, Right Triangles, & Trig	347	243.0	23.0
	Geometric Measurement and Relationships	1,203	217.2	31.0
	Overall	1,550	223.0	31.3
Operations and Algebraic Thinking	Expressions and Equations	1,177	233.2	26.0
	Use Functions to Model Relationships	480	247.2	22.0
	Overall	1,657	237.2	25.7
Statistics and Probability	Interpreting Categorical and Quantitative Data	476	207.8	29.3
	Using Sampling and Probability to Make Decisions	247	230.2	19.5
	Overall	723	215.5	28.4
The Real and Complex Number Systems	Extend and Use Properties	930	206.2	30.1
	Perform Operations	1,721	207.7	23.8
	Ratios and Proportional Relationships	644	222.5	16.2
	Overall	3,295	210.2	25.3
Mathematics K–2				
Geometry	Reason with Shapes and Their Attributes	360	153.8	27.5
	Overall	360	153.8	27.5
Measurement and Data	Represent and Interpret Data	93	165.7	27.5
	Solve Problems Involving Measurement	258	173.3	28.7
	Overall	351	171.3	28.6
Number and Operations	Number and Operations: Base Ten and Fractions	143	186.3	15.5
	Understand Place Value, Counting, and Cardinality	313	144.0	16.8
	Overall	456	157.3	25.6
Operations and Algebraic Thinking	Properties of Operations	209	170.5	19.3
	Represent and Solve Problems	253	166.1	22.4
	Overall	462	168.1	21.2

Instructional Area	Sub-Area	N	RIT Mean	RIT SD
Science 3–5				
Earth and Space Science	Earth and Human Activity	94	202.2	17.7
	Earth's Place in the Universe	140	206.1	15.0
	Earth's Systems	236	204.0	16.4
	Overall	470	204.3	16.3
Life Science	Ecosystems: Interactions, Energy, and Dynamics	111	205.4	12.3
	From Molecules to Organisms: Structures and Processes	122	195.3	17.1
	Heredity: Inheritance and Variation of Traits; Biological Evolution: Unity & Diversity	171	193.1	14.8
	Overall	404	197.1	15.8
Physical Science	Energy; Waves and Their Applications in Technologies for Information Transfer	183	198.3	13.3
	Matter and Its Interactions	122	207.9	16.3
	Motion and Stability: Forces and Interactions	112	198.5	14.5
	Overall	417	201.2	15.1
Science 6–8				
Earth and Space Science	Earth and Human Activity	135	214.9	12.2
	Earth's Place in the Universe	180	209.8	12.9
	Earth's Systems	298	211.5	13.1
	Overall	613	211.7	12.9
Life Science	Ecosystems: Interactions, Energy, and Dynamics	214	210.4	11.6
	From Molecules to Organisms: Structures and Processes	278	211.7	17.2
	Heredity: Inheritance and Variation of Traits; Biological Evolution: Unity & Diversity	291	207.6	18.5
	Overall	783	209.8	16.5
Physical Science	Energy; Waves and Their Applications in Technologies for Information Transfer	240	211.0	15.0
	Matter and Its Interactions	226	217.8	16.0
	Motion and Stability: Forces and Interactions	166	206.1	16.0
	Overall	632	212.2	16.3
Science 9–12				
Earth and Space Science	Earth and Human Activity	111	215.4	11.3
	Earth's Place in the Universe	129	212.8	13.0
	Earth's Systems	259	211.9	11.9
	Overall	499	212.9	12.1
Life Science	Ecosystems: Interactions, Energy, and Dynamics	229	213.1	12.2
	From Molecules to Organisms: Structures and Processes	250	216.6	14.1
	Heredity: Inheritance and Variation of Traits; Biological Evolution: Unity & Diversity	167	219.7	12.8
	Overall	646	216.2	13.3
Physical Science	Energy; Waves and Their Applications in Technologies for Information Transfer	165	218.2	13.5
	Matter and Its Interactions	233	223.0	14.9
	Motion and Stability: Forces and Interactions	128	215.8	13.5
	Overall	526	219.8	14.4

Chapter 4: Test Administration and Security

MAP Growth assessments are fully adaptive, and each student experiences a unique test based on their responses to each item. MAP Growth 2–12 assessments are untimed and take approximately one hour per content area. MAP Growth K–2 assessments are also untimed, and students typically take less than 30 minutes per content area. MAP Growth can be administered up to four times a year (fall, winter, and spring, with a fourth optional administration in summer). A MAP Growth administration requires a proctor computer that allows the proctor to monitor and control the student testing, as well as student devices with a lockdown browser. There are three main steps to testing:

1. Proctor creates a testing session.
2. Students sign in so they can join the testing session the proctor started.
3. Proctor supervises students and assists them with things like pausing and resuming their test if needed.

The NWEA test delivery platform supports more than 60 million student test events each year. The platform has delivered uninterrupted service with 172,000 students actively testing, defined as “concurrent” users. The most recent configuration has been certified and tested for at least 300,000 concurrent users.

4.1. Adaptive Testing

The MAP Growth adaptive testing algorithm starts item selection using items with RITs that are as suitable as possible for a student’s abilities based on known information about the student (e.g., grade level, prior RIT scores). If the student answers the item correctly, they receive a more difficult item. An incorrect response prompts an easier item. Maximum Fisher’s information method is used for item selection coupled with a random-like exposure control procedure that selects one out of a few items that can provide the most information about the student (Kingsbury & Zara, 1989).

To ensure test content validity and the comparability of different tests, a content-balancing procedure proposed by Kingsbury and Zara (1991) and commonly used in most adaptive tests is used. This content-balancing algorithm selects items from the most underrepresented content area according to its target administration value specified in the test blueprint. That is, once an item is administered by maximum information at the student’s current ability estimate, its content classification is evaluated against target values defined in advance in the test blueprint for each student. If the selected item represents a content area that is the least represented at that stage, this item is administered. The maximum likelihood estimation (MLE) method is used for final ability estimation.

Test length varies for different content areas. Tests terminate either when the maximum test length is reached or when final RIT scores meet the pre-specified measurement precision level. Struggling students who might otherwise get frustrated and stop trying and high-achieving students who might get bored by strictly grade-level assessments will remain interested as subsequent items adapt to their abilities.

4.2. Test Engagement Functionality

When students are motivated to perform on tests, they tend to do better and the results are more likely to accurately reflect what they know and can do. In 2017, NWEA introduced the test engagement capability that detects in real-time when a student is "rapid-guessing" on items and notifies proctors so they can re-engage the student with the test. In July 2018, NWEA added a rule that invalidates tests when students show disengaged responses on 30% or more of items. A summary of the test engagement functionality is as follows:

- Students receive a message at the start of the test encouraging them to remain engaged.
- When students rapid-guess, proctors are notified and the test auto-pauses so the proctor can re-engage the student and resume the test.
- MAP Growth invalidates tests when students rapid-guess on 30% of the total number of test items, at which point the test ends in order to protect instructional time.
- To better support retesting processes, educators, including proctors, have access to reports showing students with invalidated tests due to excessive rapid guessing.

MAP Growth employs a sophisticated method for stabilizing testing accuracy when a student disengages. The average amount of time that students take to answer each unique test item is used to determine if a student has rapid-guessed when answering an item. After a student rapid-guesses one item, the difficulty of the next item locks to the same level of difficulty to prevent this downward drift. After the student has rapid-guessed three items in a row, the proctor is notified so that they can intervene and re-engage the student. The data from this test event then shows in reporting the percentage of the assessment that the student rapid-guessed and the estimated impact the disengagement could have had on the student's overall RIT score.

4.3. User Roles and Responsibilities

Access to the MAP Growth system is based on multiple defined roles, as described in Table 4.1. Each role in the system has specific permissions that control levels of access to implementation, configuration, data management, testing, and reporting tasks. Each user has a unique user name to which one or more roles can be assigned. For added security, the system requires manual steps to set up user accounts and authorization levels. Only users with data administrator or proctor permissions can create or modify student profiles. This limits the ability to change student information (e.g., demographics and class assignments) to authorized users who support roster preparation or test proctoring.

Table 4.1. User Roles in the MAP Growth System

Role	Permissions & Responsibilities
System Administrator	• Assign MAP Growth roles for any user, including themselves. • Add or edit users in MAP Growth and reset user passwords. • Modify MAP Growth preferences for the organization. • Mark the test window complete.
District Assessment Coordinator	• Assign MAP Growth roles for any user except System Administrator. • View operational reports. • Add or edit users in MAP Growth and reset user passwords. • Modify MAP Growth preferences for the organization. • Mark the test window complete.

Role	Permissions & Responsibilities
Data Administrator	• Assign MAP Growth roles for any user, except System Administrator or District Assessment Coordinator. • View operational reports. • Add or edit users in MAP Growth and reset user passwords. • Add or edit students. • Import student/staff roster. • Add or edit students in MAP Growth, including permission to merge students and exclude or assign test events.
District Proctor	• Proctor any students within the district. • Set up and conduct student testing. • Add or edit students in MAP Growth.
Administrator	• Limited to assigned schools, will likely be a school principal or vice principal. • View student and class reports. • View reports for the school.
School Assessment Coordinator	• Limited to assigned school(s). • Edit students in MAP Growth.
School Proctor	• Proctor any students in assigned school(s). • Set up and conduct student testing.
Interventionist	• Limited to assigned schools, this is likely a special education teacher or similar role. • View students within their school and add them to custom groups for instruction and reporting.

4.4. Administration Training

Administration training is provided as part of the professional learning services provided by NWEA that includes in-person and online training professional development sessions. The process begins with a consulting session with an NWEA Professional Learning Consultant. NWEA then recommends four days of onsite professional learning, beginning with MAP® Growth™ Administration, Applying Reports, and MAP® Skills™ Basics workshops. During these sessions, educators learn to use MAP Growth; access, interpret, and apply MAP Growth data; and use the data to inform ongoing work, including goal-setting with students. An online MAP Growth administration workshop is also available that involves two three-hour sessions with 40 participants each who learn about administering the tests, accessing reports, and applying data.

4.5. Practice Tests

Practice tests are available online for students to familiarize themselves with the assessment. They provide the same access and functionality as the real MAP Growth tests. Students are encouraged to use the embedded universal tools or a designated feature or accommodation, if needed. To take the practice tests, users must enter a generic username and a password that determines which practice tests the user will have access to. For MAP Growth tests, the username and password are both “grow.” Practice tests specifics are as follows:

- Not adaptive
- No score
- No proctor control
- Available in any supported browser and any supported device
- Available for multiple grades and content areas
- About five items depending on the grade

4.6. Accessibility and Accommodations

MAP Growth has several features to improve test fairness and provide more precise and valid assessment measurement. These features fall within three categories:

- Universal features
- Designated features
- Accommodations

Local schools and districts may determine whether certain features are considered universal, designated, or an accommodation. Schools and districts are encouraged to follow their current state accessibility and accommodation guidelines when deciding which features are appropriate for an individual student. The policy at NWEA is aligned with the CCSSO Accessibility Manual (CCSSO, 2016). The goal is to provide a universal approach and make the use of features and accommodations as easy as possible for both the student and educator.

4.6.1. Universal Features

Table 4.2 presents the available universal features for MAP Growth. Universal features are accessibility supports that are available to all students as they access instructional or assessment content. They are either embedded and provided digitally through instructional or assessment technology (such as a keyboard) or non-embedded and provided non-digitally at the local level (such as scratch paper).

Table 4.2. Available Universal Features

Feature	Description
Embedded	
Amplifications	A student raises or lowers the volume control, as needed, using headphones.
Calculator	A student can access an on-screen digital calculator for calculator-allowed items. If the calculator is not appropriate (e.g., for a student who is blind), the student may use a calculator provided with assistive technology devices (such as a talking calculator or a braille calculator).
Highlighter	A student can mark desired text, items, or response options with a color.
Zoom	A student can increase the size of text and pictures onscreen.
Line reader	A student can use this tool as a guide when reading text.
Answer choice eliminator	A student can cross out answer choices that do not appear to be correct.
Notepad	A student can make notes or record responses virtually.
Keyboard navigation	A student can navigate through test content by using the keyboard (e.g., the arrow keys). This feature may differ depending on the testing platform.
Non-Embedded	
Breaks (frequent breaks)	A student can take breaks, when needed, to reduce cognitive fatigue.
English dictionary	A student can use an English dictionary, if necessary.
Noise buffer (headphones, audio aids)	A student can use noise buffers to minimize distractions or filter external noises during testing. Noise buffers must be compatible with the requirements of the test.

Feature	Description
Scratch paper	A student can use scratch paper or an individual erasable whiteboard to make notes or record responses. The school must also provide a marker, pen, or pencil. All scratch paper must be collected and securely destroyed at the end of each test to maintain test security. The student can use an assistive technology device to take notes instead of using scratch paper if the device is approved by the state. Test administrators must ensure that all notes taken on an assistive technology device are deleted after the test.
Spanish dictionary	A student can use a Spanish dictionary, if necessary.
Thesaurus	A student can use a thesaurus containing synonyms of terms.

4.6.2. Designated Features

Table 4.3 presents the designated features available for MAP Growth. Designated features are available when an educator (or team of educators including the parents/guardians and the student, if appropriate) indicates that there is a need for them. Designated features must be assigned to a student by trained educators or teams using a consistent process. Embedded designated features such as text-to-speech (TTS) are provided digitally through instructional or assessment technology. Non-embedded designated features (such as a magnification device) are provided locally.

Table 4.3. Available Designated Features

Feature	Description
Embedded	
Text-to-speech (TTS) (audio support, spoken audio)	A student can hear audio of the item content.
Non-Embedded	
Bilingual dictionary (word-to-word dictionary in English and native language)	A student can use a bilingual/dual language word-to-word dictionary as a language support.
Color contrast	A student can display the test content of online items in different colors.
Human reader	A qualified human reader can read the test and item content out loud.
Magnification device (low-vision aids)	A student can adjust the size of specific areas of the screen (e.g., text, formulas, tables, and graphics) with an assistive technology device. Magnification allows the student to increase the size to a level that is not provided by the zoom universal feature.
Native language translation	A test administrator who is fluent in the student's native language can translate test and question content.
Separate setting (alternate location)	A school can alter a test location so that the student is tested in a setting that's different from what's available for most students.
Student reads test aloud	A student can read the test content aloud. This feature must be administered in a one-on-one test setting.

4.6.3. Accommodations

Table 4.4 presents the accommodations available for MAP Growth. Accommodations are changes in procedures or materials that ensure equitable access to instructional and assessment content and generate valid assessment results for students who need them. Embedded accommodations are provided digitally through instructional or assessment technology. Non-embedded accommodations (such as a scribe) are provided locally.

Accommodations are generally available to students for whom there is a documented need on an Individualized Education Program (IEP) or 504 accommodation plan, although some states also offer accommodations for ELLs.

Table 4.4. Available Accommodations

Accommodation	Description
Non-Embedded	
Abacus (individual manipulatives)	May be used in place of scratch paper for students who typically use an abacus.
Assistive technology (alternate response options, word processor, or similar keyboarding device to respond to items)	A student can use assistive technology, which includes supports such as typing on customized keyboards; assistance with using a mouse, mouth or head stick, or other pointing devices; sticky keys; touch screen; and trackball.
Calculator (calculation device)	A student can use a specific calculation device (e.g., large key, talking, or other).
Extended time	Schools can allow flexible scheduling for a student test administration (e.g., testing longer than a scheduled test session, multiple breaks)
Human signer (sign language, sign interpretation of test)	A test administrator who is fluent in the language can sign test and item content. The student may also dictate responses by signing.
Multiplication table	A student can use a paper-based single digit (1–9) multiplication table.
Refreshable braille	A student can use a refreshable braille device that provides a raised-dot code that they can read with their fingertips.
Screen reader	A student with no or low vision can use a software application that identifies and interprets what is being displayed on the screen (e.g., text, images).
Scribe	A student can dictate their responses to an experienced educator who records verbatim what the student dictates.

4.6.4. Third-Party Assistive Software

Third-party software features such as those in Table 4.5 are allowed when not using the lockdown browser. If students try using these tools with the lockdown browser, they will have limited or no functionality. Therefore, NWEA recommends that students who need to use specific features use browser-based testing. If students use the lockdown browsers, NWEA recommends they launch the third-party tool prior to launching the lockdown browser.

Table 4.5. Third-Party Assistive Software

Third-Party Software	Description
ZoomText	A powerful computer access solution designed for the visually impaired. It offers a combination of magnification and reading tools, as well as enhancements to colors, pointers, and cursors. It works for both Mac® and Windows® operating systems.
Chromebook magnification	Chromebook has a built-in screen magnifier. This allows users to zoom in and out anywhere on the screen.
Windows magnifier	The magnifier in Windows is part of the Ease of Access Center and can be used to enlarge different parts of the screen. Windows 7 and 8 users can choose from either full screen or lens magnification modes.
Zoom on Mac and iPad	Mac computers and iPads have a built-in screen magnifier that can magnify a screen up to 40 times its normal display size.

Third-Party Software	Description
Chromebook color contrast	High contrast mode inverts the picture so that a white background appears black, black text appears white, and colors are inverted (for example, blue text or graphics become orange).
Windows color contrast	Windows supports high contrast themes for the OS and apps that users may choose to enable. High contrast themes use a small palette of contrasting colors that makes the interface easier to see.
Mac and iPad color contrast	Increase the readability of the screen on your MacBook or iPad by increasing the contrast of the display. Increase the contrast of the whole screen or emphasize borders between items in the Display section of the Accessibility settings.
JAWS	Job Access with Speech (JAWS) is the world's most popular screen reader, developed for computer users whose vision loss prevents them from seeing screen content or navigating with a mouse. JAWS provides speech and braille output for the most popular computer applications.
Refreshable braille device	A refreshable braille device provides a raised-dot code that individuals read with their fingertips.

4.7. Test Security

Inadequate security procedures pose a risk to assessment systems. Violations of test security may compromise the integrity of results and call into question the trustworthiness of information. A common criticism of test security relative to adaptive tests is that some tests do not use sufficiently large item pools to ensure that content on the test cannot be "poached" by groups of students or educators who memorize, compile, and share large numbers of items. However, well-designed, adaptive tests such as MAP Growth that draw from large item pools offer several advantages for ensuring test and item security. The MAP Growth systems leverage the following inherent security advantages:

- A group of students within a classroom or computer lab is likely to view hundreds of different items in any single administration of the test, making it unlikely that students will see the same content at the same time or see items used as examples in a classroom.
- Once a student has viewed an item, they will not see that item again for at least two more terms.
- Large item pools allow minor security breaches to be addressed by removing exposed items from the pool.
- Students within a program can easily be retested using a new set of items if there are questions about the integrity of their scores.

Other test security guidelines followed by NWEA include the following:

- When a student logs into a test session, the test is not started and no test items are made visible to the student until the proctor has confirmed the student and activated the test session by using the proctor dashboard.
- Item responses are not stored/cached locally. Responses are captured in real-time and stored in secure servers before presenting the next item to the student.
- A lockdown browser prevents students from initiating other browser sessions and having access to other content on the testing device unless they exit the test.

Furthermore, the processes and tools provided in Table 4.6 are used to ensure the integrity of the tests were not jeopardized, thereby providing educators and students a positive and reliable user experience.

Table 4.6. Test Security Before and During Testing

Before test administration	• Rostering of student and educator data through secure system applications. • Only specific user roles, approved and authorized within the district and school, can log into the system to access test administration features. • All testing devices are prepared with installing the secure testing browser/app.
During test administration	• Only approved and authorized proctor roles can start the test by providing a secure test session key for all students in the testing lab/classroom. The proctor has the control to start, pause, and resume testing for all students in the classroom or individual students if necessary. • Student test taking is possible with secure testing browser. • There is a district configuration that can be set to prevent retesting. • If students require any testing accommodations such as TTS, proctors can assign those specific accommodations to students based on their IEP/504 needs and ensure appropriate device setup for those tests (e.g., ear phone for TTS). • Student test-taking is only allowed during the testing window. All tests are closed and access removed upon the close of testing window.

4.7.1. Assessment Security

All MAP Growth data transmissions (i.e., testing and response data) are encrypted and secured using TLS 1.2 AES 256 encryption methods. Test data is stored in highly secure Tier 3 data centers located in the continental U.S. operating with redundant power, internet, and backup systems powered by diesel generators. All servers, disk storage, and network infrastructure within each data center are redundant, protecting against unavailability due to a single hardware failure. NWEA operates two geographically disparate data centers with data replication for failover if one data center becomes inoperable. Personally identifiable student information is encrypted at rest in the systems. More information on NWEA Information Security can be found at https://legal.nwea.org/map-growth-information-security-whitepaper.html.

4.7.2. Role-Based Access

Access management is a critical function for maintaining test security. MAP Growth uses role-based access security controls that allow partners to segregate duties in their MAP Growth accounts and grant only the amount of access to users needed to perform their jobs. This allows partners to control what actions and data individuals have access to. When planning partners' access control strategy, MAP Growth supports granting users the least privilege to perform their work. Each role in MAP Growth has specific permissions that control levels of access to implementation, configuration, data management, testing, and reporting tasks. Each user has a unique username to which one or multiple roles can be assigned. Only certain roles can create or modify student profiles, which limits the ability to change student information. More information on NWEA MAP Growth Roles and Responsibilities can be found at https://teach.mapnwea.org/impl/QRM2_Roles_and_Responsibilities_QuickRef.pdf.

Chapter 5: Test Scoring and Item Calibration

MAP Growth items are administered sequentially, with each item being selected to yield maximum information about the student's ability. Individual tests are constructed based on the student's performance while responding to items constrained in content to a set of standards. All MAP Growth items are dichotomously scored. MAP Growth results, reported as RIT scores with a range from 100 to 350, relate directly to the RIT vertical scale, an equal-interval scale that is continuous across grades. Each content area has a unique content-specific scale (i.e., there is one RIT scale each for Reading, Language Usage, Mathematics, and Science), meaning that scores cannot be compared across content areas. Using the RIT scale to report test results makes it possible to follow a student's proficiency status across time, interpreted as growth, across administrations and years. This also allows longitudinal comparison of student performance to be made. This chapter describes the practices surrounding the RIT scale with particular attention to scoring, norming, and item calibration.

5.1. Rasch Unit (RIT) Scales

Development of the RIT scale was guided by item response theory (IRT) that rests on the relationship between student achievement and item characteristics (Lord & Novick, 1968; Lord, 1980; Rasch, 1960/1980). A benefit of using an IRT model is that student scores and item difficulties are on the same scale. The scale is equal interval in the sense that the difference between any two student scores is the same regardless of item difficulty. The same is true for the difference between any two item difficulties. The difference is constant throughout the scale.

Specifically, MAP Growth assessments use the one-parameter Rasch IRT model that estimates the probability (P_{ij}) that a student (*j*) with an achievement score of θ_j will correctly answer a test item (*i*) of difficulty δ_i. It is expressed as:

$$P_{ij} = \frac{e^{(\theta_j - \delta_i)}}{1 + e^{(\theta_j - \delta_i)}}. \quad (5.1)$$

The values of the achievement score and item difficulty in Model 5.1 are on the logit metric, an arbitrary scale commonly used for academic studies of the Rasch model. To allow the MAP Growth measurement scale to be easily used in educational settings, the following linear transformation of the logit scale is performed to place it onto the RIT scale developed by NWEA for use in all MAP Growth tests:

$$RIT = (\theta_j \times 10) + 200. \quad (5.2)$$

The RIT scale ranges from 100 to 350 and is not easily mistaken for other common educational measurement scales. The RIT scale, like other IRT measurement scales, has several useful properties when applied and maintained properly. The most important properties for the development of the measurement scales and item banks include the following, which have been empirically verified for the RIT scales (Ingebo, 1997) and can be used in a variety of test development and delivery applications:

- Item difficulty calibration is sample free (i.e., if different sets of students who have had an opportunity to learn the material answer the same set of items, the resulting difficulty estimates for an item are estimates of the same parameter that differ only in the precision of the estimate's value). The accuracy will differ due to the sample size and the relative achievement of the students compared to the difficulty of the items.
- Trait score estimation is sample free (i.e., if different sets of items are given to a student who had an opportunity to learn the material, the scores are estimates of the same student trait level). Again, precision may differ due to the number of items administered and the relative difficulty of the items compared to the student's level of achievement.
- The item difficulty values define the test characteristics. This means that once the difficulty estimates for the items to be used in a test are known, the precision and the measurement range of the test are determined.

Since IRT enables the administration of different items to different students while allowing for comparable results, the development of targeted tests becomes practical. Targeted testing is the cornerstone for adaptive testing. These IRT characteristics also facilitate the building of item banks with item content that extends beyond a single grade or school district, which enables the development of vertical scales such as the RIT scales that extend from kindergarten to high school.

5.2. Calculation of RIT Scores

MAP Growth employs a common item selection and test scoring algorithm. Each student begins the test with a preliminary student score based on past test performance. If a student has no prior test score, a default starting value is assigned according to test content and the student's grade. As each test proceeds, each item is selected from a large pool of Rasch-calibrated items based on the student's interim ability estimate, content requirements, and longitudinal item exposure controls. Interim ability estimates are updated after each response using Bayesian methods (Owen, 1975) that consider all of the student's responses up to that point in the test. The updated interim ability estimate is factored into selection of the next item. As this cycle is repeated, each successive interim ability estimate is slightly more precise than the previous one. The test continues until the standard error associated with the estimate is as small as it is likely to be in the test session. The final ability estimate (i.e., RIT score) is computed via a maximum-likelihood algorithm with fencing that indicates the student's location on the RIT scale.

5.3. 2015 MAP Growth Norms

Apart from interpretations of performance and growth regarding content, how students performed or grew compared to an appropriate reference peer group (provided by norms) is important information for individualizing instruction, setting achievement goals for students or entire schools, understanding achievement patterns, and evaluating student performance. The 2015 MAP Growth norms (Thum & Hauser, 2015) provide comparative information about achievement and growth for all potential MAP Growth users from carefully defined reference populations, allowing educators to compare achievement status—and changes in achievement status (growth) between test occasions—to students' performance in the same grade at a comparable instructional stage of the school year. In achievement status norms, a student's performance on the MAP Growth test, expressed as a RIT score, is associated with a percentile ranking that shows how well the student performed in a content area compared to students in the norming group. The relative evaluation of a student's growth from one period to another (e.g., from fall to spring) is provided by growth norms.

5.3.1. Norm Reference Groups

The MAP Growth norms were created using the most recent longitudinal data from the vast archive that has been assembled by NWEA over the years. The 2015 study produced norms for Grades K–11. Each set is comprised of 200,000–800,000 scores from 110,000–200,000 students attending a random sample of 1,300–1,500 NWEA partner schools that were weighted using rigorous procedures to represent the 23,500 U.S. public schools spread across 6,000 districts in 49 states.

5.3.2. Variation in Testing Schedules and Instructional Time

School calendars can vary by state and district, which means students are likely to receive different amounts of instruction at every point in a school year. In addition, MAP Growth is administered several times each year based on schedules determined by schools and districts, so testing schedules can vary considerably between and within districts. As a result, it is very likely that students who test on the same day will not have had the same amount of instructional exposure. Variation in instructional exposure means that students' opportunity to learn is likely to be unequal (Berliner, 1990), which can be detrimental to sound measurement and fair evaluation and comparison of students' test scores. Comparing two students' RIT scores would be unfair unless they started school on the same day and shared the same testing date, and comparisons of growth would not be appropriate without considering whether students have had an equal amount of instructional exposure when they tested. Both of these issues were resolved by taking instructional time into account when creating the MAP Growth norms.

To capture instructional time, school district calendars were used to establish when schools' instructional years began, when they ended, and which days were non-instructional days. Rather than an inconvenient technical hurdle for building norms, strong variation in testing schedules actually improves the description of growth over time, leading to more accurate norms for growth. Not only does a sound model of how students grow provide the basis for producing estimates of time-specific achievement status norms, it also enables the estimation of growth norms that are tailored to student peer groups and their specific testing schedules.

5.3.3. Estimating the 2015 MAP Growth Norms

Thum and Hauser (2015) employed a three-level hierarchal linear model (HLM) to reflect the nesting of repeated observations of students within schools for modeling growth. A new growth function called the compound polynomial was introduced to better fit time-series data with marked seasonality (i.e., seasonal or periodic patterns, such as the "summer drop" from spring to fall). School-level post-stratification weights were then applied at the school level to approximate the growth patterns of students in a nationally representative population of U.S. public schools. These weights were based on the national distribution of the School Challenge Index (SCI), a measure of how U.S. public schools compare in terms of the challenges and opportunities they operate under (as reflected by an array of factors they do not control, such as student ethnicity, school type, Title 1 status, and urbanicity). The higher SCI school faces a higher level of challenge. Model estimation also considered the imprecision of the outcomes to improve precision. Estimation results were then restructured to give the joint marginal distribution of predicted scores from which achievement status and growth norms were generated for both students and schools.

5.3.4. Achievement Status and Growth Norms

The joint marginal distribution of predicted scores contains all the information necessary to produce achievement status norms for a student who is tested after any specific amount of instructional exposure (as measured by instructional week on the student's school calendar). Although achievement status and growth norms are only provided by term (fall = week 4, winter = week 20, and spring = week 32) in Appendices A and B of the norms study report (Thum & Hauser, 2015), a fuller set of norms for all instructional weeks between the first and the last week (weeks 1–36) of the school year are available in the MAP Growth reporting system and included on individual reports.

The norms include the standard deviation (SD), which is a measure of dispersion of scores around the mean. The smaller the SD, the more compact the scores are around the mean. SDs are particularly useful when comparing student-level and school-level norms. For example, knowing the spread of the data can help identify students who fall well above or below the school average. When making determinations of relative effectiveness, the SDs provided with school norms can also help determine if schools have roughly the same range of scores.

5.3.5. Measuring Growth

There is a strong tendency among stakeholders to say that an assessment measures growth. However, it should be clear that assessments measure achievement, not growth. To measure growth presupposes the following:

1. The student is observed on two or more occasions.
2. Each observation accurately measures performance on a common underlying developmental construct.

Growth is measured by comparing performances between testing occasions. The starting score is treated as a factor predicting growth. If a student's starting score was below the grade level status mean, the expected growth is typically higher. Similarly, students with starting scores above the grade level mean would typically show less growth on average. Growth norms that condition on the starting performance of the student may be achieved through direct conditioning of the joint distribution of growth and initial status. This approach results in a normative measure of growth called the conditional growth index (CGI) and its corresponding population percentile called the conditional growth percentile (CGP).

The CGI operates as a standardized effect size that expresses how much an individual student grew when compared with their academic peers. It is different from the growth index because the CGI indicates how many standard deviation units above or below the growth norm a student's growth actually was, while the growth index simply indicates how many RIT points the student grew above or below the growth projections. A CGI score of zero indicates a student grew an amount typical of his peers. Positive CGIs indicate that a student's growth exceeded the growth norms, whereas negative CGIs indicate that a student's growth was less than the growth norms. The CGI allows for growth comparisons to be made between students of differing achievement levels and across different grades and content areas. The corresponding CGP is the student's percentile rank for growth. A CGP of 50 means that the student's growth (compared to their growth projection) was greater than 50% of all students in the norm reference group.

Each set of growth norms, defined by the choice of starting performance and testing schedule, represents a different growth scale. Nationally representative growth norms for each combination of pre-test performance and instructional weeks were produced for students based on the distribution of predicted growth scale values of students in the population. Similar growth norms are also available for use with schools. Student and school conditional growth distributions and percentiles are provided in Appendices D and E of the norms study (Thum & Hauser, 2015). The NWEA reporting system should be employed when exact values are required.

Apart from how it is derived, the CGP for students is functionally equivalent to the popular growth measure for state assessments known as the Colorado Growth Model proposed by Betebenner (2008). The school-level CGI and CGP should always be employed for evaluating progress of schools. Because the variance in school means is typically only about 1/5 the variance in student scores (within schools), NWEA cautions against the use of student-level norms for evaluating schools, a practice that will generally understate the performance of the more-effective schools and overstate the performance of the less-effective ones.

5.3.6. Norms Example

Table 5.1 presents an evaluation of the fall-to-spring Reading growth of a sample of fictional Grade 4 students. As shown in the table, Peter got a RIT score of 195 on the MAP Growth Reading fall assessment. Using the student achievement status norms, a teacher can see that the student scored below the average Reading RIT score for a Grade 4 student in the fall who took the assessment during the same instructional week as Peter (i.e., an average RIT score of 199 and a standard deviation of 15.4). Peter's fall percentile is 40.

Peter then got a RIT score of 207 on MAP Growth Reading in the spring, with a gain (i.e., growth index) of 12 RIT points. Using the student growth norms, the teacher can see that the mean growth from fall to spring for a Grade 4 student on the MAP Growth Reading test with the same starting RIT score as Peter is 7.1 points with an SD of 6.1. This lets the teacher know that Peter has grown more than that expected of his peers, with a CGP of 79%. As another example, Ash and Larry took their tests during the same instructional week. In the fall, Ash scored 201 RITs (57%) while Larry scored 198 RITs (50%). Thus, their expected gains in the spring were 7.5 RITs and 7.9 RITs, respectively. Ash grew 8 RITs (53% CGP) by spring and Larry 10 RITs (62% CGP).

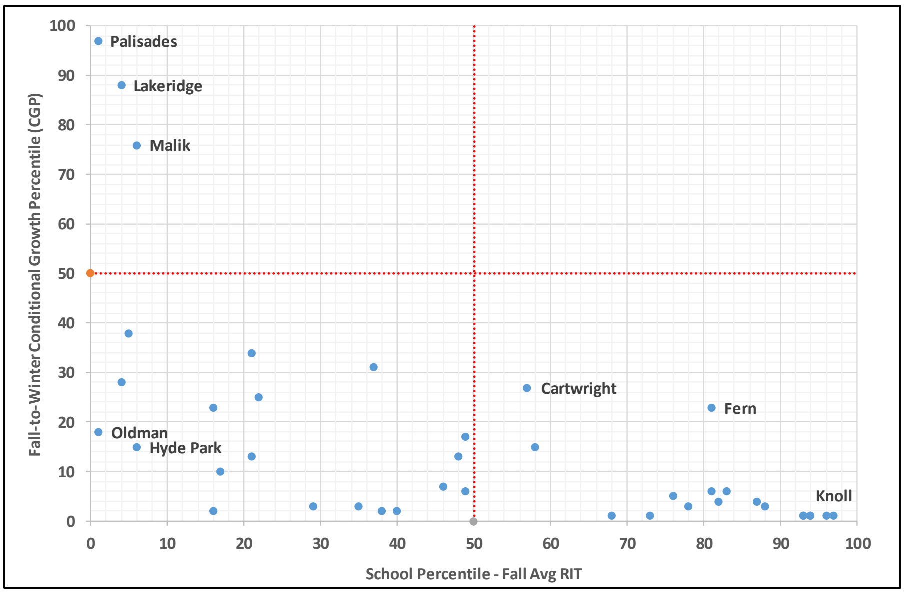
Table 5.1. Evaluation of Growth for a Sample of Grade 4 Students in MAP Growth Reading

	Fall						Spring						Fall-to-Spring Growth					
	Observed			Norms			Observed			Norms			Observed		Norms			
Student	Week	Score	SEM*	Mean	SD	%	Week	Score	SEM*	Mean	SD	%	Gain	SE	Mean	SD	CGI	CGP
Peter	6	195	3.2	199	15.4	40	30	207	3.2	206	14.9	54	12	4.5	7.1	6.1	0.79	79
Sasha	8	201	3.1	200	15.3	53	29	204	3.1	206	14.9	46	3	4.3	5.6	5.7	-0.45	32
Ash	4	201	3.3	198	15.5	57	33	209	3.1	206	14.9	58	8	4.5	7.5	6.7	0.08	53
Greg	6	196	3.2	199	15.4	42	36	204	3.3	206	15.0	44	8	4.6	7.8	7.0	0.03	51
Larry	4	198	3.1	198	15.5	50	33	208	3.2	206	14.9	55	10	4.5	7.9	6.7	0.31	62
Stan	5	196	3.3	199	15.5	43	31	203	3.2	206	14.0	43	7	4.6	7.6	6.4	-0.09	47

*SEMs lower than 3.5 indicate reliable scores on the MAP Growth scale. SEMs generally do not fall lower than 3.0 regardless of the content area.

To illustrate school growth norms, Figure 5.1 presents the growth of fictional schools in a district in terms of the average MAP Growth Reading scores of their Grade 4 students between fall and winter. The schools vary considerably in the average performance of their Grade 4 students during the fall. Growth appears to be well below expectation for most schools, except for the lower-performing schools in the fall in Palisades, Lakeridge, and Malik. The higher-performing schools in the fall, like Fern and Knoll, did not grow as strongly as expected.

Figure 5.1. Fall-to-Winter CGP for a Sample of Schools in MAP Growth Reading Grade 4



5.4. RIT Score Descriptive Statistics

Data included in the RIT score descriptive statistics analyses were from the Fall 2016, Winter 2017, Spring 2017, and Fall 2017 administrations of the MAP Growth assessments for use with the CCSS and NGSS. See Appendix A for the number of students included in the sample by state and demographics.

5.4.1. Overall Descriptive Statistics

Table 5.2 presents summary descriptive statistics of RIT scores by grade and content area, including the mean, standard deviation (SD), and the minimum and maximum RIT scores. Appendix B provides the average RIT scores by state and grade. The average RIT score at each grade varies slightly across states.

For each content area, the mean RIT score generally increases as the grade level increases. For Reading, the average RIT score increases until Grade 9 when it vacillates in subsequent grades, with the Grade 12 mean dropping as low as the Grade 7 mean. The RIT score SD steadily increases from 14 points in kindergarten to 20 points in Grade 12. Test length (i.e., the number of items) decreases from kindergarten to Grade 12, but the test duration (in minutes) is lowest in early grades and peaks in middle school. Language Usage follows a similar pattern as Reading in terms of mean RIT scores. However, the number of Language Usage items is constant across grades, and the test duration is more consistent across grades.

In Mathematics, mean RIT scores generally increase across grade levels. Exceptions include the Grade 9 mean that is lower than the Grade 8 mean and mean scores that decrease in Grades 11 and 12. RIT score SDs also increase with grade. Exceptions to this trend occur in Grades 2, 3, and 4. However, the values for these grades are still within the range of values observed across grades. The number of Mathematics items is consistent across grades, but test duration tends to decrease with grade.

Science provides an increasing trend in mean RIT scores from Grades 3–11. The SD of RIT scores also increases with values ranging from 11.8 in Grade 1 to a high of 15.5 in Grade 12. Science tests have 40–42 items, with longer tests appearing in earlier grades.

Table 5.2. Overall Descriptive Statistics of RIT Scores

Grade	#Test Events	#Items	Test Duration (minutes)	RIT Mean	RIT SD	RIT Min.	RIT Max.
Reading							
K*	865,951	49	32.0	148.2	14.3	100.1	254.5
1	1,104,917	49	34.2	167.0	16.8	100.1	251.0
2	1,351,809	42	43.5	180.3	17.8	100.1	251.9
3	1,445,055	40	53.4	191.7	17.4	106.4	253.8
4	1,440,187	40	59.1	200.7	16.9	101.9	259.9
5	1,440,237	40	62.1	207.5	16.6	102.6	259.8
6	1,374,256	39	67.9	212.3	16.3	104.3	268.1
7	1,329,350	39	66.8	216.4	16.4	108.2	268.1
8	1,288,344	39	67.3	220.2	16.3	110.6	270.3
9	543,717	39	55.9	218.9	17.9	109.3	270.3
10	424,494	39	51.5	220.4	18.1	108.4	270.1
11	194,789	39	48.6	219.2	18.9	112.1	269.5
12	76,718	40	47.2	216.2	20.2	107.1	268.8
Language Usage							
2	237,133	52	38.7	180.5	16.9	136.3	257.0
3	374,261	52	44.0	192.0	16.1	139.0	259.6
4	405,948	52	48.3	200.6	15.4	138.6	268.5
5	406,982	52	50.6	206.7	14.9	137.1	259.2
6	424,438	52	49.6	211.1	14.9	137.8	264.7
7	403,828	52	47.9	214.9	14.8	142.1	267.6
8	391,904	52	47.2	218.4	14.8	137.7	267.3
9	193,601	52	42.2	217.3	15.9	138.6	268.5
10	169,162	52	39.3	219.6	15.8	144.2	269.2
11	83,983	52	38.2	219.6	16.5	139.0	267.4
12	28,229	52	37.9	216.7	18.0	137.7	269.6

Grade	#Test Events	#Items	Test Duration (minutes)	RIT Mean	RIT SD	RIT Min.	RIT Max.
Mathematics							
K*	910,330	50	31.0	147.1	16.9	100.0	267.8
1	1,160,639	49	36.9	168.9	18.1	100.0	268.0
2	1,386,531	51	43.8	182.9	16.0	100.1	269.8
3	1,464,118	52	50.2	193.8	14.9	102.1	290.7
4	1,454,385	52	54.9	204.6	15.6	101.4	295.0
5	1,457,360	52	59.7	213.5	16.9	100.0	302.4
6	1,414,750	51	65.7	217.3	17.0	100.5	303.6
7	1,356,673	51	67.9	223.4	18.4	103.4	306.5
8	1,301,542	51	69.6	228.7	19.3	104.1	307.5
9	533,229	51	57.5	227.0	20.4	101.1	306.2
10	416,873	51	53.6	229.5	21.0	106.9	306.8
11	207,217	51	50.9	228.9	21.8	104.3	307.4
12	75,024	51	48.0	224.9	22.9	100.2	305.5
Science							
2	1,468	42	34.4	182.2	12.5	221.2	150.5
3	86,819	42	39.7	189.5	12.2	146.8	232.5
4	110,488	42	43.6	196.7	11.8	149.0	241.2
5	139,411	41	45.7	201.4	12.4	145.7	249.8
6	154,819	41	44.0	205.5	12.2	148.0	265.2
7	158,035	41	44.5	209.1	12.8	148.6	260.0
8	162,983	40	43.3	211.5	13.4	149.5	268.0
9	35,344	40	37.8	214.6	13.7	154.2	264.3
10	27,944	40	35.0	216.3	14.6	157.2	264.3
11	13,540	40	33.1	216.8	14.7	159.9	264.8
12	3,543	40	31.2	213.7	15.5	153.6	260.9

*Grade K includes kindergarten and below.

5.4.2. Descriptive Statistics by Instructional Area

Table 5.3 – Table 5.8 present the RIT score mean and SD by instructional area. Descriptive statistics for MAP Growth Reading and Mathematics K–2 are provided separately from the 2–5 and 6+ results because the instructional areas for those grade bands differ. Language Usage is designed for Grades 2–12 with three instructional areas across all grades, and Science is designed for Grades 3–5 and 6+ with three instructional areas across both levels. Summaries of the tables are as follows. Overall, the results confirm the vertical scale design and increasing difficulty of content across grades with a few exceptions in the upper grades.

RIT scores for the Reading K–2 instructional areas increase on average across grades and within each grade, as the instructional areas have similar mean RIT scores. The average RIT score for each Reading 2–12 instructional area also generally increases across grades. The pattern is most evident in lower grades and becomes irregular in high school. Each Reading instructional area is of comparable difficulty. The average scores within a grade are similar across instructional areas. In Language Usage, mean RIT scores increase across grades until high school and then level out. Mean scores for Grade 12 students tend to be the lowest in high school. There is no clear difference in the difficulty across instructional areas. Mean scores within a grade tend to be similar across instructional areas.

Mathematics K–2 average scores increase across grades for each instructional area. Operations and Algebraic Thinking is consistently the easiest instructional area, as evidenced by the consistently, albeit only slightly, higher mean scores. The SDs range from 18 to 22 points. Geometry shows the most variability in RIT scores. In Grades 2–12, average Mathematics RIT scores demonstrate a familiar trend. Means generally increase across grades. The clearest trend is for Algebraic Thinking and Geometry. Interestingly, the mean scores for Number and Operations and Measurement and Data appear to increase until about middle school and then decrease in high school. The decrease in high school may be attributed to more selective groups of students taking the test.

Mean RIT scores for each Science instructional area show an increasing trend with grade until Grade 11 or 12. The increases are most evident at the lower grades. The smallest gains occur in high school.

Table 5.3. RIT Score Descriptive Statistics by Instructional Area—Reading K–2

Grade	#Test Events	Foundational Skills		Language & Writing		Literature & Informational		Vocabulary Use & Functions	
		Mean	SD	Mean	SD	Mean	SD	Mean	SD
K*	865,760	146.4	17.4	146.7	14.7	149.8	15.0	149.9	15.5
1	1,101,775	167.0	19.3	165.9	17.2	167.6	17.6	167.3	17.6
2	350,597	179.4	19.4	179.4	17.4	180.7	17.9	180.5	17.8

*Grade K includes kindergarten and below.

Table 5.4. RIT Score Descriptive Statistics by Instructional Area—Reading 2–12

Grade	#Test Events	Literary Text		Informational Text		Vocabulary	
		Mean	SD	Mean	SD	Mean	SD
2	1,001,204	181.7	18.7	179.9	19.4	179.8	18.8
3	1,437,551	192.4	18.3	191.6	18.3	191.3	17.9
4	1,435,809	201.2	17.9	200.7	17.6	200.5	17.3
5	1,437,257	207.9	17.7	207.4	17.2	207.5	17.0
6	1,372,960	212.3	17.4	212.1	17.1	212.6	16.9
7	1,328,700	216.3	17.5	216.1	17.2	216.9	16.9
8	1,287,725	220.0	17.4	220.0	17.2	220.9	16.8
9	543,439	218.4	19.0	218.4	18.7	220.2	18.4
10	424,255	219.7	19.3	219.8	18.8	222.1	18.6
11	194,609	218.3	19.9	218.5	19.5	221.3	19.4
12	76,562	215.2	21.1	215.4	20.6	218.7	20.8

Table 5.5. RIT Score Descriptive Statistics by Instructional Area—Language Usage 2–12

Grade	#Test Events	Writing		Language: Understand, Edit for Grammar, Usage		Language: Understand, Edit for Mechanics	
		Mean	SD	Mean	SD	Mean	SD
2	237,133	180.5	16.3	181.1	18.7	180.2	17.9
3	374,261	191.4	16.3	192.7	17.2	192.1	17.1
4	405,948	199.8	16.1	201.0	16.1	200.9	16.2
5	406,982	206.2	16.0	206.7	15.4	207.1	15.6
6	424,438	210.9	16.2	210.9	15.2	211.7	15.5
7	403,828	214.8	16.3	214.3	15.1	215.5	15.3
8	391,904	218.5	16.4	217.6	15.1	219.0	15.3
9	193,601	217.3	17.7	216.5	16.0	218.2	16.2
10	169,162	219.4	17.7	218.8	15.9	220.7	16.2
11	83,983	219.2	18.4	218.8	16.8	220.9	16.9
12	28,229	216.1	19.8	215.8	18.2	218.3	18.2

Table 5.6. RIT Score Descriptive Statistics by Instructional Area—Mathematics K–2

Grade	#Test Events	Operations & Algebraic Thinking		Number & Operations		Measurement & Data		Geometry	
		Mean	SD	Mean	SD	Mean	SD	Mean	SD
K*	910,136	146.0	19.3	146.1	18.1	147.4	17.1	148.5	18.4
1	1,156,961	170.7	18.7	168.6	19.5	167.6	18.4	168.6	20.9
2	369,099	185.4	18.2	186.3	19.6	183.8	19.7	184.9	22.2

*Grade K includes kindergarten and below.

Table 5.7. RIT Score Descriptive Statistics by Instructional Area—Mathematics 2–12

Grade	#Test Events	Algebraic Thinking		Number & Operations		Measurement & Data		Geometry		The Real & Complex Number Systems		Statistics & Probability	
		Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
2	1,017,417	181.3	16.2	181.5	15.6	181.7	16.0	183.6	17.0	186.9	21.7	186.4	21.4
3	1,457,285	194.0	16.6	193.1	15.0	193.9	16.2	194.5	15.9	196.4	19.9	196.5	19.8
4	1,450,373	205.0	16.6	204.5	16.1	204.4	17.0	204.9	16.6	220.4	23.3	218.1	23.3
5	1,454,634	212.9	17.1	214.8	18.3	212.7	18.6	213.5	17.6	227.9	19.9	224.7	20.9
6	1,413,485	216.9	17.3	208.1	27.2	205.1	25.8	217.2	17.9	219.8	18.1	215.8	18.5
7	1,356,078	223.4	18.8	201.0	27.1	199.0	25.7	222.7	19.1	225.1	19.3	222.9	19.9
8	1,300,948	229.6	20.2	204.3	27.9	202.3	27.3	227.9	20.0	229.2	20.0	228.5	20.7
9	532,966	228.9	21.5	201.9	25.7	200.5	24.7	226.1	21.1	227.0	20.7	226.5	21.5
10	416,659	231.5	22.1	195.9	20.5	194.4	20.2	229.2	21.8	229.1	21.7	228.8	21.9
11	207,038	231.0	23.1	197.2	22.0	197.2	21.1	228.4	22.2	228.8	22.6	227.8	22.4
12	74,870	227.1	24.3	196.7	22.0	196.0	21.4	224.2	23.0	225.8	23.5	224.0	23.2

Table 5.8. RIT Score Descriptive Statistics by Instructional Area—Science 2–12

Grade	#Test Events	Life Science		Physical Science		Earth & Space Science	
		Mean	SD	Mean	SD	Mean	SD
2	1,468	182.2	13.9	181.8	13.3	182.9	13.2
3	86,819	189.3	13.6	189.5	13.1	189.9	12.8
4	110,488	196.5	13.4	196.9	12.6	196.8	12.4
5	139,411	201.4	14.0	201.7	13.2	201.2	12.9
6	154,819	205.4	13.3	205.6	13.0	205.6	13.1
7	158,035	209.0	13.8	209.2	13.8	209.3	13.7
8	162,983	211.7	14.6	211.6	14.3	211.3	14.1
9	35,344	214.6	14.9	214.8	14.6	214.5	14.4
10	27,944	216.9	16.3	216.4	15.4	215.7	14.8
11	13,540	217.6	16.3	217.2	16.0	215.6	14.4
12	3,543	214.2	16.8	214.2	16.8	213.0	15.3

5.5. Item Calibration

Items must be properly calibrated to the RIT scale before being added to the MAP Growth item pools. Field test items are administered in fixed positions on MAP Growth tests. Responses are continuously collected on a field test item until it successfully passes calibration. The calibration process involves three steps: filtering, calibration, and evaluation. Filtering eliminates invalid test events such as those outside valid grade ranges or students flagged as disengaged test takers. Calibration requires a minimum sample size of 1,000 responses. Items failing to meet this criterion are returned to field testing.

The calibration process follows the concept of common person equating, first presented by Masters (1985). To initiate the process, student achievement is first estimated from responses to the calibrated items in an operational test containing field test items. This estimate is used to anchor field test items to the original measurement scale. Using the fixed student achievement estimates as an anchor point, unconditional maximum likelihood is used to obtain a first estimate of the field test item's difficulty. Item calibrations are estimated from the student responses in a common grade level. Sets of responses are examined in descending order from the highest grade to the lowest grade. The first calibration estimate that is based on more than 1,000 responses and meets the calibration criteria is adopted as the item's calibration.

To improve this initial estimate, responses given by students with a probability of answering the item correctly that is at or below 10% are treated as missing during a second calibration step. This procedure is consistent with the theorem presented by Andersen (2002) and demonstrated by Andrich, Marais, and Humphry (2012) to improve item fit and reduce estimation bias. With the low probability responses removed, a second calibration is estimated using the same person anchor from the first step. These procedures are contained within a proprietary item calibration program designed for this purpose. Calibrating items in this way allows for continuous expansion of the item pool.

Calibration is automatically evaluated for certain conditions using several rules and statistics. Items remain in field testing if any of the following are observed:

- | provisional calibration – estimated calibration | ≥ 20
- Number of responses < 1,000
- Correct responses < 15%
- Correct responses > 90%
- Point-measure correlation < .20

Items are removed from the pool or are revised and re-field tested if any of the following occur:

- Any answer option receives < 5% of the responses
- Any distractor receives a positive point-measure correlation
- Any answer option receives a greater percentage of responses than the keyed option
- The keyed response has a negative point-measure correlation

Once field test items pass these checks, they are evaluated for model fit using automated processes and human review.

5.6. Field Test Item Evaluation

Good item parameter estimates are critical to the validity of a test based on IRT. The evaluation of calibrated field test items ensures that the operational items work well with students. It also allows an opportunity for items to be reworded and field tested again to improve both the content and measurement quality of the item prior to being used operationally.

To evaluate a field test item's calibration, NWEA employs various descriptive statistics (e.g., percent correct, point-measurement correlation) and calculates item infit and outfit statistics that provide useful information about how well the responses adhere to the expectation of the Rasch model. However, various forms of information collected about an item's calibration status do not necessarily result in a decision about item quality. For example, some indicators can suggest good quality while others suggest caution. In such cases, human reviewers drive the final decision. However, human reviews are expensive and inefficient, especially when large numbers of items are under consideration. Recognizing this, NWEA adopts an integrated procedure called Model of Man (MoM) by employing automated procedures and human judgment. The automated procedure uses item fit statistics to mimic human review behavior and improve the overall quality and efficiency of the calibration process.

5.6.1. Item Fit

Item fit is evaluated with multiple indices and criteria, as shown in Table 5.9. Most of the indices provide information about the fit of the Rasch model to the observed responses. Two indices, percent correct and discrimination, are classical statistics that describe item data. Percent correct criteria at this phase of evaluation are stricter than those applied during calibration to identify items in need of additional field testing.

Table 5.9. Fit Index Descriptions and Criteria

Fit Index	Description	Criterion
Infit	Rasch weighted mean square fit statistic	< 1.09
Outfit	Rasch unweighted mean square fit statistic	< 1.09
MSF	Mean square fit	< 0.9
RMSE	Root mean squared error	< 1.0
Chi-square	Tests observed count correct versus expected count correct.	N/A
Std. Chi-square	Standardized chi-square statistic (Wilson & Hilferty, 1931)	< 1.0
r	Relationship between observed and expected values	> 0.75
Percent correct	Proportion of correct responses	$0.3 < p < 0.8$
Discrimination	Correlation between RIT score and item response	> 0.25

Graphic displays of item response functions are used to further evaluate items with borderline fit statistics. The item response function is a plot that shows the probability of a correct response to an item against the achievement levels of the students who responded to the item. When reviewing an item response display, the empirical item response function is plotted on the same grid as the theoretical function. When large discrepancies exist between the two curves, there is a lack of fit between the item and the scale. A more comprehensive understanding of item performance can be gained by reviewing the response functions. For example, if an item has a borderline chi-square value (indicating that performance on the item does not track well with increases in achievement), the item is flagged for revision or deletion.

Figure 5.2 and Figure 5.3 show the theoretical and empirical response functions for two items that were both field tested by more than 4,000 students. In these graphs, the smooth curve shows the theoretical item response function from Equation 5.1, calibrated to the measurement scale based on all students responding. The vertical lines extending from the theoretical curve show the empirical proportion correct for the group of students with any final RIT score. Points not connected to the theoretical curve via a vertical line are based on small numbers of students (fewer than 10). The extent to which the empirical results deviate from the theoretical curve provides an index of item misfit. If the misfit is great, it might indicate that the item is flawed or that the model does not completely describe the item’s performance.

Specifically, Figure 5.2 shows the results for a difficult Mathematics item with poor model fit. Upon review, the item was identified as being vaguely worded and was rejected for use in the item banks. Figure 5.3 shows the results from a Reading item with good fit to the Rasch model. The empirical results match the theoretical curve quite well, except in the extremes of the measurement range. However, in both the MAP Growth and the MAP Growth K–2 systems, items are targeted to the student’s performance, so it is rare that a student would see an item in the extremes of its measurement range. This item was approved for use in the item banks.

Figure 5.2. Mathematics Item with Poor Model Fit

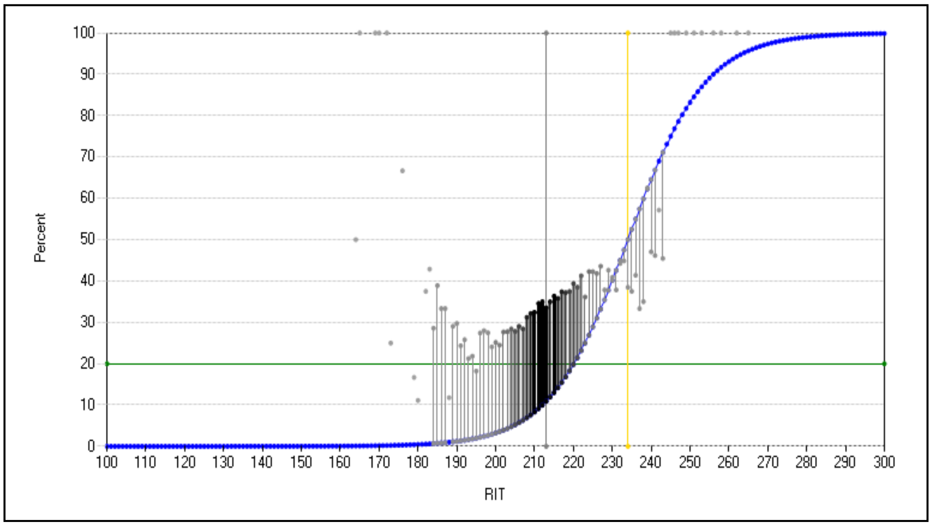
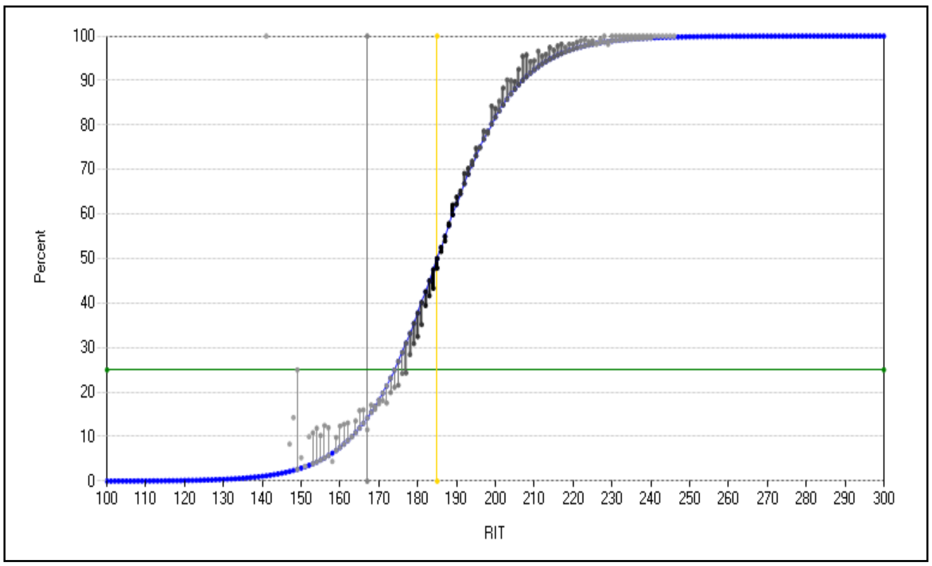


Figure 5.3. Reading Item with Good Model Fit



5.6.2. Model of Man (MoM) Procedure

The MoM procedure was developed using a set of item calibration records containing 8,017 items across the four content areas (Reading, Language Usage, Mathematics, and Science) that were reviewed by two psychometricians over a 14-month period. The items were split into training and evaluation groups. Hauser, Thum, He, and Ma (2014) provided a detailed description of the MoM development process. They used the training group to build predictive models with a logistic regression approach with stepwise selection for each outcome, each for a content area, to identify the probability associated with decisions. The independent variables were the statistical indices calculated during the item calibration process. Experts' item review decisions were used as a dependent variable. Statistically insignificant variables were dropped from the model. After the field test items calibrate through the item calibration engine, MoM is applied to the successfully calibrated items. The logistic regression model in MoM calculates the probabilities for each item that puts them into different status categories: "Auto Accept," "Keep Field Test," "Borderline Accept," "Auto Reject," and "Borderline Reject."

5.6.3. Human Review Process

The human review process is conducted by psychometricians and content specialists. Once MoM provides the status categories to the successfully calibrated field test items, a visual review process is conducted by psychometricians who review the items by comparing the empirical item response function to the model-expected IRT. An item is flagged as "Auto Accepted" if its empirical and model item response functions are close across the RIT scale. If not, a psychometrician evaluates if the range of the differences is small. If the range is small and the total response count is larger than 5,000, the item is flagged as "Auto Accepted." The item is flagged as "Keep Field Test" if the range is small and the total response count is less than 5,000. The "Auto Reject" flag is given to an item if the range of the differences is large. This visual process typically has three rounds of review involving at least two psychometricians:

1. In the first review, a psychometrician reviews all the "Borderline Reject," "Borderline Accept," "Auto Reject," and "Auto Accept" items with item-total correlations above 0.10. The first reviewer also reviews most of the "Keep Field Test" items.
2. The second reviewer examines all the "Borderline Reject" and "Auto Reject" items accepted by the first reviewer and all the "Borderline Accept" and "Auto Accept" items rejected by the first reviewer.
3. The third review is only focused on the items that received different review decisions in the first two reviews.

Once psychometricians complete the visual review, the items flagged as "Auto Rejected" move to a post-calibration content review by content specialists who decide if the items could be revised or should be kept out of the MAP Growth item bank.

5.7. Item Parameter Drift

Periodic reviews of item performance are conducted by psychometricians and content specialists to ensure scale stability across time and student subgroups. The use of IRT in scale construction requires an assumption of item parameter invariance. Item parameter drift is one condition where invariance fails to hold. It occurs when an item's parameters change over time, which can result in systematic errors in scale linking, and, ultimately, test scoring (Kolen & Brennan, 2004). NWEA periodically evaluates the presence of item parameter drift using the Robust Z method (Huynh & Rawls, 2009) calculated as:

$$z* = \frac{D - Median}{0.74 \times IQR} \tag{5.3}$$

where *D* is the difference between the original difficulty parameter and the newly calibrated difficulty parameter (on the logit scale), and *IQR* is the interquartile range for the differences.

Item RIT is transformed back to the logit scale to obtain the *b*-parameter for each item. The significance level in each direction is set at 5%, and the critical value is $z^* = \pm 1.645$, correspondingly. All items with a Robust Z smaller than the absolute value of z^* are regarded as stable, otherwise items are flagged as drifting. This approach should identify approximately 10% of items as drifting if the null hypothesis is true, which allows the identification of many items for review. This ensures that items with noticeable drift can be examined by content experts. The impact of item parameter drift on test scores is also examined. Thus far, results have shown that a large majority of MAP Growth items are stable over time and have little to no drift. Moreover, the small amount of drift has minimal impact on student test scores and scale stability.

Chapter 6: Reporting

A student's overall RIT score and instructional area scores are displayed immediately once the test has been concluded. Class- and district-level reporting are available once the testing window is closed. MAP Growth reports are accessible online and are available in a variety of formats, including PDF, HTML, and CSV. The comprehensive data file is a CSV file that can be converted into a variety of formats. HTML-based reports are available in real-time immediately after a report is requested. The time it takes to generate PDF reports depends on the report's priority, size, and volume (i.e., number of test records included in the report). The MAP Growth system performs updates to the reporting database nightly.

6.1. MAP Growth Reports

Table 6.1 presents the required roles necessary to access the different report levels, and Table 6.2 summarizes the MAP Growth reports. In addition to these reports, the district assessment coordinator can use the Data Export Scheduler to export test results as CSV files to facilitate custom analysis and reporting.

Table 6.1. Required Roles for Report Access

Report Source	Required Role
Student-Level Reports	Instructor, Administrator, or District Assessment Coordinator
Class-Level Reports	Instructor, Administrator, or District Assessment Coordinator
District-Level Reports	Administrator or District Assessment Coordinator
Skills Checklist/Screening Reports	Instructor, Administrator, or District Assessment Coordinator
Learning Continuum	Instructor, Administrator, or District Assessment Coordinator

Table 6.2. Report Summary

Report Name	Description	Prior Data	Intended Audience
Student-Level Reports			
Student Profile	Brings together the data needed to advise each student and support their growth, including learning paths and growth goals.	All years prior	• Teacher • Instructional coach • Counselor • Student • Parent
Student Progress	Shows a student's overall progress from all past terms to the selected term to show the student's term-to-term growth.	All years prior	• Teacher • Instructional coach • Counselor • Student • Parent
Student Goal Setting Worksheet	Shows a student's test history and growth projections in the selected content areas for a specific period of time to discuss the student's goals and celebrate achievements.	Up to 2 years prior	• Teacher • Instructional coach • Counselor • Student • Parent
Class-Level Reports			
Class	Shows class performance for a term, including norms status rankings, to analyze student needs.	1 year prior	• Instructional coach • Teacher

Report Name	Description	Prior Data	Intended Audience
Achievement Status and Growth (ASG)	Shows three pictures of growth, all based on national norms: *projections* to set student growth goals, *summary* comparison of two terms to evaluate efforts, and an interactive *quadrant chart* to visualize growth comparisons.	Up to 2 years prior	• Instructional coach • Teacher • Counselor
Class Breakdown by RIT	Shows the academic diversity of a class across basic content areas to modify and focus the instruction for each student.	1 year prior	• Instructional coach • Teacher • Counselor
Class Breakdown by Goal	Shows the academic diversity for specific goals within a chosen content area to modify and focus the instruction for each student.	1 year prior	• Instructional coach • Teacher • Counselor
Class Breakdown by Projected Proficiency	Shows students' projected performance on state and college readiness assessments to adjust instruction for better student proficiency.	1 year prior	• Instructional coach • Teacher • Counselor • Principal
District-Level Reports			
District Summary	Summarizes RIT score test results for the current and all historical terms to inform district-level decisions and presentations.	All years prior	• Superintendent • Curriculum specialist • Instructional coach • Principal
Student Growth Summary	Shows aggregate growth in a district or school compared to the norms for similar schools to adjust instruction and use of materials.	All years prior	• Superintendent • Curriculum specialist • Instructional coach • Principal
Projected Proficiency Summary	Shows aggregated projected proficiency data to determine how a group of students is projected to perform on separate state and college readiness tests.	1 year prior	• Superintendent • Curriculum specialist • Instructional coach • Principal
Grade	Shows students' detailed and summary test data by grade for a selected term to set goals and adjust instruction.	1 year prior	• Principal • Counselor • Instructional coach
Grade Breakdown	Provides a single spreadsheet of student achievement (both subject and goal area) to flexibly group students from across the school. Unlike the Class Breakdown reports, this report has no limit on the number of students. File format is CSV.	1 year prior	• Principal • Counselor • Instructional coach
Skills Checklist / Screening Reports			
Class	Shows overall class performance for skills and concepts included in certain Screening or Skills Checklist tests to modify and focus instruction for the whole class.	Up to 3 terms prior	• Instructional coach • Teacher • Counselor
Sub-Skill	Shows test results of individual students in a selected class to identify students who need help with specific skills.	Up to 3 terms prior	• Instructional coach • Teacher • Counselor
Student	Shows individual student results from certain Screening or Skills Checklist tests to focus instruction for each student.	Up to 3 terms prior	• Teacher • Instructional coach • Counselor • Student • Parent

Report Name	Description	Prior Data	Intended Audience
Learning Continuum			
Class View	Shows students together with the skills and concepts they need to develop.	1 year prior	• Instructional coach • Teacher • Counselor
Test View	Shows skills and concepts for all RIT bands.	1 year prior	• Instructional coach • Teacher • Counselor

6.1.1. Student-Level Reports

Student reports allow educators, parents, and students to track student data throughout the school year and across years. For example, the Student Profile dashboard report shows current and past overall RIT scores, scores for instructional areas, growth information, longitudinal data, and percentile comparisons. There are three student-level reports: Student Profile, Student Progress, and Student Goal Setting Worksheet.

- With the Student Profile Report shown in Figure 6.1, educators can share how a student is performing, develop an instructional plan, and collaboratively set goals. The “Print and Share” function allows teachers to batch print the Student Profile Report for an entire class or download a PDF for an individual student, making sharing with parents easier. From within the Student Profile, educators can access current, past, and predictive data to gain a complete picture of each student’s individual growth.

- The Student Progress Report, Figure 6.2, tracks and compares student performance with the NWEA norms and/or the district over time. Instructional area performance can be displayed as quintiles or RIT values. An optional explanatory page can be printed along with the Student Progress Report for distribution to parents and teachers.

- The Student Goal Setting Worksheet, Figure 6.3, shows measured growth and projections to support conversations regarding a student's goals and achievements. The report tracks overall RIT, instructional area RIT, and Lexile range for up to five terms. It also includes growth projections for each content area.

Figure 6.1. Student Profile Report

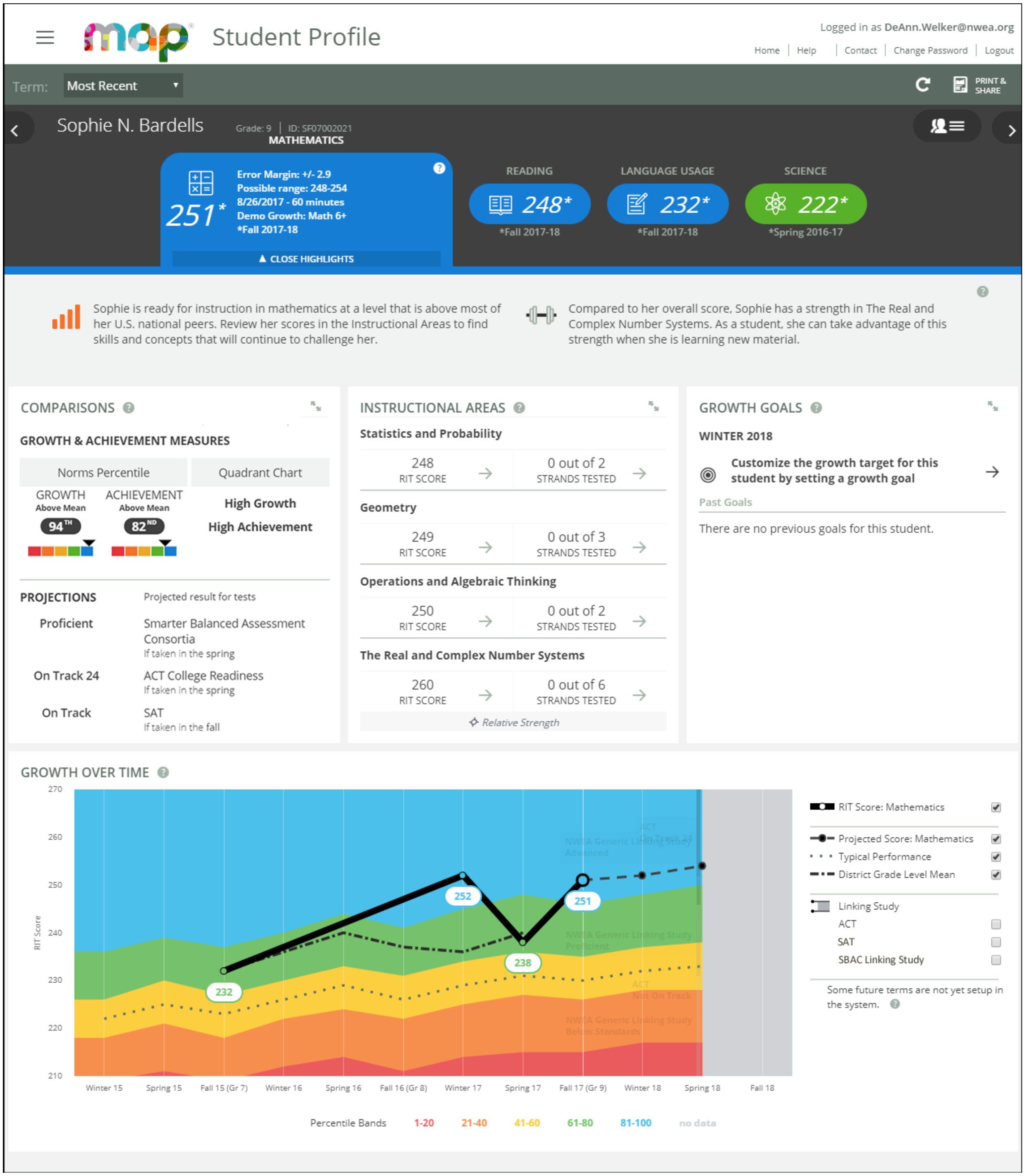
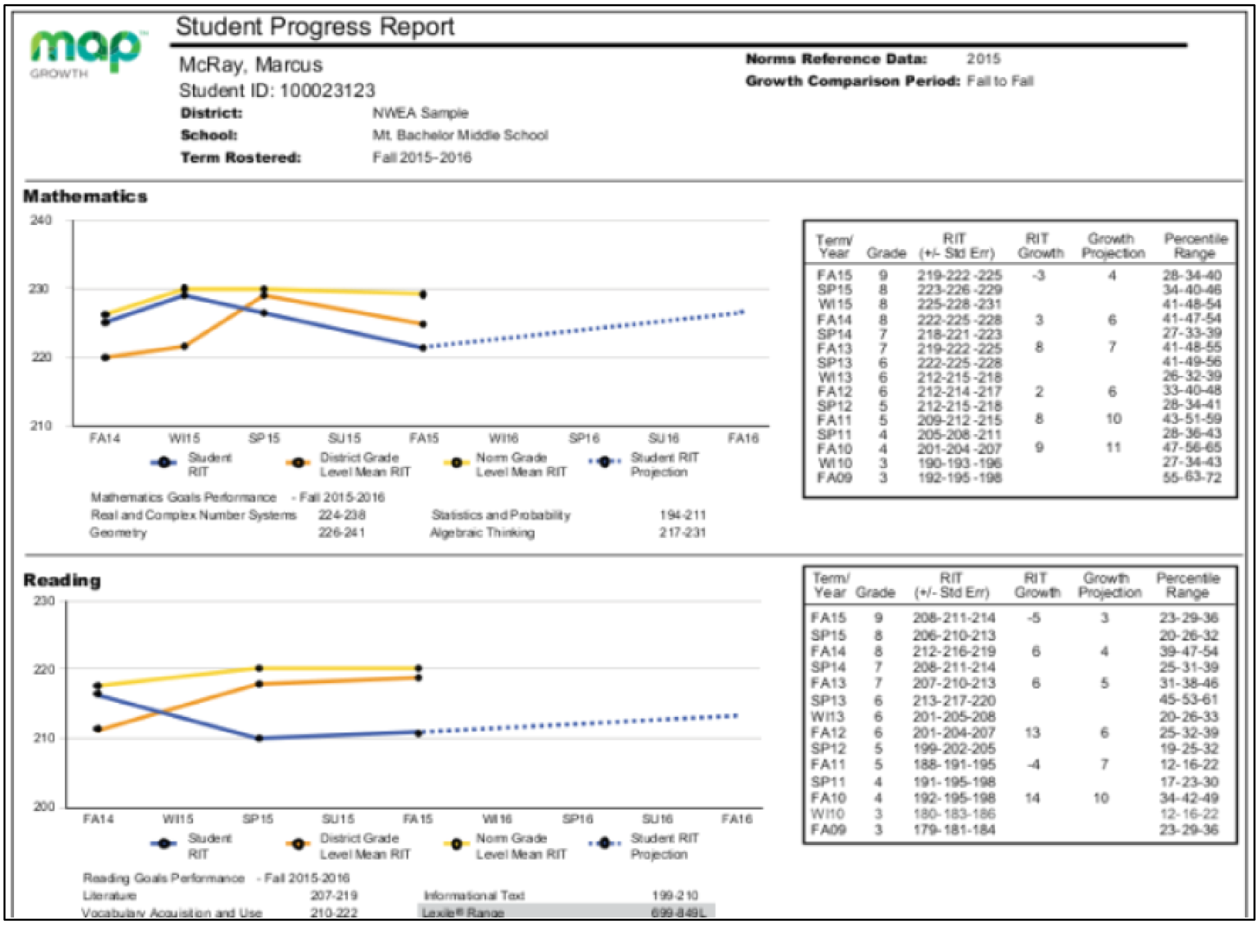


Figure 6.2. Student Progress Report



Student Progress Report

McRay, Marcus
Student ID: 100023123
District: NWEA Sample
School: Mt. Bachelor Middle School
Term Rostered: Fall 2015-2016

Norms Reference Data: 2015
Growth Comparison Period: Fall to Fall

Mathematics

Term/ Year	Grade	RIT (+/- Std Err)	RIT Growth	Growth Projection	Percentile Range
FA15	9	219-222 -225	-3	4	28-34-40
SP15	8	223-226 -229			34-40-46
WI15	8	225-228 -231			41-48-54
FA14	8	222-225 -228	3	6	41-47-54
SP14	7	218-221 -223			27-33-39
FA13	7	219-222 -225	8	7	41-48-55
SP13	6	222-225 -228			41-49-56
WI13	6	212-215 -218			26-32-39
FA12	6	212-214 -217	2	6	33-40-48
SP12	5	212-215 -218			28-34-41
FA11	5	209-212 -215	8	10	43-51-59
SP11	4	205-208 -211			28-36-43
FA10	4	201-204 -207	9	11	47-56-65
WI10	3	190-193 -196			27-34-43
FA09	3	192-195 -198			55-63-72

Mathematics Goals Performance - Fall 2015-2016

Real and Complex Number Systems	224-238	Statistics and Probability	194-211
Geometry	226-241	Algebraic Thinking	217-231

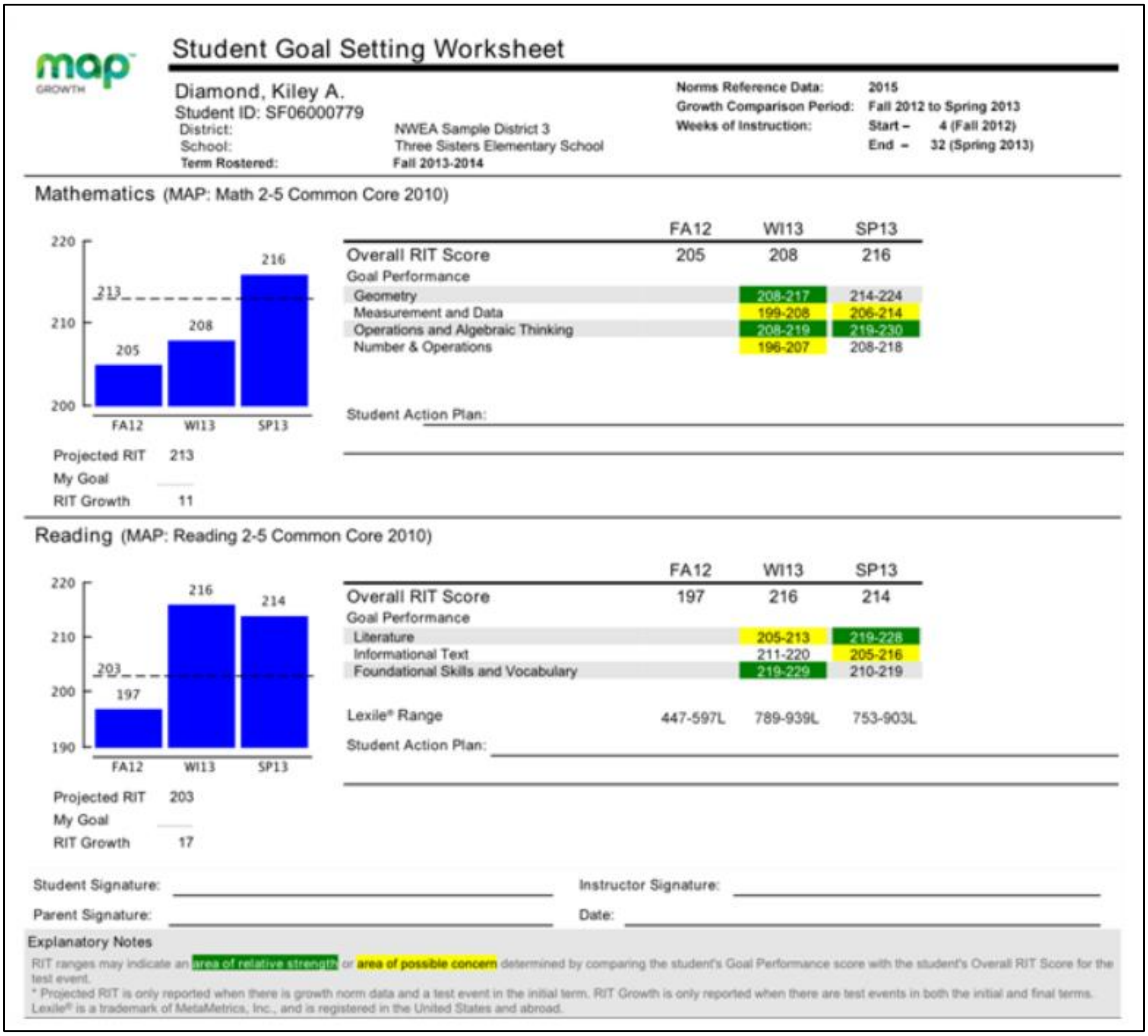
Reading

Term/ Year	Grade	RIT (+/- Std Err)	RIT Growth	Growth Projection	Percentile Range
FA15	9	208-211-214	-5	3	23-29-36
SP15	8	206-210-213			20-26-32
FA14	8	212-216-219	6	4	39-47-54
SP14	7	208-211-214			25-31-39
FA13	7	207-210-213	6	5	31-38-46
SP13	6	213-217-220			45-53-61
WI13	6	201-205-208			20-26-33
FA12	6	201-204-207	13	6	25-32-39
SP12	5	199-202-205			19-25-32
FA11	5	188-191-195	-4	7	12-16-22
SP11	4	191-195-198			17-23-30
FA10	4	192-195-198	14	10	34-42-49
WI10	3	180-183-186			12-16-22
FA09	3	179-181-184			23-29-36

Reading Goals Performance - Fall 2015-2016

Literature	207-219	Informational Text	199-210
Vocabulary Acquisition and Use	210-222	Lexile® Range	699-849L

Figure 6.3. Student Goal Setting Worksheet



6.1.2. Class-Level Reports

Class-level reports provide an overview of performance and detailed information about each student in a class. Teachers can use these reports to differentiate instruction for one student or groups of students to inform classroom practice and identify instructional areas of strength and weakness for the whole class. At the start of each term, teachers can pull previous years' assessment data for their current class. There are three class-level reports: Class, ASG, and Class Breakdown by RIT, Goal, and Projected Proficiency.

Figure 6.4 provides a sample Class Report for a middle school Mathematics class. The ASG report in Figure 6.5 is useful in measuring program effectiveness and student learning. This customizable report provides both a static and interactive summary of data. The static report shows growth projections for each student based on the NWEA norms and compares actual student growth to projected growth. With the interactive visualization of this report, teachers can see how each student is growing and achieving. The default setting for this report is to characterize achievement and growth relative to the 50th percentile, as shown in Figure 6.5.

Using this report, educators can adjust the benchmarks against which achievement and growth are compared to groups of students for more effective instruction or intervention.

The Class Breakdown reports help to focus the instruction for each student. The Class Breakdown by Projected Proficiency report, Figure 6.6, categorizes students' projected performance on state and college readiness assessments. The Class Breakdown can also be generated by RIT for a high-level view across basic content areas or by instructional area for a detailed view of instructional areas within each content area.

Figure 6.4. Class Report

map GROWTH

Class Report

Filek, Jace E
Class: Homeroom 1(A)

Term Rostered: Fall 2017-2018
Term Tested: Fall 2017-2018
District: NWEA Sample District - Partner Accoun
School: St. Helens Elementary School

Norms Reference Data: 2015
Weeks of Instruction: 5 (Fall 2017)
Small Group Display: No

Mathematics

MAP Growth: Math 2-5 / Common Core

Goal Performance
A. Operations and Algebraic Thinking
B. Number and Operations
C. Measurement and Data
D. Geometry

Name (Student ID)	Grade	Test Date	RIT (+/- Std Err)	Percentile (+/- Std Err)	Test Duration	A	B	C	D
Foglio, Nathan N. (SF06000214)	5	08/19/17	190-**193**-196	7-**10**-14	60 m	LoAvg	Low	Low	Low
Norton, Terrilynn N. (SF06000204)	5	08/22/17	190-**193**-197	7-**10**-15	60 m	Low	Low	Low	Low
Shefick, Lyndsey Marie E. (F08000118)	5	08/23/17	189-**193**-197	6-**10**-15	60 m	LoAvg	Low	Low	Low
Minez, Colton R. (F08000054)	5	08/26/17	191-**194**-197	8-**11**-16	60 m	LoAvg	Low	Low	Low
Abel, Rheanna A. (F08000089)	5	08/19/17	198-**202**-206	18-**25**-34	60 m	Avg	Low	Low	Avg
Doris, Seth E. (F08000057)	5	08/29/17	201-**204**-207	23-**30**-37	60 m	Low	LoAvg	Avg	LoAvg
Shindler, Kamery R. (SF06000027)	5	08/23/17	200-**204**-209	21-**30**-41	60 m	LoAvg	Avg	LoAvg	Low
Tresler, Shannon A. (F08000316)	5	08/18/17	201-**204**-207	22-**30**-38	60 m	LoAvg	Avg	Low	LoAvg
Arvidson, Noor R. (SF06000158)	5	08/29/17	202-**205**-208	25-**32**-41	60 m	Avg	Avg	LoAvg	Low
Mentink, Marilynn N. (SF06000220)	5	08/27/17	202-**205**-209	24-**32**-41	60 m	LoAvg	Low	LoAvg	Low
Song, MyLee G. (F08000099)	5	08/25/17	202-**206**-210	25-**35**-46	60 m	Low	LoAvg	LoAvg	HiAvg
Kane, Frisco O. (F08000163)	5	08/21/17	205-**208**-212	31-**40**-49	60 m	Avg	Avg	LoAvg	LoAvg
Schaffer, Kyle H. (SF06000145)	5	08/26/17	206-**209**-212	35-**42**-50	60 m	Avg	Avg	LoAvg	Avg
Walterscheid, Alec C. (SF06000149)	5	08/22/17	206-**209**-213	34-**42**-52	60 m	LoAvg	LoAvg	HiAvg	LoAvg
Griffith, Skyler R. (F08000246)	5	08/29/17	209-**211**-214	41-**48**-55	60 m	HiAvg	HiAvg	LoAvg	HiAvg
Landon, Donovan N. (SF06000131)	5	08/18/17	207-**211**-215	38-**48**-58	60 m	Avg	LoAvg	LoAvg	LoAvg
Stenbach, Stuart T. (F08000148)	5	08/26/17	208-**211**-215	39-**48**-57	60 m	HiAvg	Avg	Avg	HiAvg
Brown, Vince E. (SF06000122)	5	08/22/17	210-**214**-218	46-**56**-65	60 m	HiAvg	LoAvg	HiAvg	Avg
Givens, Ubadiah H. (F08000090)	5	08/18/17	210-**214**-218	45-**56**-66	60 m	HiAvg	HiAvg	HiAvg	HiAvg
Slider, Gino O. (SF06000219)	5	08/28/17	211-**215**-219	47-**58**-69	60 m	HiAvg	Avg	Avg	HiAvg
Wynnette, Gwendelin N. (SF06000086)	5	08/18/17	212-**215**-218	51-**59**-66	60 m	Avg	HiAvg	HiAvg	Avg
Bergez, Juanita I. (F08000070)	5	08/28/17	219-**222**-225	68-**75**-82	60 m	HiAvg	HiAvg	High	Avg
Cornelius, Henry N. (F08000028)	5	08/26/17	221-**223**-226	72-**78**-83	60 m	Avg	HiAvg	HiAvg	High

Figure 6.5. Achievement Status and Growth (ASG) Report

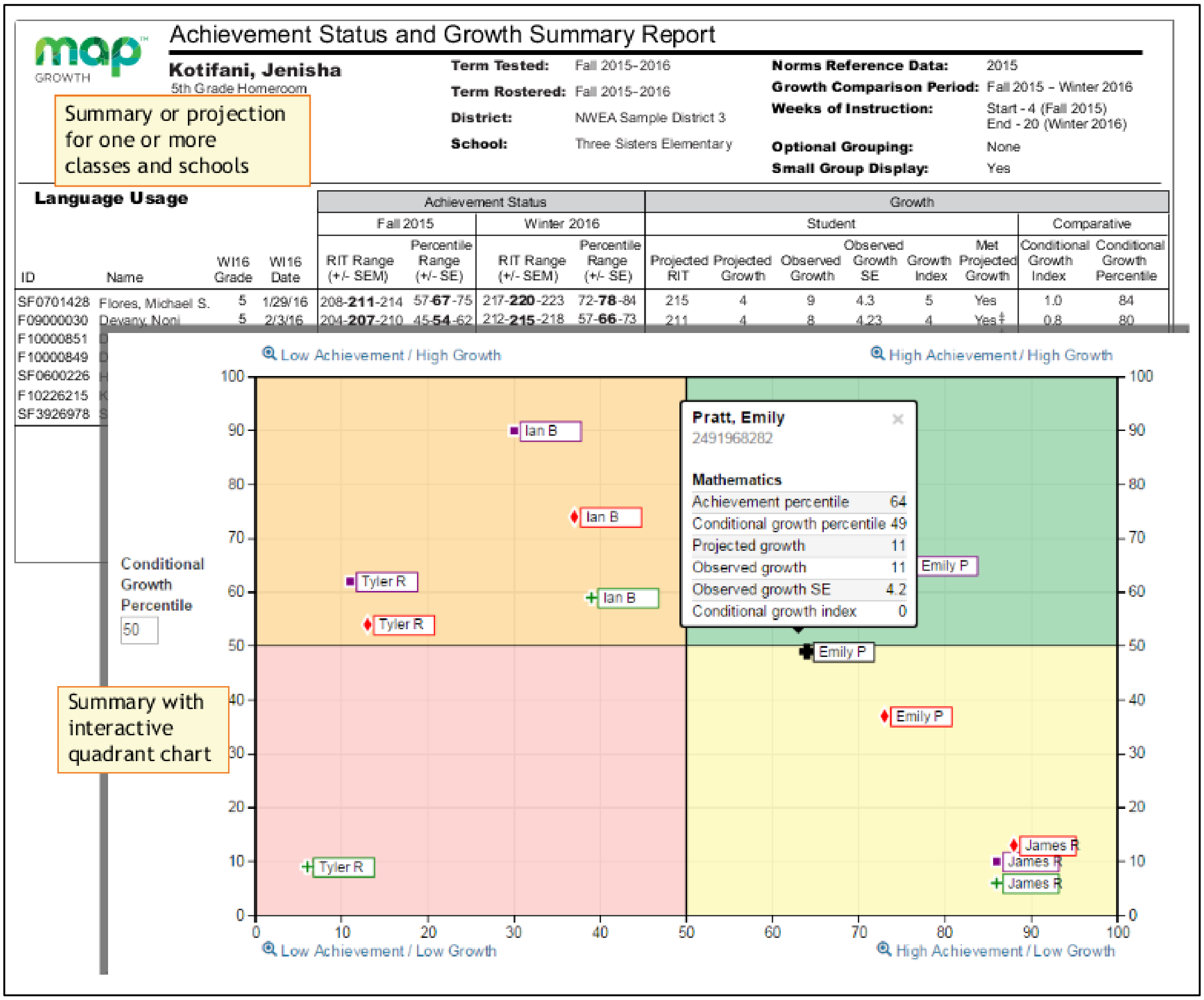


Figure 6.6. Class Breakdown by Projected Proficiency Report

Class Breakdown by Projected Proficiency | Create a PDF version of this report Letter 8 1/2x11" | Create PDF

Projected to: **ACT College Readiness** taken in **spring**.

View Linking Study: https://www.nwea.org/resources/map-college-readiness-benchmarks/

Subject	Projected Proficiency Category		
	Not On Track	On Track 22	On Track 24
Mathematics	V.E. Stoll (213) A.E. Reynolds (216) K.E. Walton (216) J.N. Rudberg (218) A.M. Ahmed (222) L.T. Stjern (222) K.M. Baumgartner (228) A.E. Cosentino (230) K.R. Dotson (230) C.R. Rust (230) J.E. Matzke (232) J.L. McMurphy (235) J.E. Carson (237)	C.R. Mariancito (238) S.A. Steinhorst (238) P.E. Payton (241) S.A. Rasmussen (242) K.R. VanMinsel (242)	C.A. Elder (246) M.O. Ponsler (246) K.A. Schubick (246) J.A. Williford (246) V.N. Sebastian (249) J.A. Gustave (256) A.E. Shadden (260)
Reading	J.A. Williford (199) K.M. Baumgartner (201) J.E. Matzke (204) J.N. Rudberg (207) A.E. Reynolds (208) K.E. Walton (210) K.R. VanMinsel (211) C.R. Rust (214) P.E. Payton (216) S.A. Rasmussen (218) J.A. Gustave (219) M.O. Ponsler (219) A.M. Ahmed (220) J.L. McMurphy (222)	A.E. Cosentino (224) C.R. Mariancito (224) S.A. Steinhorst (226)	V.E. Stoll (227) K.A. Schubick (228) V.N. Sebastian (230) J.E. Carson (231) K.R. Dotson (233) C.A. Elder (233) A.E. Shadden (233) L.T. Stjern (242)

6.1.3. District-Level Reports

To help districts assess performance trends by grade and school, NWEA provides district-level reports that present historical data for a school and are valuable in planning and monitoring school improvement plans. District-level reports include the District Summary, Student Growth Summary, Projected Proficiency Summary, Grade, and Grade Breakdown reports.

- The District Summary Report, Figure 6.7, summarizes school and grade data to help identify trends and isolate areas of strength or concern. It includes average performance and SD by instructional area.

- To help administrators assess achievement and growth performance and see the percentage of students meeting targets, the Student Growth Summary Report, Figure 6.8, gives school and district leaders aggregated and comparative data at the grade level for an entire school or district.

- Prior to taking a state or college readiness assessment, the Projected Proficiency Summary Report, Figure 6.9, provides an aggregate view of students' predicted performance. This report helps identify groups for remediation work, helps determine instructional strategy, and informs district and school improvement plans.

- The Grade Report in Figure 6.10 shows students' summary test data by grade from a selected term. Educators can use this data to determine strengths and weaknesses and set goals with departments and instructors. Educators can also compare schools within the district by looking at the grade at a whole. The Grade Report is available in multiple views, similar to the Class Report.

- Similar to the Class Breakdown report at the class level, a Grade Breakdown Report, Figure 6.11, provides a single spreadsheet of student achievement to groups of students from across the school. This data extract can be used to identify groups of students with a similar instructional level in an instructional area for differentiated instruction. Unlike the Class Breakdown reports, this report has no limit on the number of students and is available in CSV format only.

Figure 6.7. District Summary Report

map GROWTH

District Summary Report

Aggregate by School

Term: Spring 2015-2016
District: NWEA Sample District PA
Grouping: None
Small Group Display: No

Reading

St. Helens Elementary School

MAP Growth: Reading 2-5 Common Core 2010
V2 Common Core English Language Arts K-12:

						Goal Performance					
						Literature		Informational Text		Vocabulary Acquisition and Use	
Term	Grade	Student Count	Mean RIT	Std Dev	Median	Mean	Std Dev	Mean	Std Dev	Mean	Std Dev
Winter 2015-2016	13	1									
Spring 2013-2014	2	41	179.8	10.3	178	179.8	11.6	178.5	11.9	179.0	13.6
Winter 2013-2014	2	41	169.8	8.9	167	169.6	10.7	169.8	10.9	171.5	10.9
Fall 2013-2014	2	41	158.4	6.0	158	158.0	8.2	158.9	8.4	157.4	7.8
Winter 2015-2016	5	27	206.9	9.5	207	205.0	10.7	204.1	10.7	209.2	10.0
Fall 2015-2016	5	27	201.5	10.4	201	202.2	11.6	201.7	10.5	200.7	11.8
Spring 2014-2015	5	21	206.9	11.0	206	208.9	13.2	206.6	11.3	208.1	13.4
Winter 2014-2015	5	21	204.1	10.0	203	205.0	10.0	204.6	12.2	205.5	11.2
Fall 2014-2015	5	21	199.2	8.8	198	197.9	8.9	200.6	10.8	200.0	12.1
Spring 2013-2014	5	71	211.4	9.2	212	211.1	11.4	210.9	10.6	212.4	10.6
Winter 2013-2014	5	71	209.3	9.4	209	208.3	10.9	209.0	11.5	209.6	11.7
Fall 2013-2014	5	71	202.8	11.5	202	201.7	13.4	202.8	12.6	202.0	12.5

MAP Growth K-2: Reading Common Core 2010
Common Core English Language Arts K-12: 2010

						Goal Performance							
						Foundational Skills		Language and Writing		Literature and Informational		Vocabulary Use and Functions	
Term	Grade	Student Count	Mean RIT	Std Dev	Median	Mean	Std Dev	Mean	Std Dev	Mean	Std Dev	Mean	Std Dev
Winter 2015-2016	13	19	151.1	11.0	149	152.6	11.6	151.1	12.7	151.5	13.0	152.4	9.9
Fall 2015-2016	13	20	136.6	9.7	134	135.4	11.8	137.9	10.1	134.7	10.5	134.4	9.6
Spring 2014-2015	13	29	158.9	11.8	159	157.3	14.7	158.5	14.4	158.4	14.9	159.2	14.7
Winter 2014-2015	13	30	146.1	8.7	140	146.0	9.1	146.0	11.0	143.4	10.0	145.6	10.2
Fall 2014-2015	13	30	134.9	8.9	129	136.9	11.5	135.0	9.9	134.2	11.7	134.1	12.1
Winter 2015-2016	1	16	179.4	10.3	180	177.8	10.9	178.5	9.7	178.0	12.0	179.6	12.7
Fall 2015-2016	1	16	162.8	10.0	164	160.9	11.0	162.9	13.0	161.4	9.9	164.5	10.4
Spring 2014-2015	1	16	177.8	10.1	179	177.6	12.3	178.5	9.7	175.1	13.5	179.4	11.4
Winter 2014-2015	1	16	169.9	9.5	172	170.1	10.1	171.8	8.8	170.8	10.5	169.9	12.6
Fall 2014-2015	1	16	156.9	9.8	159	155.9	9.4	155.1	9.8	156.7	8.2	158.4	9.3

Figure 6.8. Student Growth Summary Report

map GROWTH

Student Growth Summary Report

Aggregate by School

Term: Spring 2016-2017
District: NWEA Sample District - Partner Accounts

Norms Reference Data: 2015 Norms
Growth Comparison Period: Fall 2016 - Spring 2017
Weeks of Instruction: Start - 4 (Fall 2016)
End - 36 (Spring 2017)
Grouping: None
Small Group Display: No

Mt. Bachelor Middle School

Language Usage

		Comparison Periods								Growth Evaluated Against						
		Fall 2016			Spring 2017			Growth		School Norms			Student Norms			
Grade (Spring 2017)	Growth Count‡	Mean RIT	SD	Percentile	Mean RIT	SD	Percentile	Observed Growth	Observed Growth SE	Projected Growth	School Conditional Growth Index	School Conditional Growth Percentile	Count with Projection	Count Met Projection	Percent Met Projection	Student Median Conditional Growth Percentile
6	83	223.4	11.5	99	232.3	12.1	99	8.9	0.8	4.3	2.21	99	83	68	82	80
7	119	230.6	12.7	99	235.2	12.4	99	4.6	0.9	3.5	0.52	70	119	91	76	75
8	58	226.7	11.0	95	233.1	13.1	98	6.4	1.5	3.2	1.48	93	58	37	64	74

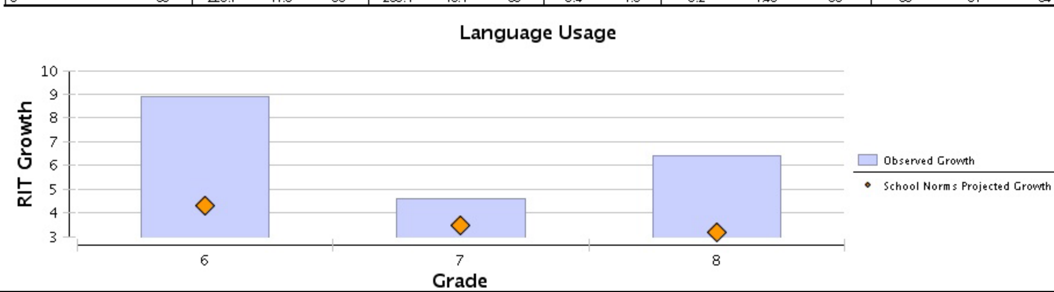


Figure 6.9. Projected Proficiency Summary Report

Projected Proficiency Summary Report

Aggregate by District by Grade

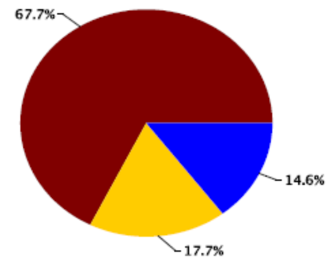
Term Tested:	Fall 2017-2018
District:	NWEA Sample District - General Use
Grouping:	None

Mathematics

Projected to: **ACT College Readiness** taken in **spring**.

View Linking Study: https://www.nwea.org/resources/map-college-readiness-benchmarks/

Grade	Student Count	Not On Track		On Track 22		On Track 24	
		Count	Percent	Count	Percent	Count	Percent
5	40	33	82.5%	3	7.5%	4	10.0%
6	93	83	89.2%	5	5.4%	5	5.4%
7	116	69	59.5%	21	18.1%	26	22.4%
8	135	82	60.7%	29	21.5%	24	17.8%
9	104	57	54.8%	32	30.8%	15	14.4%
10	38	32	84.2%	3	7.9%	3	7.9%
Total	526	356	67.7%	93	17.7%	77	14.6%



Projected to: **NWEA Generic Linking Study** taken in **spring**.

View Linking Study:

Grade	Student Count	Below Standards		Proficient		Advanced	
		Count	Percent	Count	Percent	Count	Percent
1	16	5	31.3%	8	50.0%	3	18.8%
2	30	6	20.0%	13	43.3%	11	36.7%
3	1	1	100.0%	0	0.0%	0	0.0%
4	41	17	41.5%	15	36.6%	9	22.0%
5	40	12	30.0%	22	55.0%	6	15.0%
6	93	47	50.5%	37	39.8%	9	9.7%
7	116	32	27.6%	42	36.2%	42	36.2%
8	135	37	27.4%	47	34.8%	51	37.8%
9	104	17	16.3%	40	38.5%	47	45.2%
10	38	22	57.9%	9	23.7%	7	18.4%
Total	614	196	31.9%	233	37.9%	185	30.1%

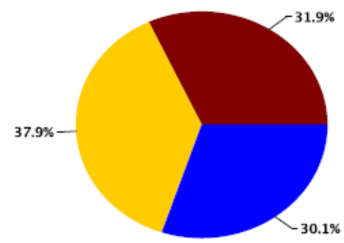


Figure 6.10. Grade Report

Grade Report

Grade 10

Term:	Fall 2017-2018	**Norms Reference Data:**	2015
District:	NWEA Sample District - Partner Accc	**Weeks of Instruction:**	5 (Fall 2017)
School:	Mt. Hood High School	**Grouping:**	None
		Small Group Display:	No

Mathematics

MAP Growth: Math 6+ / Common Core

Summary	
Total Students With Valid Growth Test Scores	27
Mean RIT	226.3
Standard Deviation	9.4
District Grade Level Mean RIT	*
Students At or Above District Grade Level Mean RIT	*
Norm Grade Level Mean RIT	230.2
Students At or Above Norm Grade Level Mean RIT	8

	Lo %ile < 21		LoAvg %ile 21-40		Avg %ile 41-60		HiAvg %ile 61-80		Hi %ile > 80		Mean RIT (+/- Smp Err)	Std Dev
Overall Performance	count	%	count	%	count	%	count	%	count	%		
Demo Growth: Math 6+ / Demonstration Tests - NWEA 2017	4	15%	6	22%	13	48%	4	15%	0	0%	224-**226**-228	9.4
Goal Area												
Statistics and Probability	3	11%	10	37%	10	37%	3	11%	1	4%	223-**226**-228	10.9
Geometry	3	11%	8	30%	10	37%	6	22%	0	0%	223-**226**-228	11.7
Operations and Algebraic Thinking	3	11%	9	33%	10	37%	5	19%	0	0%	225-**227**-229	11.4
The Real and Complex Number Systems	4	15%	6	22%	10	37%	7	26%	0	0%	225-**227**-229	11.5

Figure 6.11. Grade Breakdown Report

Includes whatever schools, grades, subjects you choose

Shows goal areas for the subjects/tests (blank if not applicable)

Student M.I.	Term Tested	Term Roster	School	Grade	Subject	Test RIT Score	Test RIT 10 Point Range	Assessment Name	Mathematics: Geometry	Measurement and Data
Michael	Fall 2014-	Fall 2014-	LaView Elem	5	Mathemati	233	231-240	MAP: Math 2-5	231-240	231-240
JaShae	Fall 2014-	Fall 2014-	LaView Elem	5	Mathemati	229	221-230	MAP: Math 2-5	241-250	221-230
Smith	Fall 2014-	Fall 2014-	LaView Elem	5	Mathemati	233	231-240	MAP: Math 2-5	251-260	231-240
Gage	Fall 2014-	Fall 2014-	Dill Middle S	6	Mathemati	165	161-170	MAP: Math 6+ (	151-160	
Reginald	Fall 2014-	Fall 2014-	Dill Middle S	6	Mathemati	157	151-160	MAP: Math 6+ (	161-170	
Michael	Fall 2014-	Fall 2014-	Dill Middle S	6	Mathemati	164	161-170	MAP: Math 6+ (	161-170	

6.1.4. Learning Continuum

The learning continuum, designed for classroom use, translates MAP Growth scores to learning statements that show what students performing at a given RIT level on MAP Growth assessments are typically ready to learn to allow teachers to set student goals and tailor instruction to student needs. The learning continuum identifies skills and concepts each student is ready to learn by showing relationships among standards, learning statements, and the student's RIT score. This helps educators bridge the gap between MAP Growth data and standards and/or intervention.

Educators can use data from the learning continuum to help develop focused, effective instructional plans and target instruction to an individual student's needs. For each identified instructional area and sub-area, the learning continuum provides a list of skills and concepts associated with a given RIT range. Educators can use the learning statements to differentiate core instruction focused on either standards or topics. Struggling students often have one or more instructional area scores that fall above or below the expected level for their grade. Teachers can identify these areas using MAP Growth reports and then incorporate the learning statements to help develop instructional interventions for struggling students or create customized learning paths.

The learning continuum has two views:

1. Class view: Groups students and learning statements by RIT score bands to show where students are and what they are ready to learn. Seeing the skills and concepts students need to develop in each sub-area can help inform teachers' decisions for grouping, differentiated instruction, and targeted interventions. The learning statements can be further organized by content standards or topics.
2. Test view: Organizes each test's learning statements by RIT band into three columns: introduce, develop, and reinforce. The teacher can view the learning statements aligned to grade-level standards or by topics.

    a. Introduce: The skills and concepts students may be able to learn with additional scaffolding or pre-teaching
    b. Develop: The closest skills and concepts students in a given RIT range are ready to learn today (i.e., their zone of proximal development)
    c. Reinforce: Skills and concepts where students show more independence, though they may need reinforcement to build consistent proficiency and confidence

Figure 6.12. Learning Continuum Class View

6.2. Quality Assurance

The NWEA Quality Assurance team validates all business rules and formulas applied when generating results for both standard reports provided via the assessment platform and all custom reports or data extracts. NWEA employs a software quality assurance process within the software development lifecycle that routinely checks the developed software to ensure that it meets desired quality measures. Software quality assurance processes test for quality in each phase of development. NWEA also employs several other approaches to ensure the integrity of the software, as described in Table 6.3.

Table 6.3. Ensuring Software Integrity

Approach	Description
Ad-Hoc Testing	A testing phase where the tester tries to “break” the system by randomly trying the system’s functionality.
Black Box Testing	Functional testing based on requirements with no knowledge of the internal program structure or data. Black box testing indicates whether a program meets required specifications by spotting faults of omission — places where the specification is not fulfilled.
Boundary Testing	Testing that focuses on the boundary or limit conditions of the software being tested.
Breadth Testing	A test suite that exercises the full functionality of a product but does not test features in detail.
Browser/Platform Testing	A test suite that exercises cross-platform web application accessibility from any of various web browsers within different operation systems.
Concurrency Testing/Group Testing	Multi-user testing geared toward determining the effects of accessing the same application code, module, or database records.
Depth Testing	A test that exercises a feature of a product in full detail.
End-to-End Testing	Testing a complete application environment in a situation that mimics real-world use, such as interacting with a database, using network communications, or interacting with other hardware, applications, or systems if appropriate.
Exploratory Testing	Exploratory testing seeks to find out how the software works and to ask questions about how it will handle difficult and easy cases. The tester configures, operates, observes, and evaluates the product and its behavior, critically investigating the result, and reporting information that seems likely to be a bug.
Functional Testing	Application test derived from the specified functional requirements without regard to the final program structure.
Reliability Testing	Confirms that the application under test recovers from expected or unexpected events without loss of data or functionality.
Negative Testing	Testing aimed at showing software does not work.
Performance Testing	Testing conducted to evaluate the compliance of a system or component with specified performance requirements. Often this is performed using an automated test tool to simulate large number of users. Also known as “load testing.”
Regression Testing	Selective retesting to detect faults introduced during modification of an application or system component, to verify that modifications have not caused unintended adverse effects, or to verify that a modified application or system component still meets its specified requirements.
Scalability Testing	Performance testing focused on ensuring the application under test gracefully handles increases in workload.
Smoke Testing	A scaled-down regression test of an applications major functionality.
Stress Testing	Testing conducted to evaluate a system or component at or beyond the limits of its specified requirements to determine the load under which it fails and how.
System Testing	System-level tests verify proper execution of all application components, including interfaces to other applications. Tests are performed to verify that the system meets both functional and nonfunctional requirements.
Unit Testing	The testing is done to show whether a unit (the smallest piece of software that can be independently compiled or assembled, loaded, and tested) satisfies its functional specification or its implemented structure matches the intended design structure.

Chapter 7: Reliability

Reliability refers to the consistency of scores obtained from the assessment. It reflects the absence of random measurement error. When the measurement error is large, reliability is small, and vice versa. Increasing reliability by minimizing error is an important goal for any test. Different sources of measurement error affect scores. The effect of each particular source of error has a corresponding reliability coefficient that describes the influence of that source on scores. One source of measurement error is time, or the instability of a construct over time, as measured by test-retest reliability. If this source of error is low, the test-retest reliability coefficient will be high. Another source of measurement error is the items selected for a test. Internal consistency, or marginal reliability, will be high if measurement error due to items is low.

It is important to report multiple reliability coefficients to describe the influence of different sources of error. Therefore, the reliability of the MAP Growth assessments was examined in the following ways:

- **Test-retest reliability** that demonstrates the consistency of MAP Growth assessments across time by administering it to a group of students two times separated by a reasonable period of time. The question being answered with this type of reliability is "To what extent does the test administered to the same students twice yield the same results from one administration to the next?"

- **Marginal reliability** that examines a test's consistency across items. The question being answered with this type of reliability is "To what extent do items in the test measure the test's construct(s) in a consistent manner?"

- **Score precision** based on the standard error of measurement (SEM) of MAP Growth scores

Data included in these analyses were from the Fall 2016, Winter 2017, Spring 2017, and Fall 2017 administrations of the MAP Growth assessments for use with the CCSS and NGSS. See Appendix A for the number of students included in the sample by state and demographics.

7.1. Test-Retest Reliability

MAP Growth affords the means to assess students on multiple occasions (e.g., fall, winter, and spring) during the school year. Thus, test-retest reliability is key as it provides insight into the consistency of MAP Growth across time. The adaptive nature of MAP Growth assessments requires reliability to be examined using non-traditional methods because dynamic item selection is an integral part of MAP Growth. Parallel forms are restricted to identical item content from a common goal structure, but the item difficulties depend on the student's responses to previous items on the test. Therefore, test-retest reliability of MAP Growth is more accurately described as a mix between test-retest reliability and a type of alternate forms reliability, both of which are spread across several months versus the typical two or three weeks. The second test (or retest) is not the same test. Rather, it is one that is comparable to the first by its content and structure, differing only in the difficulty level of its items. In other words, test-retest with alternate forms (Crocker & Algina, 1986) describes the influence of two sources of measurement error: time and item selection.

Specifically, test-retest with alternate forms reliability for MAP Growth was estimated via the Pearson correlation between MAP Growth RIT scores of students taking MAP Growth in two consecutive terms (e.g., Fall 2016 and Winter 2017, Winter 2017 and Spring 2017, and Spring 2017 and Fall 2017). Table 7.1 presents test-retest reliability results by grade, and Appendix C presents the values by state and grade for each content area with n-counts greater than 300. The grade level is based on students' actual grade levels. The coefficients in Table 7.1 are generally higher than 0.80 except at some lower grade levels such as kindergarten. Results in Appendix C suggest high correlations and similar patterns across states. These results provide evidence that students' MAP Growth scores are highly consistent for students at different grade levels and from different states.

Table 7.1. Test-Retest with Alternate Forms Reliability by Grade

	Fall 2016 – Winter 2017		Spring 2017 – Fall 2017*		Winter 2017 – Spring 2017	
Grade	N	Reliability	N	Reliability	N	Reliability
Reading						
K	177,448	0.687	154,290	0.797	209,749	0.759
1	241,392	0.824	190,741	0.789	253,565	0.857
2	292,918	0.855	242,516	0.847	310,425	0.867
3	312,725	0.857	258,650	0.861	321,320	0.862
4	314,025	0.862	264,366	0.863	321,602	0.864
5	308,664	0.863	259,945	0.855	316,185	0.864
6	281,851	0.857	239,809	0.856	282,554	0.859
7	270,295	0.855	235,353	0.854	267,978	0.856
8	261,713	0.852	86,688	0.836	252,876	0.851
9	97,345	0.834	67,889	0.839	87,972	0.841
10	79,370	0.823	27,345	0.834	70,579	0.825
11	35,972	0.807	9,564	0.818	27,794	0.795
12	11,910	0.780	–	–	7,124	0.777
Language Usage						
2	50,183	0.853	36,542	0.865	48,880	0.876
3	77,264	0.857	58,795	0.860	69,224	0.871
4	83,781	0.861	64,072	0.862	76,413	0.871
5	81,667	0.866	59,331	0.863	75,034	0.871
6	82,681	0.865	63,039	0.869	74,601	0.871
7	76,736	0.866	63,225	0.874	66,717	0.868
8	74,602	0.867	19,975	0.856	63,062	0.874
9	33,715	0.847	23,760	0.857	28,314	0.855
10	30,742	0.843	11,420	0.861	25,485	0.846
11	15,626	0.835	3,556	0.862	12,142	0.833
12	3,844	0.807	–	–	2,366	0.841

Grade	Fall 2016 – Winter 2017		Spring 2017 – Fall 2017*		Winter 2017 – Spring 2017	
	N	Reliability	N	Reliability	N	Reliability
Mathematics						
K	188,211	0.753	167,115	0.816	219,743	0.796
1	253,970	0.835	203,863	0.794	265,331	0.856
2	300,344	0.847	248,567	0.800	316,179	0.855
3	315,437	0.861	260,792	0.877	323,572	0.870
4	316,016	0.884	266,765	0.898	323,570	0.889
5	312,928	0.904	264,228	0.898	319,027	0.907
6	293,312	0.905	244,552	0.916	291,348	0.908
7	276,811	0.915	236,430	0.925	274,727	0.917
8	268,597	0.919	80,827	0.915	259,051	0.920
9	98,106	0.907	65,719	0.915	88,247	0.906
10	79,053	0.897	30,004	0.906	70,087	0.900
11	38,849	0.893	9,685	0.902	30,701	0.881
12	12,122	0.855	–	–	7,017	0.847
Science**						
3	12,631	0.792	12,088	0.806	11,012	0.812
4	16,713	0.798	15,218	0.820	15,804	0.812
5	21,045	0.825	16,436	0.813	19,865	0.841
6	21,773	0.816	21,717	0.821	20,833	0.833
7	20,496	0.830	23,055	0.840	20,316	0.844
8	22,633	0.837	4,460	0.825	21,853	0.847
9	4,854	0.835	2,876	0.859	4,424	0.846
10	3,906	0.851	1,510	0.841	3,380	0.839
11	1,321	0.829	301	0.789	986	0.846

*The Spring 2017 – Fall 2017 correlations do not include Grade 12 because all Grade 12 students that took the Spring 2017 test had graduated by Fall 2017 and did not take MAP Growth.
**Grade 12 isn't included for Science because the sample size was less than 300.

7.2. Marginal Reliability (Internal Consistency)

Internal consistency measures how well the items on a test that reflect the same construct yield similar results. Determining the internal consistency of MAP Growth tests is challenging because traditional methods depend on all test takers taking a common test consisting of the same items. Application of these methods to adaptive tests is statistically cumbersome and inaccurate. Fortunately, an equally valid alternative is available in the marginal reliability coefficient (Samejima, 1977, 1994) that incorporates measurement error as a function of the test score. In effect, it is the result of combining measurement error estimated at different points on the achievement scale into a single index. This method of calculating internal consistency, ρ_θ, yields results that are nearly identical to coefficient alpha when both methods are applied to the same fixed-form tests. The approach taken for MAP Growth was suggested by Wright (1999) and is given by:

$$\rho_\theta = \frac{\sigma_\theta^2 - M_{S_\theta^2}}{\sigma_\theta^2} \tag{7.1}$$

where σ_{θ}^{2} is the observed variance of the achievement estimates, θ, (the RIT score) and $M_{S_{\theta}^{2}}$ is the observed mean of the score's conditional error variances at each value of θ. Tests are considered of sound reliability when their marginal reliability coefficients range from 0.80 and above.

Table 7.2 presents the marginal reliabilities of RIT scores by content area and grade. Table 7.3 – Table 7.8 present the marginal reliabilities of RIT scores by instructional area. The overall marginal reliabilities for all grades and content areas are in the .90s, which suggests that MAP Growth tests have high internal consistency. Science has slightly lower reliability values, which may be due to their shorter test lengths. Marginal reliabilities are noticeably lower at the instructional area score level than the overall test scores. These reliability estimates will always be smaller in magnitude than the corresponding estimates for the overall test because instructional area scores are based on many fewer items and are therefore less precise than the overall scores.

Table 7.2. Marginal Reliability by Grade

Grade	N	Reliability	Mean SEM
Reading			
K	860,385	0.955	3.0
1	1,104,917	0.967	3.0
2	1,351,801	0.965	3.3
3	1,445,054	0.962	3.4
4	1,440,186	0.960	3.4
5	1,440,235	0.958	3.4
6	1,374,250	0.957	3.4
7	1,329,342	0.957	3.4
8	1,288,335	0.957	3.4
9	543,715	0.964	3.4
10	424,492	0.964	3.4
11	194,789	0.967	3.4
12	76,717	0.971	3.4
Language Usage			
2	237,133	0.969	3.0
3	374,261	0.966	3.0
4	405,948	0.963	2.9
5	406,982	0.961	2.9
6	424,438	0.961	2.9
7	403,828	0.961	2.9
8	391,904	0.960	2.9
9	193,601	0.965	2.9
10	169,162	0.965	3.0
11	83,983	0.968	3.0
12	28,229	0.973	3.0

Grade	N	Reliability	Mean SEM
Mathematics			
K	905,354	0.968	3.0
1	1,160,639	0.972	3.0
2	1,386,516	0.966	3.0
3	1,464,117	0.961	2.9
4	1,454,384	0.964	2.9
5	1,457,360	0.970	2.9
6	1,414,749	0.970	3.0
7	1,356,673	0.974	3.0
8	1,301,540	0.976	3.0
9	533,219	0.978	3.0
10	416,866	0.980	3.0
11	207,209	0.981	3.0
12	75,012	0.983	3.0
Science			
3	86,819	0.927	3.3
4	110,488	0.922	3.3
5	139,411	0.928	3.3
6	154,819	0.927	3.3
7	158,035	0.933	3.3
8	162,983	0.938	3.3
9	35,344	0.940	3.3
10	27,944	0.947	3.4
11	13,540	0.947	3.4
12	3,543	0.952	3.4

Table 7.3. Marginal Reliability by Instructional Area and Grade—Reading K–2

		Foundational Skills		Language & Writing		Literature & Informational		Vocabulary Use & Functions	
Grade	N	Reliability	Mean SEM	Reliability	Mean SEM	Reliability	Mean SEM	Reliability	Mean SEM
K	860,222	0.867	6.3	0.818	6.3	0.825	6.3	0.835	6.3
1	1,101,775	0.890	6.4	0.864	6.3	0.871	6.3	0.871	6.3
2	350,597	0.885	6.5	0.866	6.4	0.872	6.4	0.870	6.4

Table 7.4. Marginal Reliability by Instructional Area and Grade—Reading 2–12

		Literary Text		Informational Text		Vocabulary	
Grade	N	Reliability	Mean SEM	Reliability	Mean SEM	Reliability	Mean SEM
2	1,001,204	0.879	6.4	0.887	6.4	0.883	6.4
3	1,437,551	0.872	6.5	0.873	6.5	0.869	6.4
4	1,435,809	0.868	6.4	0.864	6.4	0.860	6.4
5	1,437,257	0.865	6.5	0.858	6.4	0.854	6.4
6	1,372,960	0.858	6.5	0.854	6.5	0.849	6.5
7	1,328,700	0.860	6.5	0.856	6.5	0.850	6.5
8	1,287,725	0.859	6.5	0.855	6.5	0.847	6.5
9	543,439	0.880	6.5	0.876	6.5	0.870	6.6
10	424,255	0.883	6.5	0.877	6.5	0.872	6.6
11	194,609	0.890	6.6	0.884	6.6	0.881	6.6
12	76,562	0.897	6.7	0.892	6.7	0.892	6.7

Table 7.5. Marginal Reliability by Instructional Area and Grade—Language Usage 2–12

		Writing		Language: Understand, Edit for Grammar, Usage		Language: Understand, Edit for Mechanics	
Grade	N	Reliability	Mean SEM	Reliability	Mean SEM	Reliability	Mean SEM
2	237,133	0.891	5.3	0.921	5.3	0.914	5.3
3	374,261	0.896	5.3	0.907	5.2	0.906	5.2
4	405,948	0.894	5.2	0.895	5.2	0.897	5.2
5	406,982	0.894	5.2	0.886	5.2	0.888	5.2
6	424,438	0.896	5.2	0.883	5.2	0.886	5.2
7	403,828	0.898	5.2	0.881	5.2	0.884	5.2
8	391,904	0.899	5.2	0.881	5.2	0.883	5.2
9	193,601	0.912	5.2	0.893	5.2	0.895	5.2
10	169,162	0.911	5.3	0.892	5.2	0.893	5.3
11	83,983	0.917	5.3	0.902	5.3	0.901	5.3
12	28,229	0.928	5.3	0.916	5.3	0.914	5.3

Table 7.6. Marginal Reliability by Instructional Area and Grade—Mathematics K–2

		Operations & Algebraic Thinking		Number & Operations		Measurement & Data		Geometry	
Grade	N	Reliability	Mean SEM	Reliability	Mean SEM	Reliability	Mean SEM	Reliability	Mean SEM
K	905,183	0.887	6.4	0.878	6.3	0.862	6.3	0.880	6.3
1	1,156,961	0.882	6.4	0.894	6.3	0.881	6.3	0.906	6.4
2	369,099	0.873	6.5	0.891	6.4	0.893	6.4	0.912	6.5

Table 7.7. Marginal Reliability by Instructional Area and Grade—Mathematics 2–12

Grade	#Test Events	Algebraic Thinking		Number & Operations		Measurement & Data		Geometry		The Real & Complex Number Systems		Statistics & Probability	
		R	Mean SEM	R	Mean SEM	R	Mean SEM	R	Mean SEM	R	Mean SEM	R	Mean SEM
2	1,017,417	0.856	6.1	0.847	6.1	0.854	6.1	0.869	6.1	0.921	6.1	0.918	6.1
3	1,457,285	0.865	6.1	0.836	6.1	0.860	6.1	0.853	6.1	0.906	6.1	0.904	6.1
4	1,450,373	0.866	6.1	0.857	6.1	0.873	6.1	0.865	6.1	0.930	6.2	0.929	6.2
5	1,454,634	0.873	6.1	0.887	6.1	0.892	6.1	0.876	6.2	0.904	6.1	0.913	6.1
6	1,413,485	0.874	6.1	0.947	6.2	0.942	6.2	0.882	6.1	0.884	6.1	0.889	6.1
7	1,356,078	0.893	6.1	0.948	6.2	0.942	6.2	0.897	6.1	0.898	6.1	0.905	6.1
8	1,300,948	0.907	6.1	0.951	6.2	0.948	6.2	0.905	6.1	0.905	6.2	0.911	6.2
9	532,966	0.917	6.2	0.941	6.2	0.937	6.2	0.914	6.2	0.910	6.2	0.917	6.2
10	416,659	0.921	6.2	0.908	6.2	0.905	6.2	0.919	6.2	0.917	6.2	0.919	6.2
11	207,038	0.927	6.2	0.920	6.2	0.914	6.2	0.922	6.2	0.923	6.2	0.922	6.2
12	74,870	0.933	6.3	0.920	6.2	0.915	6.2	0.925	6.3	0.928	6.3	0.926	6.3

Table 7.8. Marginal Reliability by Instructional Area and Grade—Science 3–12

Grade	N	Life Science		Physical Science		Earth & Space Science	
		Reliability	Mean SEM	Reliability	Mean SEM	Reliability	Mean SEM
3	86,819	0.820	5.7	0.798	5.9	0.786	5.9
4	110,488	0.811	5.8	0.783	5.9	0.776	5.8
5	139,411	0.822	5.9	0.798	5.9	0.793	5.8
6	154,819	0.810	5.8	0.794	5.9	0.796	5.9
7	158,035	0.819	5.9	0.813	5.9	0.811	5.9
8	162,983	0.835	5.9	0.826	6.0	0.821	6.0
9	35,344	0.840	5.9	0.831	6.0	0.827	6.0
10	27,944	0.864	6.0	0.848	6.0	0.834	6.0
11	13,540	0.863	6.0	0.857	6.0	0.823	6.0
12	3,543	0.871	6.0	0.869	6.1	0.843	6.1

Appendix D presents marginal reliabilities of overall RIT scores by state and grade and by instructional area and state. These results show that the marginal reliabilities are in the .90s and that the general patterns of marginal reliabilities are consistent across states. Measurement error is shown to be a minimal portion of the overall score variance of the MAP Growth tests.

7.3. Score Precision

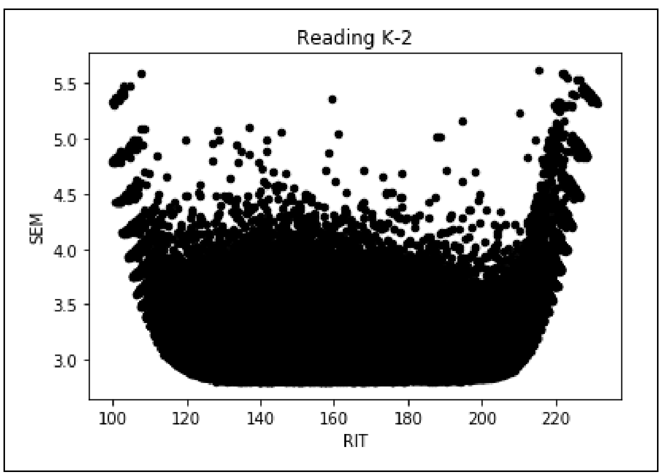
Score precision of MAP Growth scores is measured by the standard error of measurement (SEM), a function of the relationship among item parameters, the ability of the student, and the number of items administered. SEM is related to reliability in that it estimates how repeated measures of a student on the same assessment tend to be distributed around their “true” score. The SEM is the inverse of the square root of test information. Score precision is best when students are given items closely matched to their abilities. Lower values of SEM indicate greater precision in the score. With greater score precision across a broad range of ability, several benefits follow:

- Differences between similar students become more apparent. Because there is a direct mathematical relationship between test information and SEM, lower SEM indicates greater test information. This means that the level of test information observed across a group of students from even a wide grade span should be comparable across the achievement range.
- When change in student scores from one test occasion to another is of interest, measurement errors accrue with each test occasion. The greater the precision of individual scores, the greater the likelihood of drawing reliable conclusions about changes in student status over time.
- Classification accuracy will be improved as the level of score precision is increased.

The MAP Growth adaptive test algorithm selects the best items for each student, producing a significantly lower SEM than fixed-form tests. MAP Growth tests yield ability estimates with SEMs that are less than .30 of a typical large sample standard deviation (Kingsbury & Hauser, 2004). Standard errors vary minimally across more than 90% of the achievement range of a grade level. This makes MAP Growth scores well suited for use in growth models and other statistical procedures that assume additive measures.

Figure 7.1 – Figure 7.4 present the levels of SEM across the operational RIT range for MAP Growth tests by content area and grade band. Each figure has a noticeable fluctuation in SEMs at the very low and very high end of the RIT score distributions. All mean SEMs are below 4.5 RITs except at the very low and high levels of the RIT score distributions, which is to be expected. This consistency in MAP Growth SEMs across the RIT ranges of interest is particularly important when student change in performance is to be evaluated. Because MAP Growth is used to monitor students' progress over years, it is important that MAP Growth has similarly low SEMs across the RIT score range so that students at different ability levels are measured equally precisely.

Figure 7.1. Mean SEM of RIT Scores, Fall 2016 – Fall 2017—Reading



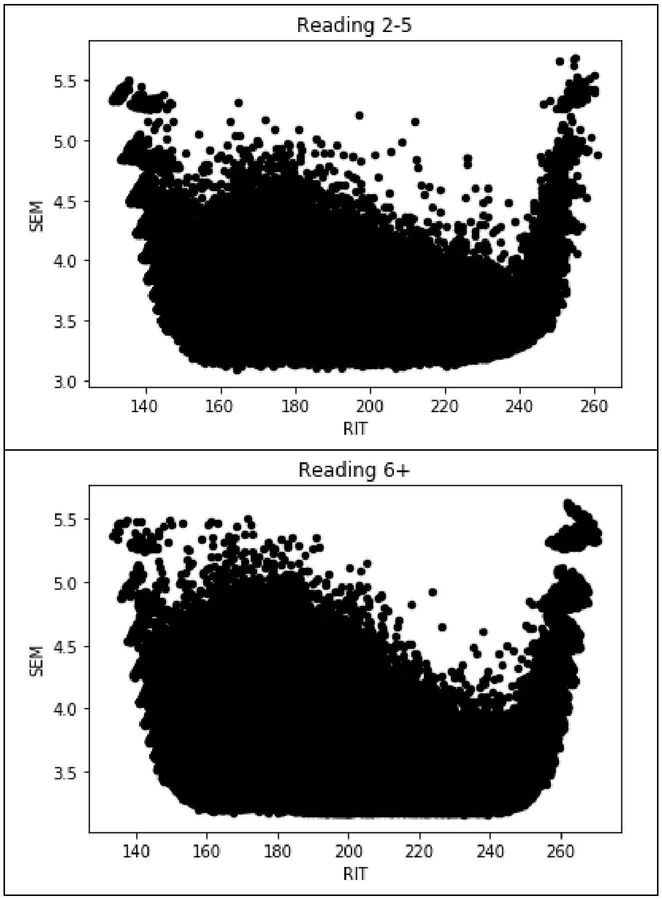


Figure 7.2. Mean SEM of RIT Scores, Fall 2016 – Fall 2017—Language Usage

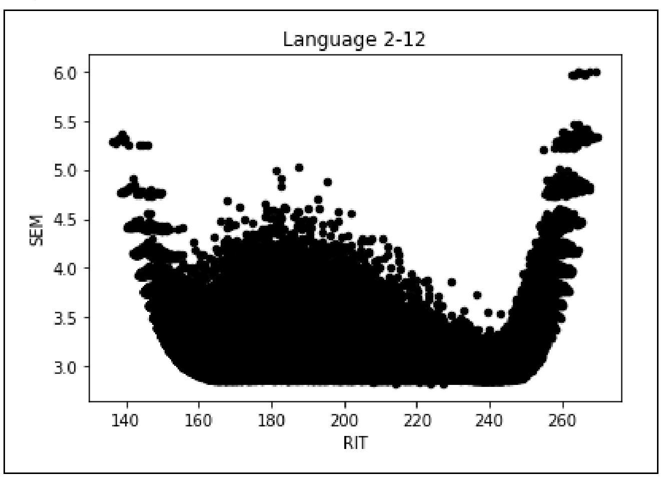


Figure 7.3. Mean SEM of RIT Scores, Fall 2016 – Fall 2017—Mathematics

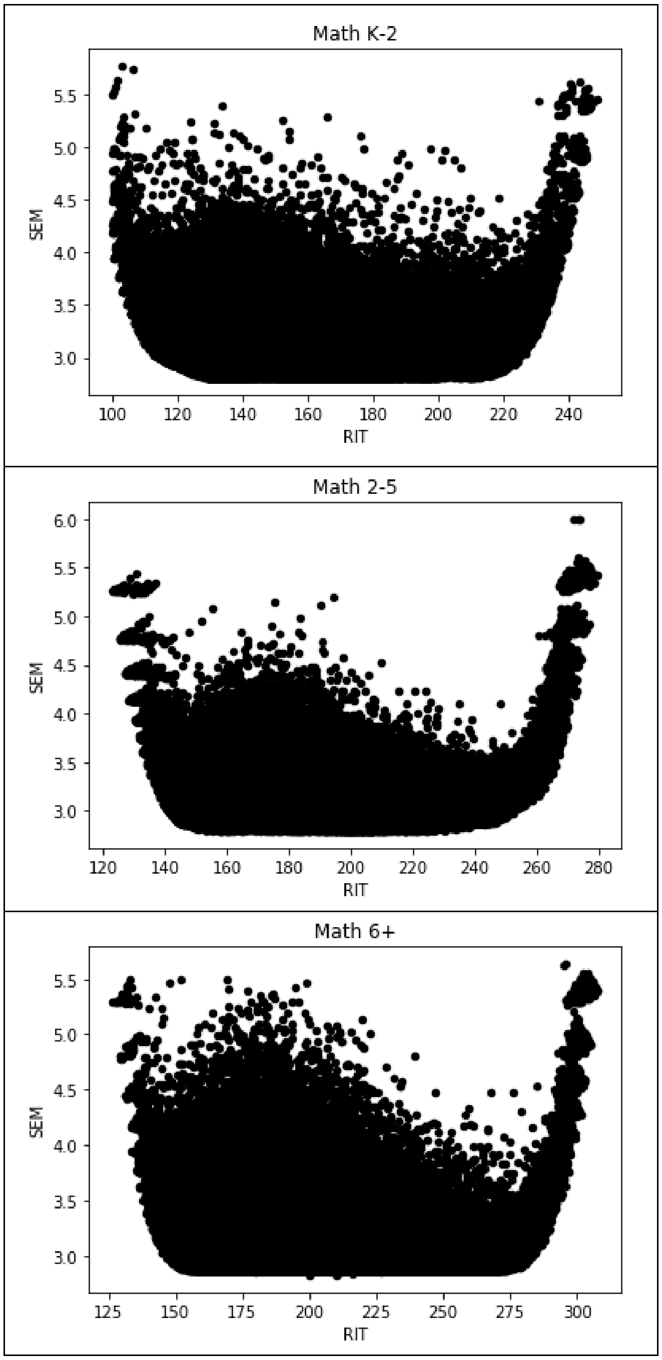
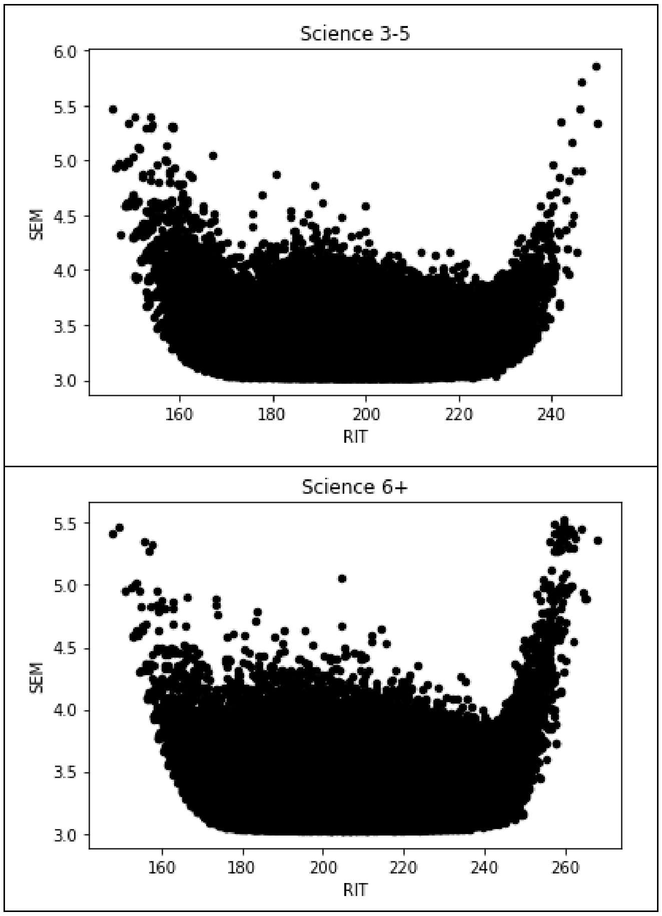


Figure 7.4. Mean SEM of RIT Scores, Fall 2016 – Fall 2017—Science



Chapter 8: Validity

Validity is defined as the "the degree to which evidence and theory support the interpretations of test scores for proposed uses. Validity is, therefore, the most fundamental consideration in developing tests and evaluating tests" (AERA, APA, & NCME, 2014, p. 11). It is not a quantifiable property but an ongoing process, beginning at initial conceptualization of the construct, continuing throughout the entire testing process, and extending into the interpretation and use of test sores. Validity evidence for MAP Growth assessments involves multiple sources including test content, internal structure, and relations to other variables.

8.1. Evidence Based on Test Content

Chapter 2 describes test content and alignment to standards, and Chapter 3 describes item development procedures. Evidence to support content validity is gathered during the internal review process for content standards and item quality. NWEA content specialists conducted an internal alignment analysis to assess how well and in what ways MAP Growth items align to the standards. This work examined and rated each item in the item bank against a content-specific rubric. It checked alignment to standards and helped to inform future item development.

EdMetric completed an external alignment study for MAP Growth (Egan & Davidson, 2017). Their study randomly sampled 20% of the MAP Growth item pools for use. Overall, 1,563 Reading items, 1,134 Language items, and 1,702 Mathematics items were evaluated. The study found that, on average, 97.4% of the items were aligned to the CCSS across all grades and content areas. The results showed that MAP Growth assessments have good alignment in terms of categorical concurrence, cognitive complexity, and range and balance of knowledge. Results also showed that there is strong evidence that the item pools cover the assessable CCSS within the NWEA blueprints (Egan & Davidson, 2017).

8.2. Evidence Based on Relations to Other Variables

Evidence based on relations to other variables (i.e., criterion-related validity) for MAP Growth includes concurrent validity and classification accuracy statistics. Table 8.1 presents a summary of the concurrent validity coefficients between MAP Growth and state test scores, as well as the overall classification accuracy results. Appendix E provides the concurrent validity estimates by state-specific assessments (including ACT Aspire, Partnership for Assessment of Readiness for College and Careers (PARCC), and Smarter Balanced Assessment Consortium (SBAC) assessments), and Appendix F presents the classification accuracy summary statistics by state. The following sections provide descriptions of concurrent validity and classification accuracy.

Table 8.1. Average Concurrent Validity (*r*) and Classification Accuracy (*p*)

Content Area	Grade	N	*r*	*p*
Reading	3	173,174	0.79	0.84
	4	170,767	0.80	0.84
	5	174,556	0.80	0.84
	6	163,305	0.79	0.84
	7	154,280	0.79	0.83
	8	138,007	0.78	0.82
	9	2,631	0.75	0.87
	10	2,791	0.78	0.87
	11	968	0.68	0.87

Content Area	Grade	N	*r*	*p*
Mathematics	3	171,233	0.82	0.86
	4	169,323	0.84	0.87
	5	173,605	0.84	0.87
	6	162,024	0.84	0.88
	7	151,649	0.84	0.88
	8	133,127	0.83	0.87
	9	2,706	0.72	0.88
	10	2,857	0.73	0.90
	11	975	0.73	0.87
Science	5	13,454	0.78	0.82
	8	4,220	0.79	0.86

8.2.1. Concurrent Validity

Concurrent validity is expressed in the form of a Pearson correlation coefficient between the total content area RIT score and the total score of another established and validated test designed to assess the same content area. It answers the question, “How well do the scores from this test that reference this scale (e.g., RIT scale) in this content area (e.g., Reading) correspond to the scores obtained from another test that references some other scale in the same content area?”

Concurrent validity requires that both tests are administered to the same students within a short amount of time. According to the National Center on Response to Intervention (NCRTI), acceptable concurrent validity is indicated when the correlations exceed 0.70 (NCRTI, 2016). Correlations in Table 8.1 are unweighted average correlation coefficients between MAP Growth scores and state assessment scores across states. As shown in the table, the average correlation coefficients range from 0.68 to 0.80 between scores on MAP Growth Reading and state tests, from 0.73 to 0.84 between MAP Growth Mathematics and state tests, and from 0.78 to 0.79 between MAP Growth Science and state tests.

8.2.2. Classification Accuracy of Predicting State Achievement Levels

NWEA produces linking studies for MAP Growth tests that allow users to predict proficiency status on state summative assessments.[6] Classification accuracy statistics indicate whether MAP Growth cut scores are good predictors of students’ proficiency status on the state summative assessment and can therefore be used as an indicator for criterion-related validity for MAP Growth, where the criterion is the observed proficiency status.

NWEA uses the equipercentile procedure to link state summative and MAP Growth scores. This procedure matches scores on the two scales that have the same percentile rank (i.e., the proportion of scores at or below each score). Consider the linked scores between two tests. Let x represent a score on Test X (e.g., a state summative assessment). Its equipercentile equivalent score on Test Y (e.g., MAP Growth), $e_y(x)$, can be obtained through a cumulative-distribution-based linking function defined in Equation 8.1:

$$e_y(x) = G^{-1}[P(x)] \tag{8.1}$$

[6] Linking study reports are available online at https://www.nwea.org/resource/type/linking-studies/.

where $e_y(x)$ is the equipercentile equivalent of score x of the state summative assessment on the scale of MAP Growth, $P(x)$ is the percentile rank of a given score on Test X, and G^{-1} is the inverse of the percentile rank function for scores on Test Y that indicates the scores on Test Y corresponding to a given percentile. Once linking tables between a state summative assessment and MAP Growth are created, the MAP Growth cut scores in the tables permit users to predict state summative proficiency status.

Table 8.2 presents the classification accuracy statistics included in Table 8.1 and Appendix F. The results show that MAP Growth accurately classified approximately 83% of Reading students, 87% of Mathematics students, and 83% of Science students. These numbers are high, suggesting that the MAP Growth cut scores are effective predictors of student proficiency status on the state summative assessments.

Table 8.2. Summary of Classification Accuracy Statistics

Classification Accuracy Statistic	Description*	Interpretation
Overall Classification Accuracy Rate	(TP + TN) / (total sample size)	The proportion of students in the study sample whose proficiency classification on the state test was correctly predicted by MAP Growth cut scores (Pommerich, Hanson, Harris, & Sconing, 2004).
False Positive (FP)	FP / (total sample size)	The proportion of below-proficient students who were incorrectly predicted by MAP Growth test to be proficient.
False Negative (FN)	FN / (total sample size)	The proportion of proficient students who were incorrectly predicted by MAP Growth test to be below proficiency.

8.3. Evidence Based on Internal Structure

The internal structure of a test should align with theoretical expectation and test design. The intended construct of MAP Growth assessments is student achievement of the content standards across time. NWEA has conducted a series of studies for MAP Growth tests, and the results indicate that the constructs underlying the tests remained consistent at different grades or time points (Wang, Jiao, & Zhang, 2013; Wang, McCall, Jiao, & Harris, 2013). These findings support using MAP Growth results to measure student achievement and learning. Other evidence based on internal structure (i.e., construct validity) includes results from test-taking engagement and differential item functioning (DIF) studies.

8.3.1. Test-taking Engagement

An implicit assumption in any testing situation is that examinees attempt each item with full engagement and effort. The absence of this productive test-taking behavior (i.e., test-taking disengagement) introduces construct-irrelevant variance and jeopardizes score interpretation. A score should be the product of the measured construct only, not a result of the measured construct and the degree of test-taking engagement. Test-taking engagement can be viewed as a prerequisite for validity arguments regarding uses of test scores for the intended purpose of testing (Hauser, Kingsbury, & Wise, 2008).

Disengaged test-taking tends to occur in low-stakes tests (Knekta, 2017; Wolf & Smith, 1995), but it rarely occurs for the full duration of a test (Wise & Kong, 2005; Wolf, Smith, & Birnbaum, 1995). Test-takers sometimes idiosyncratically engage and disengage during a test depending on the amount of reading and the cognitive demand required by test items (Wise & Kingsbury, 2016; Wolf, et al., 1995). Research has demonstrated that the structure of item response time distributions allows examinee behavior to be classified as a rapid-guessing or solution behavior (Wise & Kong, 2005) and aggregated into a composite measure of a test-taker's engagement during a test event (Wise, 2006).

A lack of student motivation has been shown to reduce mean scores by more than a half standard deviation (Wise & DeMars, 2005). Strategies for reducing this effect on a student's score include statistical score adjustments (Wang & Xu, 2015; Wise & DeMars, 2006) and effort monitoring. Score adjustments take place after a test event has concluded, but effort monitoring occurs during testing by intervening with messages to the student or prompts for a proctor to encourage test-taking engagement. Messages to disengaged students have been shown to positively affect student engagement and overall test performance (Kong, Wise, Harmes, & Yang, 2006; Wise, Bhola, & Yang, 2006). Research with MAP Growth has also shown that proctor notification improves test-taking engagement, test performance, and convergent validity evidence (Wise, Kuhfeld, & Soland, in press).

NWEA provides engagement information on score reports and employs multiple strategies for enhancing engagement, including student messages, test pauses, and proctor notification. The work of Wise, Kuhfeld, and Soland (in press) demonstrates the benefit of these strategies.

8.3.2. Differential Item Functioning (DIF)

A fundamental assumption in the Rasch model is that the probability of a correct response to a test item is a function of the item's difficulty and the student's ability. This function is expected to remain invariant to other person characteristics such as gender and ethnicity. Therefore, if two students with the same ability respond to the same item, they are assumed to have an equal probability of answering the item correctly. To test this assumption, responses to items by students sharing an aspect of a person characteristic (e.g., gender) are compared to responses to the same items by other students who share a different aspect of the same characteristic (e.g., males vs. females). The group representing students in a specific demographic group (usually a minority group) is referred to as the focal group. The group comprised of students from outside this group is referred to as the reference group.

When students with the same ability from two different groups of interest have different probabilities of correctly answering an item, the item is said to exhibit DIF, a statistical characteristic of an item that shows the extent to which the item might be measuring different ability for different student subgroups. DIF indicates a violation of a major assumption of the Rasch model, and it signals potential for a lack of fairness at the item level. The presence of DIF in an item suggests that the item is functioning unexpectedly regarding the groups included in the comparison. The cause of the unexpected functioning is not revealed in a DIF analysis. It may be that item content is inadvertently providing an advantage or disadvantage to members of one of the two groups. Content experts who have special knowledge of the groups involved are often in a good position to identify a cause of this type. DIF may also result from differential instruction closely associated with group membership.

The Mantel-Haenszel (MH) procedure (1959) is the most cited and studied method for detecting DIF. It stratifies examinees by a composite test score, compares the item performance of reference and focal group members in each strata, and then pools this comparison over all strata. The MH procedure is easy to implement and is featured in most statistical software. NWEA applied the MH method to assess DIF of the MAP Growth item pool in this report.

In the previous technical report (NWEA, 2011), NWEA conducted a large-scale DIF analysis that assessed more than 4,000 items from both the Reading and Language Usage item pools and more than 6,000 items from the Mathematics item pool. Results from that report suggested that the percentages of items that exhibit DIF related to gender and ethnicity are very small. In this technical report, instead of assessing the entire item pools, 500 items from each content area's item pool were randomly selected. DIF analysis was conducted for these randomly selected items to examine the percentages of items that exhibit DIF in the item pools and whether DIF results are similar compared to the results reported in the previous technical report.

The results are categorized based on the Educational Testing Service (ETS)'s method of classifying DIF (Zwick, 2012). Table 8.3 presents the criteria for each level of classification. This method allows items exhibiting negligible DIF (Category A) to be differentiated from those exhibiting moderate DIF (Category B) and severe DIF (Category C). Categories B and C have a further breakdown as "+" (DIF is in favor of the focal group) or "-" (DIF is in favor of the reference group).

Table 8.3. DIF Categories

ETS Category	Level of DIF	Definition
A	Negligible	• Absolute value of the Mantel-Haenszel delta difference (MH D-DIF) is not significantly different from 0 or is less than one.
B	Moderate	• Absolute value of the MH D-DIF is significantly different from 0 but not from one, and is at least 1; or • Absolute value of the MH D-DIF is significantly different from 1, but less than 1.5. • Positive values are classified as "B+" and negative values as "B-".
C	Severe	• Absolute value of the MH D-DIF is significantly different from 1, and is at least 1.5; and • Absolute value of the MH D-DIF is larger than 1.96 times the standard error of MH D-DIF. • Positive values are classified as "C+" and negative values are "C-".

Data for the DIF analyses were taken from responses to operational MAP Growth tests from Fall 2016 to Fall 2017 retrieved from the NWEA Growth Research Database (GRD)[7]. Two thousand items were included in the DIF analyses, with 500 items from each content area. Each item had more than 5,000 test records, ensuring an adequate sample size of students for each group involved in the comparison. This, in turn, ensured that each comparison had adequate power to detect DIF. Each test record included the student's recorded ethnic group, gender, and score of the item. All items exhibiting moderate (Category B) DIF are subjected to an extra review by content specialists to identify the source for DIF. For each item, these specialists decide the following:

- Remove the item from the item bank
- Revise the item and re-submit it for field testing
- Retain the item without modification

[7] The GRD was developed and is maintained by the Center for Research on Academic Growth at NWEA in Portland, OR. It currently holds data for more than 170 million test events dating back to Spring 2002. Roughly 99% of all tests results come from adaptive tests consisting of Rasch calibrated items.

Items exhibiting severe DIF (Category C) are removed from the item bank. These procedures are consistent with periodic item quality reviews that remove or flag items for revision and re-field testing problem items.

Table 8.4 presents the number of items and students who answered all 500 items for each content area that were included in this analysis. The table also presents the percentages of students by gender and ethnicity included in the DIF analyses. Data from all states and grades were combined for each content area. This aggregation was made because DIF was focused narrowly on how students of the same ability but of a different gender or ethnic group respond to items. The intent was to neutralize the effects of differential content and instructional emphasis that could potentially influence the DIF analysis. Retaining states and grades as part of the analysis could have led to conclusions that were tangential to the primary focus.

Table 8.4. Number of Students and Items Included in the Fall 2016 to Fall 2017 DIF Analysis

			%Students*						
			Gender		Ethnicity**				
Content Area	#Items	#Students	Female	Male	AI/AN	Asian	Black	Hispanic	White
Reading	500	63,362,963	48.8	51.1	1.7	4.1	17.4	16.8	46.2
Language Usage	500	41,383,859	47.8	52.1	2.5	3.7	13.8	15.8	46.2
Mathematics	500	75,945,605	48.7	51.2	1.6	4.1	17.3	17.6	45.5
Science	500	19,240,698	49.0	50.8	2.7	3.9	19.0	14.5	44.5

*Because gender and ethnicity information of some students was not available, the total % may not add up to 100.0.
**AI/AN = American Indian or Alaskan Native. Besides the ethnicity groups listed in the table, there are three other ethnicity groups with smaller proportions of students: Multiethnic, Native Hawaiian or other Pacific Islander (NH/PI), and Not Specified or Other.

Table 8.5 presents the number of items and percentage of items exhibiting DIF by gender or ethnicity for each MAP Growth content area. As shown in the table, DIF related to gender is rare. The percentage of Category C DIF ranged from 0.4% to 1.4% across content areas. Language Usage had the highest percentage of items showing negligible DIF, or Category A (99.2%), and Mathematics had the lowest percentage of items showing negligible DIF (94.8%). DIF related to ethnicity shares the following three patterns for all content areas:

- Most items are classified in Category A.
- Only 0.2–5.2% of items are classified as Category C.
- The prevalence of B and C classifications are fewer than expected by chance.

Table 8.5. DIF Results for Gender and Ethnicity

Focal Group*	ETS Class***	Reading		Language Usage		Mathematics		Science	
		#Items	%	#Items	%	#Items	%	#Items	%
Female	A	491	98.2	496	99.2	474	94.8	478	95.6
	B+	2	0.4	–	–	4	0.8	8	1.6
	B-	4	0.8	2	0.4	15	3.0	11	2.2
	C+	–	–	–	–	–	–	–	–
	C-	3	0.6	2	0.4	7	1.4	3	0.6

Focal Group*	ETS Class***	Reading		Language Usage		Mathematics		Science	
		#Items	%	#Items	%	#Items	%	#Items	%
AI/AN**	A	468	99.2	471	95.0	444	93.3	438	98.2
	B+	–	–	8	1.6	16	3.4	2	0.4
	B-	2	0.4	12	2.4	11	2.3	5	1.1
	C+	–	–	–	–	–	–	–	–
	C-	2	0.4	5	1.0	5	1.1	1	0.2
Asian	A	444	88.8	431	86.4	445	89.0	463	93.2
	B+	29	5.8	19	3.8	25	5.0	8	1.6
	B-	18	3.6	23	4.6	15	3.0	21	4.2
	C+	7	1.4	3	0.6	5	1.0	1	0.2
	C-	2	0.4	23	4.6	10	2.0	4	0.8
Black	A	489	97.8	473	94.8	414	83.0	476	95.2
	B+	3	0.6	7	1.4	39	7.8	2	0.4
	B-	7	1.4	11	2.2	27	5.4	18	3.6
	C+	–	–	1	0.2	11	2.2	–	–
	C-	1	0.2	7	1.4	8	1.6	4	0.8
Hispanic	A	491	98.2	478	95.6	456	91.2	490	98.0
	B+	1	0.2	2	0.4	23	4.6	2	0.4
	B-	6	1.2	7	1.4	10	2.0	6	1.2
	C+	–	–	–	–	1	0.2	1	0.2
	C-	2	0.4	13	2.6	10	2.0	1	0.2

*For the DIF analysis by gender, the reference group is male. For all other analyses, the reference group is White. The number of items includes items with 500 or more responses from both the focal and the reference groups and 200 or more responses form the focal group.
**AI/AN = American Indian or Alaskan Native.
***B- and C- = DIF is against the focal group. B+ and C+ = DIF is against the reference group.

References

Achieve. (2018, April). *A framework to evaluate cognitive complexity in mathematics assessments*. Retrieved from https://www.achieve.org/files/Cognitive%20Complexity%20Mathematics%20Assessment_FINAL_0.pdf.

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

Andersen, E. B. (2002). Residual diagrams based on a remarkably simple result concerning the variances of maximum likelihood estimators. *Journal of Educational and Behavioral Statistics, 27,* (1), 19–30.

Anderson, L. W., & Krathwohl, D. R. (eds.) (2001). *A taxonomy for learning, teaching, and assessing: A revision of bloom's taxonomy of educational objectives*. New York: Longman.

Andrich, D., Marais, I., & Humphry, S. (2012). Using a theorem by Andersen and the dichotomous Rasch model to assess the presence of random guessing in multiple choice items*. Journal of Educational and Behavioral Statistics, 37*(3), 417–442.

Berliner, D. (1990). What's all the fuss about instructional time? In M. Ben-Peretz & R. Bromme (Eds.), *The nature of time in schools: Theoretical concepts, practitioner perceptions* (pp. 3–35). New York: Teachers College Press. Retrieved from http://courses.ed.asu.edu/berliner/readings/fuss/fuss.htm.

Betebenner, D. W. (2008). Toward a normative understanding of student growth. In K. E. Ryan & L. A. Shepard (Eds.), *The future of test-based educational accountability* (pp. 155–170). New York: Taylor & Francis.

Center for Applied Special Technology (CAST). (2018). *Universal design for learning guidelines version 2.2* (graphic organizer). Wakefield, MA: CAST. Retrieved from http://udlguidelines.cast.org/binaries/content/assets/udlguidelines/udlg-v2-2/udlg_graphicorganizer_v2-2_numbers-yes.pdf.

Council of Chief State School Officers (CCSSO). (2016, August). *CCSSO accessibility manual: How to select, administer, and evaluate use of accessibility supports for instruction and assessment of all students*. Washington, DC: Author.

Crocker, L., & Algina, J. (1986). Introduction to classical and modern test theory. New York : Holt, Rinehart, and Winston.

Egan, K. L., & Davidson, A. H. (2017, Nov. 14). *Alignment of the NWEA MAP Growth & MAP Growth K–2 to the Common Core State Standards: English language arts & mathematics*. EdMetric.

Hauser, C., Kingsbury, G. G., & Wise, S. L. (2008, March). *Individual validity: Adding a missing link.* Paper presented at the annual meeting of the American Educational Research Association (AERA), New York, NY.

Hauser, C., Thum, Y. M., He, W., & Ma, L. (2014). Using a model of analysts' judgments to augment an item calibration process. *Educational and Psychological Measurement, 75*(5), 826–849.

Huynh, H., & Rawls, A. (2009). A comparison between robust z and 0.3-logit difference procedures in assessing stability of linking items for the Rasch model. In Everett V. Smith Jr. & Greg E. Stone (Eds.), *Applications of Rasch Measurement in Criterion-Referenced Testing: Practice Analysis to Score Reporting*. Maple Grove, MN: JAM Press.

Ingebo, G. S. (1997). *Probability in the measure of achievement*. Chicago, IL: MESA Press.

Jiban, C. (2017). *MAP Growth Reading and Language Usage literature review*. Portland, OR: NWEA.

Kingsbury G. G., & Hauser, C. (2004, April). *Computerized adaptive testing and No Child Left Behind.* Paper presented at the annual meeting of the American Educational Research Association (AERA), San Diego, CA.

Kingsbury, G. G., & Weiss, D. J. (1980). *An alternate-forms reliability and concurrent validity comparison of Bayesian adaptive and conventional ability tests* (Research Report 80-5). Minneapolis, MN: University of Minnesota, Department of Psychology, Psychometric Methods Program, Computerized Adaptive Testing Laboratory.

Kingsbury, G. G., & Zara, A. (1989). Procedures for selecting items for computerized adaptive tests. *Applied Measurement in Education, 2*(4), 359–375.

Kingsbury, G. G., & Zara, A. (1991). A comparison of procedures for content-sensitive item selection in computerized adaptive tests. *Applied Measurement in Education, 4*(3), 241–261.

Kolen, M. J., & Brennan, R. L. (2004). *Test equating, scaling, and linking*. New York: Springer.

Kong, X. J., Wise, S. L., Harmes, J. C., & Yang, S. (2006, April). *Motivational effects of praise in response-time based feedback: A follow-up study of the effort-monitoring CBT.* Paper presented at the annual meeting of the National Council on Measurement in Education, San Francisco.

Knekta, E. (2017). Are all pupils equally motivated to do their best on all tests? Differences in reported test-taking motivation within and between tests with different stakes. *Scandinavian Journal of Educational Research, 61*(1), 95–111.

Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores*. Menlo Park, CA: Addison-Wesley.

Lord, F. M. (1980). *Applications of item response theory to practical testing problems.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Mantel, N., & Haenszel, W. (1959). Statistical aspects of the analysis of data from retrospective studies of disease. *Journal of the National Cancer Institute*, 22, 719–748.

Masters, G. N. (1985). Common person equating with the Rasch model. *Applied Psychological Measurement, 9*(1), 73–82.

National Center on Response to Intervention (NCRTI). (2016). *Screening tools chart rating system*. Retrieved from https://rti4success.org/resources/tools-charts/screening-tools-chart/screening-tools-chart-rating-system.

National Governors Association Center for Best Practices & Council of Chief State School Officers (CCSSO). (2010). *Common core state standards*. Washington, DC: Authors.

NGSS Lead States. (2013). *Next Generation Science Standards: For states, by states*. Washington, DC: The National Academic Press.

NWEA. (2011, January). *Technical manual for Measures of Academic Progress® (MAP®) and Measures of Academic Progress for Primary Grades (MPG)*. Portland, OR: NWEA.

Owen, R. J. (1975). A Bayesian sequential procedure for quantal response in the context of adaptive testing. *Journal of the American Statistical Association, 70,* 229–244.

Pommerich, M., Hanson, B., Harris, D., & Sconing, J. (2004). Issues in conducting linkage between distinct tests. *Applied Psychological Measurement, 28*(4), 247–273.

Rasch, G. (1960/1980). *Probabilistic models for some intelligence and attainment tests*. Chicago, IL: MESA Press.

Samejima, F. (1977). A use of the information function in tailored testing. *Applied Psychological Measurement, 1*(3), 233–247.

Samejima, F. (1994). Estimation of reliability coefficients using the test information function and its modifications. *Applied Psychological Measurement, 18*(3), 229–244.

Thompson, S. J., Johnstone, C. J., & Thurlow, M. L. (2002). *Universal design applied to large scale assessments* (Synthesis Report 44). Minneapolis: University of Minnesota, National Center on Educational Outcomes (NCEO). http://education.umn.edu/NCEO/OnlinePubs/Synthesis44.html

Thum, Y. M., & Hauser, C. H. (2015). *NWEA 2015 MAP norms for student and school achievement status and growth*. Portland, OR: NWEA.

Wang, S., Jiao, H., & Zhang, Z. (2013). Validation of longitudinal achievement constructs of vertically scaled computerized adaptive tests: A multiple-indicator, latent-growth modelling approach. *International Journal of Quantitative Research in Education, 1*(4), 383–407.

Wang, S., McCall, M., Jiao, H., & Harris, G. (2013). Construct validity and measurement invariance of computerized adaptive testing: Application to Measures of Academic Progress (MAP) using confirmatory factor analysis. *Journal of Educational and Developmental Psychology, 3*(1), 88–100.

Wang, C., & Xu, G. (2015). A mixture hierarchical model for response times and response accuracy. *British Journal of Mathematical and Statistical Psychology*, *68*, 456-477.

Webb, N. (1997). *Alignment of science and mathematics standards and assessments in four states*. Research Monograph Number 6: Washington, D.C.: CCSSO.

Weiss, D. J., & Vale, C. D. (1987). Adaptive testing. *Applied Psychology, 36*(3–4), 249–262.

Weiss, D. J. (1974). *Strategies of adaptive ability measurement* (Research Report 74-5). Minneapolis, MN: University of Minnesota, Department of Psychology, Psychometric Methods Program, Computerized Adaptive Testing Laboratory.

Wilson, E. B., & Hilferty, M. M. (1931). The distribution of chi-square. *Proceedings of the National Academy of Sciences of the United States of America, 17*, 684–688.

Wise, S. L., Bhola, D., & Yang, S. (2006). Taking the time to improve the validity of low-stakes tests: The effort-monitoring CBT. *Educational Measurement: Issues and Practice 25(2),* 21–30.

Wise, S. L., & DeMars, C. E. (2005). Low examinee effort in low-stakes assessment: Problems and potential solutions. *Educational Assessment, 10,* 1–17.

Wise, S. L., & DeMars, C. E. (2006). An application of item response time: The effort-moderated IRT model. *Journal of Educational Measurement, 43,* 19-38.

Wise, S. L., & Kingsbury, G. G. (2016). Modeling student test-taking motivation in the context of an adaptive achievement test. *Journal of Educational Measurement, 53,* 86–105.

Wise. S. L., & Kong, X. (2005). Response time effort: A new measure of examinee motivation in computer-based tests. *Applied Measurement in Education, 18*, 163–183.

Wise, S. L., Kuhfeld, M. R., & Soland, J. (in press). The effects of effort monitoring with proctor notification on test-taking engagement, test performance, and validity. *Applied Measurement in Education*.

Wolf, L. F., & Smith, J. K. (1995). The consequence of consequence: Motivation, anxiety, and test performance. *Applied Measurement in Education, 8*, 227–242.

Wolf, L. F., Smith, J. K., & Birnbaum, M. E. (1995). Consequence of performance, test motivation, and mentally taxing items. *Applied Measurement in Education, 8*, 341–351.

Wright, B. D. (1999). Rasch measurement models. In G. N. Masters & J. P. Keeves (Eds.), *Advances in measurement in educational research and assessment* (pp. 85–97). Oxford, UK: Elsevier Science Ltd.

Zwick, R. (2012). *A review of ETS differential item functioning assessment procedures: Flagging rules, minimum sample size requirements, and criterion refinement* (ETS RR-12-08). Princeton, NJ: ETS.

Appendix A: Student Sample by State and Demographics

Table A.1. Number of Test Events and Students by State

State	Reading			Language Usage			Mathematics			Science		
	#Test Events	Students N	Students %*	#Test Events	Students N	Students %*	#Test Events	Students N	Students %*	#Test Events	Students N	Students %*
AK	51,421	26,163	0.6	1,639	582	0.0	51,386	25,933	0.5	–	–	–
AL	6,334	3,171	0.1	4,646	2,359	0.2	6,385	3,149	0.1	–	–	–
AR	–	–	–	–	–	–	–	–	–	45,034	20,398	4.1
AZ	27,535	14,665	0.3	12,345	5,343	0.4	27,465	14,550	0.3	234	234	0.0
CA	638,281	220,835	4.7	216,675	85,896	6.7	650,604	227,426	4.7	62,513	35,506	7.1
CO	31,200	12,297	0.3	2,671	1,096	0.1	33,421	13,328	0.3	36,749	14,921	3.0
CT	329,546	123,816	2.6	73,719	29,010	2.2	360,844	132,550	2.8	19,086	10,137	2.0
DC	69,617	26,419	0.6	1,412	891	0.1	89,528	35,384	0.7	1,372	690	0.1
DE	53,312	20,082	0.4	1,786	779	0.1	55,039	19,931	0.4	1,354	858	0.2
FL	147,409	54,450	1.2	3,829	2,177	0.2	146,590	54,245	1.1	336	310	0.1
GA	3,876	1,518	0.0	1,953	822	0.1	8,353	3,321	0.1	43,593	43,515	8.7
HI	20,329	7,734	0.2	3,387	1,610	0.1	21,034	7,995	0.2	438	296	0.1
IA	–	–	–	–	–	–	–	–	–	47,217	38,768	7.7
ID	57,322	23,134	0.5	36,848	14,781	1.1	62,264	24,933	0.5	1,121	999	0.2
IL	2,822,342	997,935	21.1	362,527	144,213	11.2	2,854,548	1,006,407	20.9	115,402	63,988	12.8
IN	4,816	2,077	0.0	1,471	706	0.1	6,291	3,092	0.1	617	305	0.1
KS	735	334	0.0	351	148	0.0	686	335	0.0	22,705	13,926	2.8
KY	1,175,197	414,495	8.8	348,899	144,314	11.2	1,178,857	413,151	8.6	31,761	18,579	3.7
LA	160,951	62,132	1.3	64,851	25,567	2.0	159,766	61,881	1.3	192	111	0.0
MA	6,965	6,912	0.1	124	91	0.0	8,444	7,788	0.2	5,437	3,583	0.7
MD	6,594	3,783	0.1	3,289	1,564	0.1	7,231	3,993	0.1	3,085	1,958	0.4
ME	232,463	90,235	1.9	53,703	24,654	1.9	235,286	90,470	1.9	424	424	0.1
MI	2,544,570	870,566	18.4	907,606	355,580	27.6	2,551,864	866,713	18	371,595	178,984	35.7
MN	850	718	0.0	487	378	0.0	1,447	1,119	0.0	455	313	0.1
MO	143,505	57,295	1.2	47,673	20,161	1.6	144,391	57,999	1.2	5,656	2,900	0.6
MS	235,431	92,116	1.9	93,406	41,760	3.2	234,739	92,144	1.9	–	–	–
MT	181,739	64,526	1.4	105,100	41,086	3.2	182,937	64,165	1.3	5,369	4,152	0.8
NC	524,790	177,097	3.7	25,254	11,511	0.9	564,309	190,358	4.0	663	388	0.1
ND	–	–	–	–	–	–	–	–	–	657	398	0.1
NE	19,747	7,554	0.2	–	–	–	19,310	7,537	0.2	–	–	–
NH	138,381	57,894	1.2	20,672	11,213	0.9	143,572	58,587	1.2	1,047	1,047	0.2
NJ	288,833	127,998	2.7	70,509	34,172	2.6	340,498	150,255	3.1	9,369	5,370	1.1
NM	158,036	67,000	1.4	66,615	32,040	2.5	159,968	67,723	1.4	–	–	–
NV	403,289	198,018	4.2	41,753	19,502	1.5	394,379	185,841	3.9	9,453	7,850	1.6
NY	10,202	4,101	0.1	309	238	0.0	13,513	5,422	0.1	2,624	2,390	0.5
OH	–	–	–	–	–	–	–	–	–	5,867	3,986	0.8
OK	5,167	3,668	0.1	852	786	0.1	6,915	4,286	0.1	1,919	850	0.2
OR	83,789	32,591	0.7	23,212	10,717	0.8	88,828	34,774	0.7	2,669	1,751	0.3
PA	17,023	6,841	0.1	7,805	2,971	0.2	17,248	6,986	0.1	368	342	0.1
RI	25,422	9,798	0.2	4,498	2,244	0.2	25,665	9,893	0.2	2,865	1,281	0.3
SC	536	271	0.0	393	213	0.0	421	211	0.0	–	–	–

State	Reading			Language Usage			Mathematics			Science		
	#Test Events	Students N	Students %*	#Test Events	Students N	Students %*	#Test Events	Students N	Students %*	#Test Events	Students N	Students %*
SD	168,882	67,090	1.4	77,276	32,950	2.6	171,975	67,124	1.4	4,168	2,196	0.4
TN	368,456	144,046	3.0	73,112	36,290	2.8	369,353	142,980	3.0	136	136	0.0
TX	11,063	5,367	0.1	2,726	1,319	0.1	11,286	5,522	0.1	725	640	0.1
UT	44,550	16,853	0.4	30,802	11,677	0.9	44,654	17,000	0.4	–	–	–
VA	2,104	1,430	0.0	1,837	1,275	0.1	2,205	1,509	0.0	755	538	0.1
VT	29,085	11,552	0.2	14,661	5,622	0.4	31,262	12,235	0.3	37	37	0.0
WA	552,106	217,019	4.6	68,476	29,790	2.3	557,851	220,718	4.6	23,053	13,902	2.8
WI	874,360	300,275	6.3	172,284	69,310	5.4	892,911	305,803	6.4	6,203	2,668	0.5
WV	1,684	1,389	0.0	579	579	0.0	1,660	1,370	0.0	–	–	–
WY	202,621	77,836	1.6	66,311	30,584	2.4	204,149	78,711	1.6	129	67	0.0
Total	**12,882,466**	**4,733,096**	**100.0**	**3,120,333**	**1,290,571**	**100.0**	**13,141,332**	**4,806,847**	**100.0**	**894,452**	**501,692**	**100.0**

*Percentages are out of the total number of students across all states.

Table A.2. Number of Students by State, Gender, and Ethnicity—Reading

State	N-Count	Gender %*			Race and Ethnicity %**								
		Female	Male	N/A	AI/AN	Asian	Black	Hispanic	Multiethnic	NH/PI	NS/Other	White	N/A
AK	26,163	49.1	50.9	0.0	9.5	16.8	5.5	11.1	15.7	0.0	0.9	40.6	–
AL	3,171	47.5	52.2	0.3	0.2	0.7	5.4	4.7	0.2	0.4	11.3	77.1	0.1
AZ	14,665	48.7	51.2	0.1	53.5	0.1	0.3	33.9	0.5	0.0	2.5	9.2	–
CA	220,835	48.9	50.8	0.3	0.9	8.6	8.0	47.3	2.3	0.4	10.8	21.7	0.0
CO	12,297	47.7	52.1	0.2	1.9	1.3	1.6	43.6	2.7	0.1	5.9	42.9	–
CT	123,816	48.7	51.1	0.2	3.2	4.3	13.3	24.3	2.2	0.4	9.1	43.2	0.0
DC	26,419	50.5	48.5	1.1	0.2	0.6	60.0	7.4	0.9	0.0	27.9	2.9	0.0
DE	20,082	48.7	51.0	0.2	0.8	4.7	34.1	3.8	1.9	0.2	5.1	49.6	–
FL	54,450	49.8	50.0	0.2	0.4	3.1	24.8	36.6	3.9	0.0	9.4	21.8	0.0
GA	1,518	46.2	51.5	2.3	0.1	0.6	61.7	1.2	1.1	–	30.6	4.7	–
HI	7,734	50.1	49.8	0.0	0.7	1.9	0.3	0.2	0.6	6.1	84.0	6.3	–
ID	23,134	48.2	51.6	0.2	1.6	0.9	0.7	14.3	1.9	0.2	15.5	65.0	–
IL	997,935	48.9	51.0	0.1	1.0	4.6	18.7	22.9	3.6	0.3	10.5	38.5	0.0
IN	2,077	46.4	52.2	1.3	0.1	1.3	33.8	11.5	2.8	0.1	13.9	36.4	–
KS	334	48.2	51.8	–	–	–	2.1	2.1	4.5	–	0.3	91.0	–
KY	414,495	48.7	51.3	0.1	0.2	1.3	7.4	5.3	2.9	0.1	22.7	60.1	0.0
LA	62,132	48.2	51.2	0.6	0.3	1.7	54.2	5.6	0.3	0.0	9.6	28.3	0.0
MA	6,912	49.2	50.6	0.2	–	0.5	0.1	10.2	0.1	–	88.1	0.9	–
MD	3,783	48.4	49.6	2.0	0.1	1.0	67.7	4.3	1.6	0.0	4.8	20.4	–
ME	90,235	48.7	51.3	0.1	0.9	1.1	4.3	1.6	1.5	0.1	17.5	73.1	0.0
MI	870,566	48.6	51.2	0.2	1.0	3.6	24.8	6.8	2.0	0.1	5.9	55.9	0.0
MN	718	51.4	48.6	–	–	–	19.1	–	–	–	80.9	–	–
MO	57,295	48.3	51.3	0.3	0.6	1.7	23.6	11.7	3.5	0.3	4.2	54.4	0.0
MS	92,116	48.7	50.9	0.4	0.1	4.5	40.7	3.5	0.3	0.1	4.2	46.6	0.1
MT	64,526	48.8	51.1	0.1	11.0	0.6	0.9	4.2	3.3	0.5	13.2	66.2	–
NC	177,097	48.8	51.0	0.2	1.1	5.5	31.2	17.9	2.6	0.2	10.8	30.8	0.0
NE	7,554	48.1	51.9	0.0	1.1	1.6	5.2	49.6	0.0	0.0	0.7	41.7	0.0
NH	57,894	48.6	51.3	0.1	0.3	1.7	1.2	2.3	1.0	0.2	21.4	72.0	0.0

State	N-Count	Gender %*			Race and Ethnicity %**								
		Female	Male	N/A	AI/AN	Asian	Black	Hispanic	Multiethnic	NH/PI	NS/Other	White	N/A
NJ	127,998	48.3	51.5	0.2	0.2	7.7	17.1	16.8	2.3	0.2	9.0	46.7	0.0
NM	67,000	49.3	50.6	0.1	22.1	1.0	1.6	43.6	0.1	0.2	14.6	16.8	0.0
NV	198,018	48.8	51.2	0.0	1.4	3.7	8.1	34.1	5.5	1.2	22.6	23.7	0.0
NY	4,101	49.1	50.8	0.1	0.2	1.2	43.8	38.7	1.8	0.1	6.5	8.0	0.0
OK	3,668	47.2	52.5	0.3	11.8	1.6	7.4	25.5	1.4	0.2	26.6	25.6	–
OR	32,591	47.8	52.0	0.2	0.7	2.7	1.5	13.4	4.7	0.4	13.4	63.2	–
PA	6,841	46.1	53.1	0.7	0.1	3.2	32.7	14.9	2.9	0.0	8.0	38.2	–
RI	9,798	49.8	50.0	0.2	1.0	1.3	5.2	11.6	2.8	0.1	44.9	33.1	–
SC	271	53.9	46.1	–	–	–	4.8	4.1	–	1.5	0.4	89.3	–
SD	67,090	48.7	51.0	0.3	23.9	2.2	3.4	6.2	3.7	0.1	0.8	59.7	–
TN	144,046	48.1	49.4	2.5	0.1	1.5	61.4	12.0	2.2	0.1	1.6	18.8	2.4
TX	5,367	47.8	51.8	0.4	0.3	2.6	5.0	60.3	1.8	0.1	11.6	18.4	0.0
UT	16,853	47.9	51.7	0.4	2.9	1.7	0.9	11.4	1.9	0.5	6.3	74.3	–
VA	1,430	47.6	52.3	0.1	0.4	3.6	23.9	4.3	1.2	0.1	44.7	21.8	–
VT	11,552	48.1	51.9	0.0	0.1	0.8	0.9	0.8	1.6	0.1	14.0	81.7	–
WA	217,019	48.7	51.2	0.1	2.7	3.9	4.2	19.0	5.3	0.8	14.2	49.9	0.0
WI	300,275	48.9	51.0	0.1	1.6	3.3	9.9	11.2	2.9	0.1	6.5	64.4	0.0
WV	1,389	46.3	53.7	–	–	–	–	–	–	–	–	100.0	–
WY	77,836	48.4	51.5	0.1	4.5	1.0	1.3	13.2	1.1	0.1	1.8	77.2	0.0
Total	**4,733,096**	**48.7**	**51.0**	**0.2**	**2.0**	**3.7**	**17.6**	**16.4**	**2.9**	**0.3**	**11.0**	**46.1**	**0.1**

*N/A = Gender information is not available.
**AI/AN = American Indian or Alaskan Native. NH/PI = Native Hawaiian or Other Pacific Islander. NS/Other = Not Specified or Other. N/A = Race and ethnicity information is not available.

Table A.3. Number of Students by State, Gender, and Ethnicity—Language Usage

State	N-Count	Gender %*			Race and Ethnicity %**								
		Female	Male	N/A	AI/AN	Asian	Black	Hispanic	Multiethnic	NH/PI	NS/Other	White	N/A
AK	582	60.7	39.3	–	33.9	1.4	0.2	–	33.7	0.2	28.4	2.4	–
AL	2,359	46.6	53.0	0.4	0.1	0.7	4.4	4.9	–	0.5	12.9	76.4	0.1
AZ	5,343	50.2	49.5	0.3	89.8	0.2	0.2	1.0	0.1	–	3.7	5.1	–
CA	85,896	48.6	51.2	0.2	0.9	10.1	4.5	48.8	3.3	0.3	6.5	25.5	0.0
CO	1,096	45.5	54.5	–	0.9	1.6	0.4	24.0	0.1	–	43.8	29.2	–
CT	29,010	48.9	51.0	0.1	3.1	3.9	12.7	29.3	1.5	0.1	9.7	39.8	–
DC	891	58.5	41.0	0.6	0.2	2.7	71.2	6.0	1.4	0.1	6.6	11.9	–
DE	779	48.4	51.6	–	0.1	2.2	32.1	30.7	0.8	0.1	0.1	33.9	–
FL	2,177	49.6	50.4	–	0.1	1.1	13.0	6.3	2.0	–	61.8	15.7	–
GA	822	46.8	52.1	1.1	–	0.2	57.7	0.5	0.1	–	39.1	2.4	–
HI	1,610	50.4	49.6	–	0.4	0.9	0.2	0.4	0.5	7.8	87.4	2.4	–
ID	14,781	48.3	51.4	0.3	1.7	1.2	0.8	12.2	1.4	0.2	19.6	62.8	–
IL	144,213	48.4	51.5	0.1	0.7	4.2	9.4	13.5	4.8	0.1	15.4	52.0	0.0
IN	706	44.5	52.0	3.5	0.3	0.1	31.3	10.2	3.8	–	17.7	36.5	–
KS	148	49.3	50.7	–	–	–	4.1	3.4	–	–	0.7	91.9	–
KY	144,314	48.7	51.3	0.1	0.2	0.9	5.2	4.6	2.7	0.1	15.4	71.1	0.0
LA	25,567	49.4	50.6	0.0	0.6	2.1	41.6	6.2	0.1	0.0	4.8	44.5	0.0
MA	91	84.6	15.4	–	–	1.1	4.4	16.5	9.9	–	17.6	50.6	–

State	N-Count	Gender %*			Race and Ethnicity %**								
		Female	Male	N/A	AI/AN	Asian	Black	Hispanic	Multiethnic	NH/PI	NS/Other	White	N/A
MD	1,564	52.0	47.9	0.1	0.1	2.1	34.5	6.1	3.4	–	10.3	43.6	–
ME	24,654	47.7	52.2	0.1	1.1	0.7	1.5	1.1	1.0	0.1	15.1	79.4	–
MI	355,580	48.7	51.1	0.2	1.1	3.0	23.5	5.4	1.9	0.1	5.7	59.3	0.0
MN	378	51.1	48.9	–	–	–	30.7	–	–	–	69.3	–	–
MO	20,161	48.0	51.7	0.2	0.9	1.4	17.7	11.3	3.1	0.4	2.2	63.0	
MS	41,760	49.2	50.6	0.2	0.1	5.5	45.6	2.7	0.3	0.0	6.6	39.1	0.1
MT	41,086	49.0	50.9	0.1	11.3	0.5	0.9	4.6	3.0	0.3	11.9	67.4	–
NC	11,511	48.9	51.0	0.1	0.8	2.0	25.2	6.9	3.0	0.5	21.7	40.0	–
NH	11,213	47.5	52.3	0.2	0.3	1.8	1.5	3.6	1.2	0.1	17.5	74.0	–
NJ	34,172	47.9	51.9	0.2	0.1	5.7	16.6	18.3	2.5	0.2	9.2	47.5	–
NM	32,040	49.4	50.5	0.1	25.2	0.8	0.9	42.3	0.1	0.1	15.2	15.5	0.0
NV	19,502	48.9	50.9	0.2	4.5	3.6	5.1	26.9	3.9	0.7	5.1	50.3	–
NY	238	42.4	57.1	0.4	–	0.4	1.7	–	0.4	–	74.8	22.7	–
OK	786	45.7	54.3	–	30.2	5.2	0.9	–	0.1	0.5	0.4	62.7	–
OR	10,717	48.0	51.9	0.1	1.0	3.1	1.8	9.5	4.2	0.5	20.7	59.4	–
PA	2,971	46.1	53.5	0.4	0.0	5.5	26.7	5.1	4.7	–	2.4	55.7	–
RI	2,244	51.8	47.7	0.5	0.2	0.5	4.3	9.3	0.9	–	79.6	5.3	–
SC	213	57.3	42.7	–	–	–	3.8	3.8	–	1.9	–	90.6	–
SD	32,950	48.4	51.3	0.4	21.7	2.5	3.8	6.6	3.3	0.1	0.8	61.3	–
TN	36,290	48.1	48.8	3.1	0.1	1.2	58.0	11.4	1.7	0.0	1.0	23.6	3.0
TX	1,319	47.2	52.5	0.4	0.4	9.0	3.8	7.1	6.0	0.4	30.7	42.8	–
UT	11,677	48.0	51.7	0.3	2.4	1.3	0.8	12.2	2.1	0.5	7.4	73.4	–
VA	1,275	45.8	54.2	–	0.5	2.7	23.0	4.9	0.9	0.2	45.1	22.8	–
VT	5,622	48.4	51.6	0.0	0.1	1.0	1.3	0.7	2.2	0.1	8.5	86.1	–
WA	29,790	49.1	50.9	0.0	3.3	5.8	3.3	9.4	5.7	0.9	15.7	55.9	–
WI	69,310	49.2	50.7	0.1	3.5	1.9	5.9	6.2	1.4	0.2	10.8	70.1	0.0
WV	579	46.6	53.4	–	–	–	–	–	–	–	–	100.0	–
WY	30,584	48.2	51.7	0.1	5.6	0.9	1.5	12.0	1.2	0.1	2.7	76.1	0.0
Total	**1,290,571**	**48.7**	**51.1**	**0.2**	**3.1**	**3.1**	**14.6**	**11.8**	**2.4**	**0.2**	**9.9**	**54.9**	**0.1**

*N/A = Gender information is not available.
**AI/AN = American Indian or Alaskan Native. NH/PI = Native Hawaiian or Other Pacific Islander. NS/Other = Not Specified or Other. N/A = Race and ethnicity information is not available.

Table A.4. Number of Students by State, Gender, and Ethnicity—Mathematics

State	N-Count	Gender %*			Race and Ethnicity %**								
		Female	Male	N/A	AI/AN	Asian	Black	Hispanic	Multiethnic	NH/PI	NS/Other	White	N/A
AK	25,933	49.1	50.9	0.0	9.2	16.6	5.5	11.1	16.1	0.0	0.7	40.8	–
AL	3,149	47.5	52.2	0.3	0.2	0.7	5.4	4.5	0.2	0.4	11.5	77.1	0.1
AZ	14,550	48.6	51.2	0.1	53.9	0.1	0.2	34.4	0.5	0.0	1.8	9.2	–
CA	227,426	48.9	50.8	0.3	0.9	8.9	8.0	46.6	2.5	0.4	10.9	21.9	0.0
CO	13,328	50.0	49.8	0.2	1.8	1.3	2.4	42.8	2.7	0.1	7.9	41.0	–
CT	132,550	48.8	51.0	0.2	3.0	4.2	14.8	24.4	2.1	0.4	8.5	42.6	0.0
DC	35,384	50.1	49.1	0.8	0.2	1.0	62.3	10.1	1.1	0.0	21.3	4.1	0.0
DE	19,931	48.8	50.9	0.2	0.8	4.7	34.5	3.2	1.9	0.2	5.0	49.7	–
FL	54,245	49.8	50.0	0.2	0.5	3.1	24.8	36.5	3.9	0.0	9.2	21.9	0.0

State	N-Count	Gender %*			Race and Ethnicity %**								
		Female	Male	N/A	AI/AN	Asian	Black	Hispanic	Multiethnic	NH/PI	NS/Other	White	N/A
GA	3,321	61.6	35.1	3.3	0.2	0.5	52.0	0.6	0.6	–	41.7	4.5	–
HI	7,995	50.0	50.0	0.0	1.0	1.8	0.3	0.2	1.5	6.0	82.6	6.7	–
ID	24,933	48.2	51.5	0.3	1.5	1.0	0.7	13.7	1.8	0.2	15.1	66.0	0.0
IL	1,006,407	48.9	51.0	0.1	1.0	4.6	19.0	23.0	3.6	0.2	10.3	38.2	0.0
IN	3,092	48.4	50.7	0.9	0.4	3.0	24.4	18.6	3.5	0.4	11.5	38.2	–
KS	335	48.4	51.6	–	–	–	2.1	2.1	4.5	–	0.3	91.0	–
KY	413,151	48.6	51.3	0.1	0.2	1.3	7.4	5.5	3.0	0.1	22.5	60.1	0.0
LA	61,881	48.2	51.2	0.6	0.3	1.7	54.2	5.6	0.3	0.0	9.5	28.4	0.0
MA	7,788	50.1	49.7	0.2	0.1	0.7	5.2	10.4	0.4	0.1	81.5	1.6	–
MD	3,993	48.2	49.9	1.9	0.1	0.9	61.8	3.2	1.6	0.0	12.3	20.1	–
ME	90,470	48.6	51.3	0.1	0.9	1.2	4.6	1.7	1.5	0.1	17.0	73.2	0.0
MI	866,713	48.6	51.2	0.2	1.0	3.6	24.9	6.8	2.0	0.1	5.9	55.8	0.0
MN	1,119	47.2	52.7	0.1	0.1	0.5	21.6	3.8	1.0	–	59.4	13.6	–
MO	57,999	48.4	51.3	0.3	0.6	2.1	23.1	11.4	3.7	0.2	4.2	54.7	0.0
MS	92,144	48.7	50.9	0.4	0.1	4.3	41.7	3.6	0.3	0.1	4.0	45.8	0.1
MT	64,165	48.8	51.1	0.1	11.2	0.6	0.9	4.2	3.4	0.4	13.2	66.1	–
NC	190,358	48.8	51.0	0.2	1.0	5.7	30.7	18.1	2.7	0.2	9.7	31.9	0.0
NE	7,537	48.1	51.9	0.0	1.1	1.6	5.2	49.6	0.0	0.0	0.7	41.8	0.0
NH	58,587	48.6	51.3	0.1	0.3	1.7	1.2	2.3	1.0	0.2	21.1	72.3	0.0
NJ	150,255	48.7	51.1	0.2	0.2	9.2	17.2	20.4	2.2	0.2	8.4	42.4	0.0
NM	67,723	49.5	50.4	0.1	22.0	1.1	1.6	41.1	0.1	0.2	17.1	16.9	0.0
NV	185,841	48.7	51.3	0.1	1.4	3.6	7.9	34.2	5.4	1.2	23.5	23.0	–
NY	5,422	48.9	51.0	0.1	0.2	1.1	42.1	39.3	1.3	0.1	9.7	6.1	0.0
OK	4,286	46.7	52.1	1.1	11.0	1.6	12.1	25.7	2.8	0.4	22.2	24.2	–
OR	34,774	47.8	52.0	0.2	1.4	2.7	1.5	14.4	4.7	0.4	12.8	62.2	–
PA	6,986	46.7	52.6	0.7	0.1	3.1	31.5	17.4	2.8	0.0	7.9	37.3	0.0
RI	9,893	49.9	49.9	0.2	1.0	1.4	6.2	14.3	2.8	0.1	40.8	33.4	–
SC	211	55.0	45.0	–	–	–	4.7	3.8	–	1.0	0.5	90.1	–
SD	67,124	48.7	51.0	0.3	24.0	2.2	3.4	6.2	3.7	0.1	0.8	59.6	–
TN	142,980	48.1	49.5	2.4	0.1	1.5	61.5	12.0	2.2	0.1	1.5	18.7	2.3
TX	5,522	47.9	51.7	0.4	0.3	2.5	5.3	59.2	1.8	0.1	12.2	18.6	0.0
UT	17,000	48.1	51.7	0.3	3.0	1.8	0.9	11.4	1.9	0.5	5.6	75.0	–
VA	1,509	47.3	52.6	0.1	0.3	3.1	21.7	3.6	1.1	0.1	47.8	22.3	–
VT	12,235	47.9	52.0	0.0	0.1	0.8	1.1	0.8	1.5	0.1	12.8	83.0	–
WA	220,718	48.8	51.1	0.1	2.7	4.2	4.4	19.1	5.3	0.8	13.8	49.7	0.0
WI	305,803	48.9	51.1	0.1	1.6	3.4	9.8	11.1	2.9	0.1	6.6	64.4	0.0
WV	1,370	46.0	54.0	–	–	–	–	–	–	–	–	100.0	–
WY	78,711	48.5	51.4	0.1	4.6	1.0	1.2	13.1	1.1	0.1	1.8	77.1	0.0
Total	**4,806,847**	**48.7**	**51.0**	**0.2**	**2.0**	**3.8**	**17.8**	**16.6**	**2.9**	**0.3**	**10.9**	**45.7**	**0.1**

*N/A = Gender information is not available.
**AI/AN = American Indian or Alaskan Native. NH/PI = Native Hawaiian or Other Pacific Islander. NS/Other = Not Specified or Other. N/A = Race and ethnicity information is not available.

Table A.5. Number of Students by State, Gender, and Ethnicity—Science

State	N-Count	Gender %*			Race and Ethnicity %**								
		Female	Male	N/A	AI/AN	Asian	Black	Hispanic	Multiethnic	NH/PI	NS/Other	White	N/A
AR	20,398	49.0	50.6	0.4	5.2	2.0	15.3	1.5	0.6	0.2	2.3	72.8	0.0
AZ	234	51.7	48.3	–	0.4	1.3	–	7.7	–	–	78.6	12.0	–
CA	35,506	48.6	51.3	0.1	2.5	12.3	6.7	49.4	1.8	0.6	10.6	16.2	–
CO	14,921	48.3	51.5	0.2	0.3	1.6	5.5	24.2	2.3	0.1	45.0	21.2	–
CT	10,137	50.2	49.7	0.1	0.3	3.5	30.3	18.3	0.8	0.1	6.2	40.7	–
DC	690	52.5	47.2	0.3	–	0.6	17.1	29.3	0.3	–	52.3	0.4	–
DE	858	53.0	47.0	–	0.1	12.0	29.3	–	–	0.5	–	58.2	–
FL	310	59.0	41.0	–	0.3	1.3	1.0	0.3	0.3	–	75.2	21.6	–
GA	43,515	48.7	51.3	0.0	0.3	6.3	61.1	18.3	1.9	–	0.0	12.1	–
HI	296	51.4	48.6	–	0.7	7.8	1.7	–	–	27.4	38.9	23.7	–
IA	38,768	49.1	50.9	0.0	0.4	1.1	2.7	5.1	1.3	0.2	8.2	81.0	–
ID	999	42.8	57.1	0.1	–	3.0	1.1	7.3	3.5	0.1	0.4	84.6	–
IL	63,988	49.7	50.2	0.1	0.3	3.6	30.3	21.2	4.9	0.1	9.9	29.7	0.0
IN	305	44.3	55.7	–	–	1.0	2.6	15.7	2.3	–	1.0	77.4	–
KS	13,926	48.5	51.5	0.0	4.5	1.5	2.7	6.3	2.5	0.2	2.3	80.1	0.0
KY	18,579	48.5	51.4	0.1	0.7	1.0	2.9	2.3	2.6	0.2	17.1	73.3	0.0
LA	111	46.8	53.2	–	–	–	98.2	–	–	–	0.9	0.9	–
MA	3,583	50.4	49.5	0.1	–	0.3	1.1	14.9	0.5	–	77.7	5.5	–
MD	1,958	39.5	59.9	0.6	0.3	2.6	35.0	17.7	6.7	0.3	9.7	27.8	–
ME	424	51.2	48.8	–	–	0.2	1.9	4.5	1.7	0.2	3.1	88.4	–
MI	178,984	48.9	50.8	0.3	1.6	3.1	21.5	5.5	1.9	0.1	7.0	59.3	0.0
MN	313	53.4	46.6	–	–	1.9	2.2	1.0	3.5	0.3	4.8	86.3	–
MO	2,900	50.1	49.9	–	0.5	3.0	20.4	8.2	4.9	0.3	0.1	62.6	–
MT	4,152	49.1	50.8	0.0	16.0	0.6	0.8	3.5	1.5	0.3	11.5	65.9	–
NC	388	41.8	58.2	–	–	2.8	31.7	12.4	7.7	0.8	2.6	42.0	–
ND	398	46.5	53.5	–	1.5	0.8	2.8	1.3	0.8	–	1.8	91.2	–
NH	1,047	49.6	50.2	0.2	0.5	2.3	1.3	3.2	2.1	0.1	1.1	89.5	–
NJ	5,370	49.4	50.3	0.3	0.1	3.5	38.3	19.7	0.2	0.0	15.6	22.7	–
NV	7,850	47.9	51.8	0.3	2.9	5.7	4.5	23.3	5.4	0.8	3.0	54.4	–
NY	2,390	56.1	43.8	0.0	0.2	5.4	20.3	24.6	0.1	0.1	0.1	49.3	–
OH	3,986	48.7	51.3	–	0.1	2.0	3.7	2.6	3.0	0.1	24.0	64.4	–
OK	850	48.0	52.0	–	1.3	0.2	0.5	0.5	0.5	–	87.1	10.0	–
OR	1,751	51.6	48.3	0.1	1.4	2.9	3.0	16.1	3.8	0.3	11.4	61.1	–
PA	342	51.2	48.8	–	–	4.4	7.3	–	0.6	1.2	–	86.6	–
RI	1,281	49.3	50.7	–	–	–	–	0.2	0.1	–	99.1	0.6	–
SD	2,196	50.4	49.4	0.3	24.5	0.3	0.5	5.3	5.2	–	0.3	63.9	–
TN	136	36.8	59.6	3.7	0.7	8.1	13.2	5.9	1.5	0.7	10.3	59.6	–
TX	640	44.4	55.6	–	–	4.5	3.1	8.9	0.6	–	77.3	5.5	–
VA	538	52.2	47.8	–	–	3.2	2.0	–	0.4	–	89.4	5.0	–
VT	37	45.9	54.1	–	–	–	–	–	–	–	–	100.0	–
WA	13,902	50.2	49.8	0.1	6.4	2.8	1.5	18.2	3.5	1.0	17.3	49.2	–
WI	2,668	49.6	50.4	0.0	0.8	1.7	1.5	8.8	0.4	0.0	16.5	70.2	0.0
WY	67	61.2	38.8	–	–	–	–	1.5	–	–	–	98.5	–
Total	**501,692**	**49.0**	**50.8**	**0.2**	**1.7**	**3.7**	**20.2**	**13.2**	**2.3**	**0.2**	**9.9**	**48.8**	**0.0**

*N/A = Gender information is not available.
**AI/AN = American Indian or Alaskan Native. NH/PI = Native Hawaiian or Other Pacific Islander. NS/Other = Not Specified or Other. N/A = Race and ethnicity information is not available.

Appendix B: Average RIT Scores by State

Table B.1. Average RIT Scores by State and Grade—Reading

State		K	1	2	3	4	5	6	7	8	9	10	11	12
AK	RIT	–	173.6	192.7	187.8	197.5	207.4	211.6	215.9	219.8	210.6	216.7	222.3	226
	N	–	343	359	3,904	3,833	6,944	8,655	12,495	12,200	862	566	513	451
AL	RIT	146.8	164	178.5	188.3	199.3	205.5	209.6	211.3	215.6	215.5	214.2	–	–
	N	341	660	686	573	648	674	702	619	601	336	306	–	–
AZ	RIT	139.6	156.9	168.8	180.3	188.2	195.8	200.9	204.7	209.9	210.9	210.8	213.8	214.8
	N	2,117	2,481	2,753	3,242	3,020	2,969	2,893	2,615	2,507	962	732	636	608
CA	RIT	145.3	165.4	177.4	188.9	197.4	204.1	208.7	212.8	217.2	217.6	218.4	218.2	214.3
	N	41,776	52,598	63,656	65,176	67,247	68,155	64,557	63,036	60,510	38,187	30,818	15,575	6,989
CO	RIT	151.4	169.4	180.4	193.4	201.3	208.0	210.1	215.0	217.9	218.7	219.7	209.4	210.6
	N	412	864	3,485	3,749	3,777	3,629	3,171	2,946	2,913	2,702	2,399	638	503
CT	RIT	149.9	166.7	181.9	192.4	201.8	208.6	213.3	217.4	221.5	221.3	221.7	221.2	213.0
	N	14,839	26,571	30,511	32,697	35,833	36,269	37,622	36,128	35,517	22,123	16,253	3,860	1,323
DC	RIT	148.9	166.4	179.5	189.0	197.5	202.4	206.1	210.2	214.7	212.2	212.7	215.2	212.9
	N	8,927	8,265	7,871	7,272	6,417	6,015	6,008	5,525	4,857	3,584	2,513	1,505	832
DE	RIT	144.2	166.2	182.3	194.9	204.8	212.0	212.9	214.4	219.1	223.6	223.5	224.8	225.5
	N	3,054	7,199	7,011	6,385	6,045	6,485	4,044	3,516	3,185	2,453	2,175	1,219	541
FL	RIT	151.3	170.6	183.6	194.7	204.3	209.9	213.2	217.0	220.5	220.2	223.0	223.1	211.5
	N	16,611	16,533	16,626	16,769	15,414	15,114	16,382	14,174	12,728	2,819	2,703	1,160	376
GA	RIT	156.7	175.2	187.4	198.0	–	–	216.6	219.3	–	–	–	–	–
	N	637	670	573	328	–	–	417	417	–	–	–	–	–
HI	RIT	155.0	174.4	185.9	198.1	206.0	213.0	220.5	225.5	229.1	230.4	231.1	231.2	226.1
	N	641	967	1,034	1,453	1,808	1,850	2,011	2,701	2,627	2,872	1,292	606	467
ID	RIT	145.8	164.6	181.2	193.2	202.5	208.7	214.2	218.7	223.1	221.8	224.8	223.7	–
	N	3,364	4,731	5,888	5,861	6,226	6,193	6,065	5,917	5,744	3,308	2,639	1,212	–
IL	RIT	148.1	167.2	180.5	192.2	201.4	208.4	213.5	218.1	222.1	219.1	220.3	220.3	215.0
	N	14,4843	190,274	303,993	332,108	335,970	333,372	331,355	328,623	323,368	90,022	65,527	31,344	10,655

Reading														
State		Grade												
		K	1	2	3	4	5	6	7	8	9	10	11	12
IN	RIT	–	–	–	–	–	–	–	208.0	209.6	209.7	213.4	212.8	–
	N	–	–	–	–	–	–	–	853	763	719	666	594	–
KY	RIT	148.4	168.1	180.3	192.7	201.5	208.8	213.5	217.3	221.0	221.0	224.2	222.0	213.8
	N	103,289	117,157	126,429	131,838	129,857	126,711	114,563	116,372	114,004	51,333	33,069	9,603	834
LA	RIT	147.6	165.3	177.6	188.0	196.4	201.6	205.3	209.7	213.0	213.1	215.2	213.7	216.5
	N	18,477	19,837	20,026	16,343	15,130	13,994	13,490	12,652	11,537	10,302	6,884	1,516	761
MA	RIT	136.4	152.5	166.7	180.2	188.3	194.0	199.9	201.0	206.2	–	–	–	–
	N	816	763	917	857	904	810	580	564	592	–	–	–	–
MD	RIT	148.0	165.1	179.8	194.0	198.3	204.4	211.3	215.8	221.3	221.4	218.1	220.6	–
	N	455	588	429	360	480	588	615	756	593	762	402	358	–
ME	RIT	150.0	166.4	180.9	191.8	201.2	208.2	213.7	218.1	222.0	224.0	224.4	221.9	221.2
	N	8,681	14,715	20,873	26,145	26,531	25,934	26,922	27,699	26,790	14,650	9,045	2,828	1,641
MI	RIT	146.7	165.1	178.9	189.3	198.2	205.1	209.5	213.3	216.7	216.4	218.6	217.2	214.4
	N	214,348	237,535	252,892	256,232	266,776	271,413	256,737	244,719	233,190	124,305	112,172	54,742	19,047
MO	RIT	148.8	166.9	180.8	190.6	201.0	206.8	210.5	214.9	218.0	221.5	223.2	223.7	220.1
	N	11,329	13,640	19,462	16,439	18,880	15,380	13,834	11,925	11,878	4,627	3,394	1,829	888
MS	RIT	150.4	172.3	184.5	193.4	201.8	208.9	212.6	215.3	218.7	217.5	220.4	215.2	210.2
	N	22,675	26,687	27,059	21,085	21,502	19,682	22,213	24,138	23,176	12,271	11,106	3,146	379
MT	RIT	149.9	168.7	181.4	192.0	201.4	208.1	213.0	217.1	220.9	220.9	224.1	222.8	221.4
	N	10,007	11,414	14,658	21,841	21,943	22,029	21,062	17,609	17,222	8,267	11,391	3,156	1,140
NC	RIT	149.5	169.9	183.2	195.4	204.2	210.7	215.6	219.0	222.1	225.6	227.8	226.5	221.8
	N	40,365	55,442	58,029	65,457	64,837	63,710	58,536	54,941	54,054	4,096	2,723	1,895	705
NE	RIT	–	–	–	189.9	199.7	206.1	209.1	211.2	217.2	216.5	217.4	220.2	–
	N	–	–	–	2,682	2,552	2,544	2,295	2,002	2,336	1,924	1,796	1,616	–
NH	RIT	151.4	168.5	183.0	194.8	203.8	211.0	216.0	220.1	224.0	225.3	226.2	222.7	220.4
	N	4,707	11,318	15,519	16,813	17,111	17,379	15,713	14,668	13,758	5,417	4,126	1,199	653
NJ	RIT	150.8	170.6	184.9	195.7	204.0	210.5	215.4	218.5	221.9	218.1	219.7	219.8	213.9
	N	19,351	27,577	34,994	34,160	35,505	34,145	33,519	26,977	25,344	6,263	5,267	3,542	1,784

Reading														
State		Grade												
		K	1	2	3	4	5	6	7	8	9	10	11	12
NM	RIT	145.9	163.3	175.5	186.3	195.0	202.2	207.3	212.1	216.6	214.3	217.6	219.8	220.4
	N	8,684	9,725	14,045	16,979	17,159	17,229	18,538	15,511	15,158	8,702	7,128	5,730	3,448
NV	RIT	146.3	162.1	175.8	189.1	199.2	206.2	211.2	215.4	219.9	220.3	219.4	219.1	218.3
	N	20,758	59,903	61,780	65,875	42,335	40,669	32,885	28,571	27,563	10,099	5,675	4,372	2,794
NY	RIT	145.4	163.7	175.5	188.6	198.4	204.9	209.5	214.2	219.1	–	–	–	–
	N	1,352	1,323	1,404	1,106	1,009	953	992	1,016	808	–	–	–	–
OK	RIT	149.7	–	–	–	201.7	201.9	208.9	216.8	–	230.3	–	–	–
	N	301	–	–	–	550	747	1,102	629	–	345	–	–	–
OR	RIT	150.8	167.6	182.3	193.8	203.0	211.0	213.9	218.5	222.5	222.7	225.1	225.0	219.1
	N	3,363	5,449	7,860	8,327	9,030	8,347	9,432	9,086	8,789	5,734	5,250	2,203	875
PA	RIT	148.7	170.3	186.0	192.2	202.2	208.4	212.3	217.3	222.0	205.0	206.3	206.0	–
	N	629	1,774	1,675	1,962	1,882	1,852	2,100	2,061	1,781	534	394	302	–
RI	RIT	152.8	175.4	186.8	198.2	205.8	210.4	212.5	216.6	219.0	213.6	217.4	221.8	–
	N	1,430	1,578	2,017	2,049	2,075	2,521	2,693	2,887	2,597	2,613	1,893	835	–
SD	RIT	146.1	163.6	178.2	188.4	197.8	205.4	210.1	213.5	217.0	217.0	220.4	223.5	222.0
	N	14,026	15,468	15,534	16,936	16,873	21,059	15,187	12,943	12,306	9,929	8,979	6,553	3,018
TN	RIT	148.3	167.0	177.7	188.9	195.5	202.6	206.4	209.9	214.2	212.9	216.8	216.1	215.8
	N	36,135	35,032	35,159	35,793	32,582	36,454	32,203	31,064	30,091	22,470	20,220	13,533	7,703
TX	RIT	146.7	166.4	179.7	195.3	205.5	204.3	211.0	218.6	220.5	228.4	230.7	–	–
	N	1,305	982	990	1,140	822	1,878	1,149	897	1,218	338	322	–	–
UT	RIT	149.8	166.6	180.3	189.8	199.2	206.8	212.9	217.1	221.3	223.4	225.0	225.3	215.7
	N	3,762	4,591	4,860	3,654	3,868	3,583	3,808	3,932	3,608	3,138	3,018	2,397	331
VT	RIT	151.3	166.9	180.7	190.6	199.9	207.5	212.9	216.6	221.0	221.8	222.6	220.4	222.3
	N	1,331	1,771	2,184	3,073	2,942	3,124	3,193	3,042	3,089	2,475	1,878	590	388
WA	RIT	149.7	167.4	181.4	191.8	201.1	208.2	213.3	217.7	221.6	220.7	218.5	215.2	212.6
	N	26,558	43,070	62,844	69,895	68,801	67,763	57,735	57,709	57,391	21,262	10,736	5,221	3,121
WI	RIT	152.1	170.7	183.1	194.3	203.1	209.9	215.0	219.5	223.4	223.5	224.0	221.4	220.4
	N	38,217	52,662	82,226	104,532	108,002	108,603	108,703	106,972	103,085	31,557	21,484	5,858	2,457
WY	RIT	154.0	174.0	185.0	196.8	205.3	212.1	216.0	219.3	223.0	224.7	226.3	224.4	218.8
	N	15,424	21,988	22,496	22,729	22,789	22,422	19,801	17,915	17,801	9,047	6,989	2,317	666

Table B.2. Average RIT Scores by State and Grade—Language Usage

Language Usage												
State		Grade										
		2	3	4	5	6	7	8	9	10	11	12
AK	RIT	–	–	–	–	–	–	–	218.6	223.0	228.0	229.0
	N	–	–	–	–	–	–	–	438	401	411	389
AL	RIT	–	189.4	199.1	206.0	209.7	211.2	214.9	214.5	216.7	–	–
	N	–	573	638	655	671	590	581	308	300	–	–
AZ	RIT	171.6	182.0	190.4	197.6	203.3	206.2	210.6	209.7	212.6	215.2	214.6
	N	1,199	1,632	1,572	1,598	1,459	1,242	1,116	840	658	559	469
CA	RIT	181.1	193.0	200.8	206.7	212.8	216.4	219.3	216.6	218.3	217.2	217.7
	N	30,453	31,960	34,319	33,917	24,329	22,179	21,357	7,414	6,880	2,104	1,683
CO	RIT	179.9	195.0	203.9	210.5	–	–	–	–	–	–	–
	N	396	532	501	467	–	–	–	–	–	–	–
CT	RIT	179.9	192.3	200.8	206.1	211.8	216.4	220.5	218.4	220.6	216.8	215.4
	N	5,185	5,240	9,045	8,618	12,025	12,421	12,322	4,127	3,813	506	408
DE	RIT	–	–	–	–	–	–	–	–	215.0	–	–
	N	–	–	–	–	–	–	–	–	371	–	–
FL	RIT	183.8	195.3	203.5	207.8	212.9	216.3	220.7	222.8	–	–	–
	N	363	451	536	505	424	407	366	319	–	–	–
GA	RIT	–	200.0	210.3	–	217.6	219.3	–	–	–	–	–
	N	–	321	303	–	408	417	–	–	–	–	–
HI	RIT	–	–	–	–	–	–	–	225.2	228.7	229.5	226.5
	N	–	–	–	–	–	–	–	628	814	453	453
ID	RIT	184.2	194.5	203.2	209.3	213.7	217.6	221.8	222.8	226.0	223.3	–
	N	2,488	4,366	4,501	4,812	4,622	4,344	4,236	3,340	2,970	964	–
IL	RIT	182.5	193.5	202.2	208.4	211.7	216.1	219.9	217.3	219.5	221.1	212.9
	N	24,995	40,075	41,090	45,189	53,038	54,293	53,924	20,748	17,314	9,512	2,209
IN	RIT	–	–	–	–	–	208.1	208.7	–	–	–	–
	N	–	–	–	–	–	489	493	–	–	–	–
KY	RIT	180.8	193.1	201.8	208.0	212.8	216.2	219.3	218.4	221.1	221.7	–
	N	30,737	45,199	60,637	49,440	54,217	41,487	41,020	12,133	9,708	4,091	–

Language Usage												
State		Grade										
		2	3	4	5	6	7	8	9	10	11	12
LA	RIT	179.7	191.4	199.7	203.8	207.2	211.3	213.9	213.0	217.5	–	–
	N	7,596	9,017	8,344	8,048	7,364	6,539	6,194	6,344	5,040	–	–
MD	RIT	–	–	–	–	218.6	221.9	224.5	221.2	217.2	218.8	–
	N	–	–	–	–	320	319	333	719	387	347	–
ME	RIT	180.5	192.3	202.1	208.5	212.2	216.0	219.8	219.0	219.7	219.3	220.0
	N	2,786	5,249	5,824	6,191	8,033	7,930	7,866	4,294	3,360	1,307	861
MI	RIT	177.1	189.5	198.2	204.4	208.4	212.1	215.4	215.7	218.2	218.2	214.2
	N	58,348	104,048	109,915	110,979	117,329	118,678	116,178	69,621	61,266	33,420	7,721
MO	RIT	179.9	190.8	199.5	205.9	209.6	215.5	218.4	222.5	223.2	223.4	219.0
	N	1,973	6,457	6,385	6,308	6,261	5,902	5,242	3,932	2,806	1,756	623
MS	RIT	182.4	192.8	201.6	208.2	212.4	215.4	218.6	216.9	219.5	219.1	–
	N	10,179	9,907	10,555	10,810	13,006	13,062	12,302	5,163	5,674	2,452	–
MT	RIT	181.3	191.8	200.8	207.2	211.8	215.9	219.7	219.9	222.5	222.2	219.7
	N	3,671	12,719	12,906	13,461	14,329	14,713	14,751	6,487	8,707	2,545	779
NC	RIT	185.5	196.1	202.6	209.5	214.9	218.7	222.6	222.9	226.8	226.3	223.0
	N	3,362	3,437	3,527	3,312	2,941	2,971	2,503	1,067	888	705	532
NH	RIT	179.5	194.0	202.1	208.9	214.8	217.5	221.2	222.0	223.8	219.6	–
	N	1,299	2,536	2,311	2,814	2,388	2,686	2,782	1,709	1,522	439	–
NJ	RIT	186.8	196.6	204.8	210.2	214.2	215.6	219.3	216.3	217.6	216.6	214.7
	N	4,795	10,457	11,639	10,771	10,000	8,020	7,335	2,928	2,197	1,191	1,013
NM	RIT	174.1	186.3	193.7	200.2	205.7	208.7	212.6	213.8	215.9	217.9	217.6
	N	4,794	8,434	8,628	8,728	9,496	6,808	6,589	4,956	3,826	2,792	1,564
NV	RIT	179.5	190.5	199.2	204.7	210.5	214.7	218.0	216.3	219.9	220.1	218.9
	N	5,356	6,407	6,150	5,296	4,322	2,829	2,455	2,253	2,540	2,278	1,850
OR	RIT	181.8	192.6	200.8	208.3	210.9	215.0	219.1	219.8	222.2	220.7	218.6
	N	1,498	2,300	2,329	2,319	3,103	3,096	3,084	1,962	1,929	1,065	497
PA	RIT	187.6	197.1	205.4	214.5	215.2	220.2	225.3	–	–	–	–
	N	322	682	986	694	1,761	1,735	1,381	–	–	–	–

Language Usage												
State		Grade										
		2	3	4	5	6	7	8	9	10	11	12
RI	RIT	–	196.1	205.4	210.2	215.7	217.5	221.5	219.9	225.1	226.4	–
	N	–	527	484	506	476	564	579	465	443	404	–
SD	RIT	178.0	187.9	196.8	204.9	209.6	213.4	216.3	217.2	219.5	221.9	221.0
	N	1,907	8,817	8,330	14,062	8,580	7,484	7,080	7,536	6,636	4,669	2,167
TN	RIT	179.8	189.6	196.9	203.0	208.1	211.6	216.3	216.2	215.2	217.7	214.4
	N	6,980	10,792	9,904	10,766	9,355	9,353	8,667	2,284	2,170	1,952	861
TX	RIT	–	204.0	210.0	216.7	–	223.9	224.9	–	–	–	–
	N	–	483	451	415	–	340	354	–	–	–	–
UT	RIT	180.7	191.1	200.4	206.9	212.3	215.2	219.0	220.6	222.9	224.0	215.4
	N	3,386	3,502	3,816	3,560	3,318	3,293	3,061	2,411	2,304	1,845	305
VT	RIT	179.1	190.3	198.9	205.6	210.2	213.9	218.2	220.3	221.6	–	–
	N	836	1,625	1,491	1,512	1,775	1,926	1,962	1,658	1,483	–	–
WA	RIT	186.8	198.0	206.0	212.0	215.5	219.2	223.2	213.5	214.7	215.3	211.2
	N	6,102	9,284	9,663	9,188	10,056	9,613	8,723	2,150	1,854	1,154	672
WI	RIT	184.5	196.4	204.8	210.8	215.4	219.6	223.4	221.9	224.8	222.1	219.3
	N	9,845	19,563	20,911	22,257	27,092	27,120	26,919	9,607	6,109	2,051	706
WY	RIT	185.3	196.6	203.7	209.9	214.0	217.1	219.8	221.3	223.3	221.8	221.1
	N	5,605	6,444	7,045	7,858	10,315	9,607	8,638	4,831	3,997	1,437	532

Table B.3. Average RIT Scores by State and Grade—Mathematics

State		K	1	2	3	4	5	6	7	8	9	10	11	12
AK	RIT	–	179.0	195.5	188.6	199.8	213.2	216.8	222.5	227.6	222.0	232.2	241.6	241.7
	N	–	350	351	3,891	3,829	6,926	8,607	12,582	12,028	1,195	495	434	402
AL	RIT	145.2	164.3	183.1	189.7	201.8	210.3	215.1	217.9	224.0	223.8	228.0	–	–
	N	334	659	685	565	655	677	693	621	588	320	366	–	–
AZ	RIT	136.2	158.4	172.8	184.8	194.1	203.0	208.1	213.0	218.0	220.6	223.1	227.5	229.1
	N	2,191	2,662	2,750	3,156	3,018	2,940	2,873	2,594	2,432	959	688	597	605
CA	RIT	144.0	167.4	180.1	191.9	202.3	211.1	213.9	219.3	224.3	224.8	226.5	227.7	224.9
	N	41,709	52,921	65,035	67,279	69,929	70,770	68,842	63,735	60,095	36,954	29,604	15,753	7,977
CO	RIT	150.2	170.5	181.3	195.0	205.4	213.5	213.0	219.0	223.6	228.4	230.7	225.3	224.1
	N	404	863	3,465	3,743	3,786	3,647	3,893	3,821	3,890	2,542	2,262	746	347
CT	RIT	148.1	167.7	184.9	193.9	204.9	213.7	217.7	223.9	229.5	229.9	232.5	234.8	223.3
	N	17,933	30,244	34,422	38,213	39,152	38,569	38,918	37,907	37,667	22,851	18,225	5,512	1,231
DC	RIT	147.8	168.6	183.8	193.0	203.0	209.0	211.2	216.8	222.4	218.9	220.8	220.0	220.4
	N	9,234	8,532	8,208	7,432	6,455	6,102	6,089	5,594	5,160	11,526	8,574	5,354	1,152
DE	RIT	146.7	168.1	184.0	195.9	207.2	216.8	217.0	220.0	226.8	232.0	232.4	231.7	227.9
	N	3,823	7,619	7,562	6,479	6,072	6,674	4,108	3,683	3,196	2,200	2,040	1,164	419
FL	RIT	150.3	173.0	184.2	196.1	207.6	216.0	217.1	221.9	226.5	227.3	230.3	231.4	–
	N	16,542	16,464	16,561	16,674	15,431	15,137	16,374	14,249	12,631	2,591	2,525	1,125	–
GA	RIT	156.9	176.5	190.3	199.5	–	–	214.7	218.2	221.2	–	–	–	–
	N	636	667	588	326	–	–	1,849	2,078	1,617	–	–	–	–
HI	RIT	154.0	176.1	185.6	197.6	208.5	219.4	226.0	232.8	239.5	242.8	242.4	244.2	241.7
	N	921	1,242	1,197	1,665	1,876	1,885	2,016	2,731	2,610	2,700	1,196	533	462
ID	RIT	144.1	165.7	182.6	194.0	205.5	214.9	219.3	225.2	231.1	232.3	236.9	234.4	229.8
	N	3,322	4,860	5,957	5,945	6,200	6,197	6,583	7,285	7,113	4,036	3,148	1,301	317
IL	RIT	146.7	169.1	182.9	194.7	205.4	214.2	218.4	225.0	230.7	226.3	228.6	230.1	224.1
	N	160,523	211,693	306,580	329,942	335,258	332,835	338,729	330,412	326,860	81,035	59,039	31,290	9,472
IN	RIT	–	–	–	–	204.4	215.0	215.9	217.9	222.3	218.6	223.1	224.7	–
	N	–	–	–	–	330	473	531	1,023	1,196	717	659	612	–

Mathematics														
State		Grade												
		K	1	2	3	4	5	6	7	8	9	10	11	12
KY	RIT	147.3	170.1	182.1	194.5	204.9	214.0	217.7	223.7	229.0	229.4	233.1	230.2	219.9
	N	103,144	119,042	126,819	130,406	129,867	127,215	117,161	118,577	116,433	48,497	30,425	9,953	1,199
LA	RIT	146.1	166.8	180.3	190.7	200.2	207.2	210.2	216.7	221.3	222.1	228.8	219.5	–
	N	18,442	19,839	20,066	16,414	15,219	14,154	13,896	13,056	11,589	9,806	6,156	853	–
MA	RIT	132.2	153.5	170.4	183.1	194.0	202.5	206.9	211.7	216.4	–	–	–	–
	N	810	763	920	853	911	809	968	974	1,265	–	–	–	–
MD	RIT	145.8	165.3	190.8	199.2	208.5	213.4	215.4	223.4	227.7	226.4	223.5	227.0	–
	N	526	614	447	534	625	879	829	655	528	628	392	359	–
ME	RIT	149.0	168.4	184.6	193.9	204.7	213.9	218.3	224.6	230.4	232.6	234.0	231.5	228.2
	N	7,954	14,463	20,656	26,288	27,250	26,592	27,722	27,952	26,885	14,390	9,434	3,939	1,751
MI	RIT	145.4	167.3	182.4	191.6	202.1	210.9	214.2	219.9	224.7	224.3	227.5	226.8	222.2
	N	212,836	237,434	252,717	260,011	267,239	272,418	258,803	247,069	234,212	121,550	111,024	58,029	18,076
MO	RIT	148.5	170.0	183.9	193.2	204.7	212.3	215.6	222.8	226.4	233.0	234.2	236.3	–
	N	11,429	14,008	19,888	16,677	18,931	15,354	13,834	12,763	11,966	4,424	3,074	1,845	–
MS	RIT	148.8	173.1	185.2	194.4	204.2	213.6	217.1	222.8	228.0	226.6	226.9	223.4	217.9
	N	22,962	26,971	28,022	21,773	21,863	20,046	22,314	24,379	23,293	12,397	7,302	2,655	447
MT	RIT	149.3	170.6	183.1	193.5	204.4	213.4	217.9	224.2	230.0	230.6	235.9	236.5	235.2
	N	9,702	10,992	14,658	21,807	21,949	21,974	21,603	18,131	17,653	8,613	11,336	3,392	1,127
NC	RIT	147.0	169.9	183.5	196.3	208.3	218.5	221.4	227.9	233.3	235.7	240.5	240.4	235.1
	N	58,419	64,717	66,748	69,952	64,997	61,517	60,102	55,490	53,966	3,457	2,484	1,765	695
NE	RIT	–	–	–	190.2	203.2	212.6	215.6	220.3	226.0	225.2	228.1	233.8	–
	N	–	–	–	2,663	2,551	2,472	2,112	1,999	2,201	1,922	1,768	1,622	–
NH	RIT	151.3	170.2	185.4	196.2	206.6	216.1	221.1	227.8	233.4	234.8	237.7	234.4	230.7
	N	4,731	11,292	15,993	17,096	17,257	17,597	16,589	15,931	14,215	6,174	4,542	1,520	635
NJ	RIT	150.2	172.2	187.4	197.1	208.3	217.4	221.8	227.5	230.5	226.1	228.5	229.7	224.7
	N	19,269	30,748	40,603	37,978	39,372	42,105	42,809	36,181	29,094	8,394	6,816	4,669	2,056
NM	RIT	143.5	165.1	180.7	190.5	200.8	209.2	213.9	218.9	224.0	222.2	226.5	228.7	229.2
	N	10,254	11,545	15,467	16,592	16,615	17,079	18,975	15,856	14,969	7,934	6,559	5,243	2,880

Mathematics														
State		Grade												
		K	1	2	3	4	5	6	7	8	9	10	11	12
NV	RIT	144.5	163.1	177.2	190.4	203.0	212.0	216.5	222.4	228.2	227.8	226.8	228.8	229.4
	N	19,325	61,466	60,810	62,443	41,995	40,623	33,567	29,208	27,480	7,458	4,021	3,222	2,750
NY	RIT	145.8	168.7	183.5	190.1	201.8	209.9	211.8	218.2	225.4	–	–	–	–
	N	2,260	2,463	2,425	1,137	1,009	929	1,065	1,077	892	–	–	–	–
OK	RIT	147.6	–	–	192.9	202.5	208.2	211.5	217.7	216.4	–	–	–	–
	N	301	–	–	307	545	763	1,409	1,039	1,533	–	–	–	–
OR	RIT	150.4	170.2	182.8	194.1	205.8	215.4	219.0	226.2	231.8	230.9	234.3	232.9	226.5
	N	4,741	6,138	8,345	8,557	9,213	8,876	9,268	9,048	9,195	5,673	5,098	3,286	1,349
PA	RIT	148.0	171.2	188.6	193.1	205.2	214.4	217.3	223.2	225.1	213.4	212.3	–	–
	N	629	1,755	1,664	1,994	1,909	1,801	2,111	2,036	2,282	431	346	–	–
RI	RIT	151.3	175.4	188.5	199.0	208.2	215.3	218.8	225.1	229.8	224.8	228.7	230.4	–
	N	1,774	1,897	2,408	2,188	2,165	2,456	2,401	2,529	2,505	2,444	1,778	878	–
SD	RIT	145.0	165.8	182.1	190.7	201.6	211.1	215.3	220.8	225.4	227.2	231.8	236.2	234.6
	N	13,991	15,475	15,534	17,080	16,941	20,977	15,560	13,310	12,694	10,892	9,816	6,599	3,038
TN	RIT	146.3	168.3	179.5	190.8	199.2	207.7	210.8	215.5	220.9	220.5	223.3	223.4	222.9
	N	36,056	35,066	35,348	35,821	32,601	36,991	32,202	30,929	29,724	22,474	19,340	14,031	8,754
TX	RIT	144.3	168.7	181.3	195.9	208.3	210.6	216.5	225.3	228.4	233.6	237.4	–	–
	N	1,286	972	992	1,113	827	1,807	1,177	951	1,293	425	372	–	–
UT	RIT	148.9	169.0	183.6	192.8	204.5	213.7	218.3	223.6	230.0	233.4	237.6	238.8	–
	N	3,816	4,738	5,103	3,718	3,895	3,562	3,752	3,969	3,629	3,148	2,876	2,218	–
VT	RIT	151.7	168.5	184.2	192.0	202.5	212.5	217.1	222.6	229.4	231.6	233.3	232.9	232.6
	N	1,479	1,925	2,391	3,335	3,214	3,389	3,533	3,094	3,184	2,493	2,001	832	387
WA	RIT	149.6	170.0	184.0	193.7	205.0	214.3	218.7	224.8	229.6	228.0	227.5	224.0	219.2
	N	28,372	45,298	65,371	71,340	69,805	69,311	60,233	57,271	50,942	18,334	11,954	6,356	3,264
WI	RIT	152.4	173.6	186.1	196.9	207.8	216.9	221.5	228.5	234.6	234.0	235.5	230.5	222.2
	N	42,144	59,507	86,262	106,899	109,522	109,188	110,028	106,208	103,034	31,391	21,649	5,783	1,296
WY	RIT	153.8	176.5	186.9	199.2	210.0	219.2	222.2	227.3	232.3	235.0	237.8	236.5	232.3
	N	15,503	21,916	22,403	22,729	22,862	22,672	19,913	18,075	17,395	9,678	6,999	2,951	875

Table B.4. Average RIT Scores by State and Grade—Science

Science												
State		Grade										
		2	3	4	5	6	7	8	9	10	11	12
AR	RIT	–	189.6	196.7	202.9	206.9	210.3	211.7	214.4	210.3	208.9	–
	N	–	5,227	6,398	7,475	7,475	7,597	7,447	1,947	923	466	–
CA	RIT	–	186.5	192.2	194.7	204.3	207.7	207.2	211.1	212.9	210.3	210.3
	N	–	1,475	1,736	15,237	8,507	8,754	19,599	3,214	2,388	1,002	547
CO	RIT	–	–	199.4	203.2	206.5	211.5	214.7	217.3	219.8	217.3	–
	N	–	–	3,678	4,688	7,335	7,113	7,684	2,763	2,605	661	–
CT	RIT	–	–	202.5	203.5	208.0	210.1	213.4	218.2	221.2	224.3	–
	N	–	–	496	3,083	3,430	3,662	3,833	1,634	1,530	1,170	–
DC	RIT	–	–	–	–	199.5	201.3	204.9	–	–	–	–
	N	–	–	–	–	446	459	454	–	–	–	–
DE	RIT	–	–	–	–	–	–	–	219.7	–	–	–
	N	–	–	–	–	–	–	–	346	–	–	–
GA	RIT	–	184.1	191.6	196.9	201.2	204.1	206.8	–	–	–	–
	N	–	8,108	7,425	7,791	6,892	6,684	6,693	–	–	–	–
IA	RIT	–	193.2	199.7	204.6	207.2	211.0	214.2	216.1	218.1	218.8	214.8
	N	–	2,603	3,524	5,134	6,301	8,227	8,540	4,438	4,444	3,407	577
IL	RIT	–	189.6	195.6	200.9	203.5	207.3	210.4	217.0	218.3	217.2	–
	N	–	12,796	15,088	18,895	21,916	22,866	21,846	902	504	360	–
KS	RIT	–	192.8	200.3	204.7	207.9	211.3	215.0	216.3	218.6	218.8	220.5
	N	–	507	972	2,576	4,313	4,843	4,820	1,611	1,400	1,145	498
KY	RIT	182.1	191.4	198.3	204.2	208.0	211.7	215.0	214.8	–	–	–
	N	437	3,665	6,274	3,270	4,972	7,245	4,393	1,501	–	–	–
MA	RIT	–	–	193.1	197.0	–	–	208.2	–	–	–	–
	N	–	–	312	2,775	–	–	1,704	–	–	–	–
MD	RIT	–	–	–	204.0	214.0	217.7	218.6	214.5	–	–	–
	N	–	–	–	349	646	650	633	440	–	–	–
MI	RIT	180.0	189.6	196.6	202.2	205.1	208.6	211.6	213.4	215.0	215.1	211.7
	N	624	45,092	55,427	54,543	65,537	60,461	58,554	13,932	11,876	4,466	1,059

Science												
State		Grade										
		2	3	4	5	6	7	8	9	10	11	12
MO	RIT	–	–	–	206.7	208.0	210.9	214	–	–	–	–
	N	–	–	–	1,450	1,327	1,288	1,238	–	–	–	–
MT	RIT	–	193.3	200.4	205.9	209.1	212.3	215.1	218.0	220.5	–	–
	N	–	583	737	702	703	808	988	363	417	–	–
NC	RIT	–	–	–	–	210	–	–	–	–	–	–
	N	–	–	–	–	311	–	–	–	–	–	–
NJ	RIT	–	190.2	195.4	200.9	205.2	207.5	210.1	–	–	–	–
	N	–	1,091	1,134	1,053	1,657	1,860	1,946	–	–	–	–
NV	RIT	–	190.8	197.1	201.6	205.9	208.0	211.3	216.8	–	–	–
	N	–	674	926	1,440	1,694	1,879	1,813	581	–	–	–
NY	RIT	–	–	–	–	201.6	206.4	208.7	–	–	–	–
	N	–	–	–	–	634	981	430	–	–	–	–
OH	RIT	–	196.6	203.8	208.7	211.2	215.4	219.0	–	–	–	–
	N	–	747	938	1036	1,129	1,083	910	–	–	–	–
OK	RIT	–	–	–	205.2	204.8	206.9	212.5	–	–	–	–
	N	–	–	–	485	393	442	362	–	–	–	–
OR	RIT	–	–	205.3	–	206.8	210.0	215.1	212.8	217.9	–	–
	N	–	–	312	–	373	354	401	355	357	–	–
RI	RIT	–	194.1	201.7	205.5	210.0	214.0	219.1	–	–	–	–
	N	–	442	465	495	552	483	428	–	–	–	–
SD	RIT	–	–	–	–	209.9	213.9	216.9	–	–	–	–
	N	–	–	–	–	1,274	1,284	1,172	–	–	–	–
WA	RIT	–	194.2	200.8	204.5	208.5	211.6	214.9	215.2	215.5	–	–
	N	–	1,427	1,927	3924	4,008	5,673	4,312	696	622	–	–
WI	RIT	–	–	202.7	207.5	210.9	215.2	218.7	–	–	–	–
	N	–	–	1,037	1121	1,295	1,219	1,319	–	–	–	–

Appendix C: Test-Retest Reliability by State

Table C.1. Test-Retest with Alternate Forms Reliability by State—Reading Overall

State	Fall 2016–Winter 2017		Spring 2017–Fall 2017		Winter 2017–Spring 2017	
	N	Reliability	N	Reliability	N	Reliability
AK	7,528	0.904	9,768	0.868	7,470	0.892
AL	1,084	0.920	933	0.875	966	0.887
AZ	3,803	0.937	3,990	0.924	4,115	0.933
CA	149,531	0.944	109,431	0.933	122,029	0.940
CO	8,645	0.913	1,762	0.896	7,114	0.899
CT	67,303	0.938	47,776	0.933	78,686	0.937
DC	14,773	0.930	11,367	0.911	14,771	0.926
DE	10,753	0.933	9,689	0.932	10,736	0.939
FL	45,860	0.942	1,098	0.921	44,887	0.933
GA	1,173	0.962	–	–	1,164	0.957
HI	3,895	0.945	3,470	0.905	3,457	0.949
ID	10,033	0.936	9,779	0.936	10,144	0.946
IL	543,929	0.946	514,288	0.933	660,222	0.936
IN	1,343	0.825	–	–	1,272	0.833
KY	254,890	0.951	219,462	0.932	258,211	0.946
LA	47,702	0.927	366	0.816	47,086	0.922
MD	533	0.948	869	0.859	542	0.938
ME	28,795	0.938	48,324	0.931	30,812	0.937
MI	518,120	0.939	506,251	0.923	495,175	0.933
MO	41,468	0.940	–	–	39,878	0.939
MS	75,613	0.940	–	–	64,740	0.940
MT	33,372	0.936	36,340	0.922	34,242	0.932
NC	123,060	0.950	91,190	0.938	122,912	0.950
NE	5,917	0.898	1,196	0.899	1,374	0.883
NH	22,370	0.940	19,321	0.928	19,149	0.935
NJ	58,838	0.941	905	0.796	61,214	0.938
NM	28,428	0.934	23,113	0.932	25,256	0.928
NV	69,788	0.944	58,607	0.930	60,881	0.939
NY	1,598	0.949	1,733	0.930	1,593	0.946
OK	881	0.950	–	–	354	0.884
OR	16,417	0.932	14,536	0.924	14,874	0.930
PA	3,215	0.934	2,593	0.895	3,421	0.925
RI	4,632	0.914	4,493	0.913	4,852	0.907
SD	33,294	0.941	29,705	0.928	32,595	0.934
TN	109,494	0.936	1,298	0.882	106,578	0.924
TX	916	0.954	1,356	0.918	1,278	0.964
UT	9,548	0.944	7,745	0.935	8,612	0.946
VT	5,539	0.925	4,821	0.920	5,324	0.931
WA	104,066	0.938	87,945	0.933	95,228	0.938
WI	181,922	0.941	161,533	0.926	186,303	0.934
WY	43,164	0.941	13,069	0.932	44,404	0.940

Table C.2. Test-Retest with Alternate Forms Reliability by State—Reading K–2

	Fall 2016–Winter 2017		Spring 2017–Fall 2017		Winter 2017–Spring 2017	
State	N	Reliability	N	Reliability	N	Reliability
AK	372	0.920	–	–	323	0.912
AL	408	0.863	308	0.829	401	0.836
AZ	1,621	0.858	1,429	0.836	1,818	0.863
CA	61,766	0.903	38,044	0.896	51,326	0.906
CO	4,394	0.886	470	0.873	4,311	0.889
CT	25,351	0.890	14,488	0.870	28,679	0.888
DC	5,374	0.844	3,102	0.857	5,038	0.851
DE	5,498	0.896	3,495	0.870	5,587	0.891
FL	19,998	0.878	360	0.853	19,715	0.871
GA	316	0.868	–	–	313	0.847
HI	1,342	0.891	650	0.854	836	0.890
ID	3,820	0.882	2,985	0.862	3,448	0.874
IL	243,370	0.905	187,486	0.892	309,464	0.896
KY	113,028	0.901	80,416	0.874	114,468	0.899
LA	16,825	0.858	–	–	17,297	0.857
ME	13,574	0.893	14,551	0.883	13,940	0.890
MI	193,484	0.883	154,451	0.866	188,391	0.880
MO	17,372	0.881	–	–	16,919	0.884
MS	27,902	0.869	–	–	23,548	0.876
MT	15,288	0.876	12,676	0.858	15,797	0.877
NC	60,429	0.908	39,143	0.898	60,413	0.911
NE	2,193	0.858	562	0.899	943	0.872
NH	11,730	0.891	7,354	0.869	9,353	0.883
NJ	25,942	0.884		–	25,918	0.882
NM	11,585	0.896	6,075	0.877	10,888	0.887
NV	34,582	0.906	26,164	0.895	34,163	0.903
NY	718	0.880	586	0.836	712	0.883
OK	387	0.855	–	–	–	–
OR	5,903	0.895	4,952	0.877	6,193	0.891
PA	1,255	0.867	723	0.837	1,240	0.867
RI	1,612	0.868	1,264	0.847	1,731	0.864
SD	12,446	0.873	7,549	0.853	12,393	0.876
TN	42,005	0.879	589	0.814	41,567	0.864
TX	522	0.837	696	0.893	526	0.804
UT	3,159	0.873	1,956	0.860	2,710	0.891
VT	2,182	0.885	1,368	0.854	2,036	0.883
WA	53,326	0.896	32,947	0.877	48,559	0.890
WI	82,306	0.895	59,121	0.878	84,697	0.890
WY	23,229	0.893	4,898	0.871	23,346	0.892

Table C.3. Test-Retest with Alternate Forms Reliability by State—Reading 2–5

	Fall 2016–Winter 2017		Spring 2017–Fall 2017		Winter 2017–Spring 2017	
State	N	Reliability	N	Reliability	N	Reliability
AK	6,922	0.873	6,463	0.851	6,910	0.860
AL	488	0.765	356	0.750	381	0.779
AZ	1,663	0.825	1,268	0.808	1,651	0.822
CA	64,691	0.863	36,396	0.846	46,290	0.850
CO	3,983	0.839	910	0.804	2,529	0.829
CT	29,864	0.845	16,422	0.847	35,550	0.856
DC	4,213	0.786	2,692	0.780	4,540	0.816
DE	2,681	0.754	2,388	0.843	2,390	0.802
FL	15,359	0.796	425	0.890	14,688	0.778
GA	308	0.878	–	–	305	0.876
HI	2,225	0.827	2,349	0.797	2,203	0.825
ID	4,758	0.857	3,837	0.826	4,373	0.854
IL	219,650	0.864	174,817	0.860	260,709	0.857
IN	1,129	0.702	–	–	1,062	0.748
KY	91,270	0.850	65,244	0.846	90,510	0.852
LA	16,810	0.775	360	0.797	15,616	0.786
MD	–	–	391	0.812	–	–
ME	9,689	0.862	18,870	0.856	9,703	0.861
MI	198,986	0.830	165,997	0.828	176,099	0.832
MO	13,770	0.840	–	–	12,472	0.846
MS	30,402	0.814	–	–	24,050	0.829
MT	12,699	0.843	12,711	0.840	12,569	0.833
NC	39,604	0.872	23,014	0.878	37,233	0.875
NE	3,724	0.891	354	0.912	431	0.891
NH	6,802	0.845	5,224	0.853	5,339	0.844
NJ	18,103	0.841	623	0.771	17,792	0.828
NM	13,191	0.843	8,760	0.843	10,792	0.844
NV	23,923	0.851	11,704	0.837	13,496	0.848
NY	489	0.828	346	0.805	492	0.823
OK	360	0.875	–	–	313	0.851
OR	8,593	0.854	5,757	0.847	6,440	0.857
PA	1,159	0.839	950	0.833	1,386	0.845
RI	2,264	0.808	1,842	0.848	2,166	0.805
SD	13,335	0.837	10,583	0.835	12,321	0.834
TN	44,909	0.841	–	–	42,747	0.853
TX	–	–	–	–	395	0.816
UT	4,196	0.830	3,109	0.855	3,667	0.856
VT	2,463	0.817	2,103	0.851	2,255	0.838
WA	35,100	0.861	26,300	0.863	27,157	0.863
WI	77,766	0.865	56,001	0.855	76,430	0.858
WY	10,856	0.841	3,498	0.840	10,745	0.842

Table C.4. Test-Retest with Alternate Forms Reliability by State—Reading 6+

	Fall 2016–Winter 2017		Spring 2017–Fall 2017		Winter 2017–Spring 2017	
State	N	Reliability	N	Reliability	N	Reliability
AZ	496	0.823	520	0.790	637	0.862
CA	22,699	0.870	10,393	0.833	24,275	0.889
CT	11,232	0.893	6,577	0.883	14,134	0.903
DC	5,124	0.886	2,952	0.843	5,137	0.859
DE	2,542	0.861	1,046	0.848	2,750	0.904
FL	10,464	0.850	–	–	10,466	0.862
GA	527	0.904	–	–	545	0.901
HI	312	0.877	–	–	414	0.886
ID	1,411	0.888	1,386	0.852	2,261	0.901
IL	78,283	0.884	44,383	0.860	87,750	0.892
KY	49,683	0.880	26,182	0.822	52,602	0.884
LA	13,845	0.874	–	–	13,886	0.882
ME	5,223	0.877	5,077	0.856	6,968	0.899
MI	122,471	0.884	75,035	0.846	127,060	0.887
MO	9,574	0.894	–	–	9,871	0.904
MS	16,928	0.888	–	–	16,807	0.906
MT	5,006	0.878	3,416	0.845	5,633	0.887
NC	22,559	0.874	8,055	0.836	24,775	0.895
NH	3,771	0.877	2,383	0.861	4,421	0.890
NJ	14,178	0.894	–	–	17,038	0.904
NM	3,580	0.870	3,555	0.861	3,452	0.886
NV	10,896	0.858	5,475	0.833	13,036	0.881
NY	385	0.825	435	0.832	387	0.843
OR	1,728	0.861	1,174	0.793	2,070	0.852
PA	797	0.868	–	–	794	0.899
RI	753	0.911	523	0.885	951	0.912
SD	7,305	0.888	4,524	0.858	7,766	0.899
TN	22,282	0.855	–	–	22,048	0.821
TX	350	0.870	–	–	357	0.894
UT	2,166	0.882	1,149	0.857	2,209	0.892
VT	882	0.846	448	0.842	1,026	0.895
WA	14,908	0.885	10,297	0.879	18,758	0.899
WI	21,243	0.883	11,359	0.845	24,459	0.893
WY	8,972	0.878	1,757	0.847	10,123	0.887

Table C.5. Test-Retest with Alternate Forms Reliability by State—Language Usage Overall

	Fall 2016–Winter 2017		Spring 2017–Fall 2017		Winter 2017–Spring 2017	
State	N	Reliability	N	Reliability	N	Reliability
AK	401	0.822	–	–	366	0.783
AL	771	0.872	659	0.826	678	0.834
AZ	2,292	0.905	2,093	0.908	2,493	0.911
CA	51,493	0.932	27,457	0.930	32,108	0.926
CO	454	0.912	366	0.877	437	0.927
CT	16,072	0.918	9,009	0.910	16,193	0.920
DE	–	–	–	–	577	0.844
FL	–	–	599	0.916	–	–
GA	575	0.914	–	–	547	0.918
HI	–	–	589	0.936	–	–
ID	6,265	0.913	6,916	0.906	5,771	0.910
IL	61,664	0.908	62,633	0.905	62,313	0.907
IN	324	0.786	–	–	–	–
KY	68,179	0.918	47,210	0.905	64,141	0.917
LA	19,787	0.874	–	–	18,736	0.874
MD	428	0.865	369	0.876	418	0.869
ME	3,262	0.896	9,964	0.897	3,412	0.899
MI	184,299	0.905	129,946	0.888	161,281	0.901
MO	14,352	0.907	–	–	11,751	0.908
MS	28,551	0.904	–	–	20,528	0.906
MT	15,335	0.909	20,322	0.901	14,825	0.907
NC	5,254	0.924	2,878	0.930	4,640	0.940
NH	2,136	0.916	1,738	0.900	1,471	0.922
NJ	12,652	0.892	841	0.851	11,296	0.892
NM	14,967	0.915	4,879	0.883	11,831	0.903
NV	7,281	0.922	5,083	0.901	6,354	0.906
OR	3,941	0.900	3,271	0.903	3,460	0.911
PA	1,478	0.910	1,195	0.895	1,677	0.890
RI	–	–	881	0.913	–	–
SD	15,387	0.908	12,634	0.907	13,774	0.907
TN	18,180	0.915	512	0.865	16,295	0.904
TX	–	–	612	0.880	–	–
UT	6,701	0.921	5,102	0.915	5,570	0.926
VT	2,624	0.902	2,595	0.903	2,820	0.894
WA	9,121	0.909	12,135	0.899	8,554	0.905
WI	28,833	0.917	29,874	0.902	29,468	0.908
WY	7,634	0.903	3,919	0.889	7,749	0.905

Table C.6. Test-Retest with Alternate Forms Reliability by State—Mathematics Overall

	Fall 2016–Winter 2017		Spring 2017–Fall 2017		Winter 2017–Spring 2017	
State	N	Reliability	N	Reliability	N	Reliability
AK	7,520	0.943	9,976	0.916	7,297	0.934
AL	1,096	0.960	981	0.922	1,015	0.940
AZ	4,024	0.965	3,963	0.956	4,289	0.961
CA	149,648	0.963	113,016	0.954	123,977	0.957
CO	9,419	0.950	1,930	0.931	7,519	0.936
CT	76,101	0.963	52,802	0.954	87,123	0.956
DC	17,800	0.949	14,029	0.929	17,174	0.933
DE	11,561	0.956	10,215	0.955	11,686	0.953
FL	45,548	0.960	1,263	0.956	44,370	0.948
GA	2,515	0.961	–	–	2,479	0.953
HI	3,788	0.968	3,751	0.960	3,236	0.969
ID	10,842	0.955	10,502	0.959	11,333	0.962
IL	556,718	0.965	518,537	0.952	667,540	0.954
IN	1,319	0.902	–	–	1,281	0.908
KY	256,609	0.968	221,440	0.952	259,765	0.962
LA	47,326	0.954	–	–	46,465	0.949
MA	–	–	–	–	314	0.830
MD	460	0.965	1,081	0.922	464	0.961
ME	30,017	0.956	49,406	0.950	31,779	0.952
MI	521,298	0.959	508,794	0.943	499,523	0.951
MO	40,560	0.959	319	0.936	39,631	0.955
MS	75,235	0.965	–	–	64,168	0.962
MT	34,830	0.960	36,411	0.951	35,344	0.957
NC	132,723	0.970	100,169	0.961	130,792	0.970
NE	5,938	0.942	839	0.920	957	0.914
NH	23,691	0.957	20,351	0.947	20,060	0.954
NJ	71,459	0.955	997	0.863	71,817	0.952
NM	29,412	0.960	23,509	0.947	25,863	0.951
NV	70,511	0.964	60,143	0.948	62,200	0.955
NY	2,368	0.959	2,182	0.941	2,375	0.946
OK	1,400	0.931	–	–	931	0.925
OR	17,326	0.958	14,965	0.949	16,492	0.953
PA	3,235	0.953	2,618	0.926	3,474	0.941
RI	4,733	0.954	4,515	0.948	4,847	0.944
SD	34,374	0.963	30,487	0.952	33,619	0.956
TN	111,485	0.960	1,399	0.919	108,159	0.943
TX	1,018	0.974	1,254	0.934	1,451	0.974
UT	9,628	0.965	7,689	0.956	8,651	0.963
VT	6,032	0.957	5,244	0.946	5,696	0.953
WA	105,678	0.957	87,225	0.948	96,254	0.953
WI	182,671	0.963	166,878	0.950	187,185	0.958
WY	43,651	0.963	13,215	0.956	44,700	0.959

Table C.7. Test-Retest with Alternate Forms Reliability by State—Mathematics K–2

	Fall 2016–Winter 2017		Spring 2017–Fall 2017		Winter 2017–Spring 2017	
State	N	Reliability	N	Reliability	N	Reliability
AK	355	0.910	–	–	308	0.900
AL	318	0.913	–	–	309	0.923
AZ	1,673	0.905	1,427	0.881	1,863	0.910
CA	61,969	0.933	39,690	0.931	52,407	0.939
CO	4,398	0.923	471	0.905	4,316	0.936
CT	28,557	0.919	16,097	0.909	31,307	0.921
DC	5,182	0.894	3,255	0.892	5,007	0.893
DE	5,839	0.935	3,574	0.919	5,924	0.934
FL	19,936	0.920	403	0.924	19,627	0.920
GA	319	0.926	–	–	305	0.918
HI	1,550	0.937	814	0.923	937	0.937
ID	3,714	0.906	2,847	0.904	3,424	0.922
IL	242,445	0.930	184,863	0.915	306,586	0.920
KY	112,699	0.928	80,613	0.903	114,422	0.929
LA	17,064	0.893	–	–	17,389	0.904
MD	–	–	334	0.897	–	–
ME	13,732	0.912	15,353	0.901	13,978	0.914
MI	194,461	0.912	153,880	0.895	188,574	0.912
MO	17,220	0.913	–	–	16,738	0.915
MS	28,215	0.918	–	–	23,822	0.923
MT	15,891	0.910	12,755	0.894	16,058	0.920
NC	61,276	0.937	39,062	0.928	60,964	0.942
NE	2,191	0.907	556	0.908	856	0.910
NH	11,868	0.909	7,405	0.885	9,993	0.915
NJ	29,600	0.924	–	–	29,259	0.927
NM	11,309	0.914	6,350	0.891	10,579	0.911
NV	34,715	0.933	26,557	0.922	34,033	0.932
NY	716	0.914	598	0.886	718	0.919
OK	383	0.885	–	–	–	–
OR	6,209	0.914	4,743	0.900	6,592	0.917
PA	1,245	0.921	730	0.895	1,236	0.914
RI	1,690	0.911	1,314	0.881	1,734	0.907
SD	12,382	0.916	7,523	0.904	12,134	0.918
TN	42,814	0.915	620	0.899	42,214	0.901
TX	460	0.877	683	0.926	527	0.910
UT	3,224	0.907	1,959	0.901	2,766	0.930
VT	2,343	0.911	1,549	0.884	2,174	0.907
WA	54,118	0.922	32,878	0.907	48,047	0.921
WI	81,603	0.922	60,559	0.907	83,412	0.925
WY	23,720	0.924	4,869	0.904	23,782	0.927

Table C.8. Test-Retest with Alternate Forms Reliability by State—Mathematics 2–5

	Fall 2016–Winter 2017		Spring 2017–Fall 2017		Winter 2017–Spring 2017	
State	N	Reliability	N	Reliability	N	Reliability
AK	6,910	0.930	6,682	0.919	6,752	0.923
AL	503	0.884	409	0.862	432	0.871
AZ	1,564	0.897	1,240	0.909	1,526	0.909
CA	64,757	0.919	37,268	0.919	47,198	0.912
CO	4,758	0.918	1,076	0.903	2,928	0.903
CT	32,358	0.920	18,489	0.918	38,552	0.923
DC	7,318	0.851	5,143	0.864	6,898	0.864
DE	2,644	0.855	2,323	0.919	2,377	0.887
FL	15,196	0.868	541	0.940	14,348	0.834
GA	1,638	0.921	–	–	1,626	0.921
HI	1,804	0.898	2,352	0.908	1,767	0.895
ID	5,594	0.912	4,362	0.912	5,413	0.915
IL	225,359	0.924	171,387	0.926	261,840	0.915
IN	1,105	0.819	–	–	1,079	0.861
KY	93,158	0.917	66,293	0.914	92,115	0.916
LA	16,260	0.860	–	–	14,878	0.871
MA	–	–	–	–	314	0.830
MD	–	–	449	0.893	–	–
ME	11,055	0.913	19,464	0.923	11,299	0.917
MI	200,508	0.904	166,009	0.908	179,343	0.904
MO	13,134	0.909	–	–	12,413	0.906
MS	29,500	0.894	–	–	23,044	0.899
MT	13,865	0.920	13,207	0.927	13,823	0.918
NC	41,235	0.926	22,897	0.932	37,848	0.934
NE	3,747	0.930	–	–	–	–
NH	7,950	0.912	6,028	0.914	5,509	0.898
NJ	26,605	0.879	743	0.844	25,059	0.887
NM	13,756	0.907	8,467	0.899	11,188	0.900
NV	23,382	0.922	11,865	0.911	14,331	0.909
NY	490	0.905	315	0.888	494	0.921
OK	884	0.895	–	–	872	0.929
OR	8,740	0.907	6,105	0.909	7,079	0.910
PA	1,193	0.879	971	0.902	1,445	0.888
RI	2,011	0.856	1,722	0.899	1,905	0.862
SD	14,383	0.910	11,435	0.919	13,463	0.912
TN	46,088	0.897	–	–	43,760	0.897
TX	–	–	–	–	559	0.917
UT	4,219	0.903	3,014	0.921	3,673	0.915
VT	2,723	0.908	2,120	0.908	2,395	0.916
WA	34,615	0.909	24,736	0.917	26,658	0.914
WI	77,497	0.928	56,018	0.930	76,360	0.926
WY	10,971	0.905	3,817	0.915	10,686	0.910

Table C.9. Test-Retest with Alternate Forms Reliability by State—Mathematics 6+

	Fall 2016–Winter 2017		Spring 2017–Fall 2017		Winter 2017–Spring 2017	
State	N	Reliability	N	Reliability	N	Reliability
AZ	751	0.868	509	0.876	888	0.907
CA	22,617	0.888	10,641	0.845	24,174	0.902
CT	14,338	0.919	8,056	0.891	16,896	0.910
DC	5,199	0.903	2,904	0.847	5,210	0.883
DE	3,066	0.888	1,566	0.861	3,352	0.905
FL	10,383	0.864	–	–	10,387	0.884
GA	556	0.930	–	–	546	0.905
HI	424	0.867	–	–	527	0.918
ID	1,445	0.901	1,473	0.891	2,451	0.921
IL	86,020	0.901	48,599	0.874	96,543	0.900
KY	50,073	0.899	25,944	0.843	52,422	0.896
LA	13,774	0.893	–	–	13,808	0.900
ME	4,989	0.902	4,837	0.881	6,321	0.907
MI	122,799	0.903	74,683	0.868	127,368	0.904
MO	9,403	0.903	–	–	9,827	0.913
MS	17,190	0.909	–	–	17,178	0.921
MT	4,720	0.884	3,187	0.864	5,210	0.902
NC	29,759	0.899	14,443	0.860	31,489	0.914
NH	3,723	0.877	2,527	0.860	4,488	0.906
NJ	14,600	0.900	–	–	17,065	0.907
NM	4,191	0.898	3,810	0.874	3,952	0.903
NV	12,120	0.868	5,266	0.861	13,686	0.900
NY	1,160	0.913	903	0.887	1,162	0.901
OR	2,154	0.879	1,424	0.849	2,616	0.885
PA	778	0.886	–	–	773	0.912
RI	1,029	0.929	670	0.892	1,207	0.922
SD	7,352	0.907	4,560	0.881	7,803	0.916
TN	22,213	0.882	–	–	22,012	0.838
TX	342	0.892	–	–	365	0.889
UT	2,157	0.915	1,284	0.894	2,174	0.908
VT	903	0.888	568	0.860	1,102	0.894
WA	16,219	0.901	11,291	0.892	20,125	0.912
WI	22,830	0.903	13,544	0.866	26,537	0.912
WY	8,924	0.889	1,673	0.866	10,209	0.907

Table C.10. Test-Retest with Alternate Forms Reliability by State—Science Overall

	Fall 2016–Winter 2017		Spring 2017–Fall 2017		Winter 2017–Spring 2017	
State	N	Reliability	N	Reliability	N	Reliability
AR	8,427	0.873	6,622	0.857	8,970	0.876
CA	8,552	0.853	4,926	0.847	9,020	0.860
CO	7,887	0.847	5,804	0.836	7,845	0.855
CT	2,577	0.873	3,066	0.864	3,150	0.867
IA	1,008	0.800	2,635	0.846	690	0.822
IL	15,852	0.880	11,981	0.874	17,653	0.879
KS	2,186	0.865	2,103	0.854	1,146	0.868
KY	3,938	0.873	3,373	0.880	4,573	0.876
MA	1,061	0.857	–	–	634	0.844
MD	–	–	455	0.889	–	–
MI	65,572	0.866	48,323	0.860	56,407	0.867
MO	1,308	0.841	–	–	1,416	0.837
MT	409	0.871	–	–	405	0.861
NJ	1,473	0.849	855	0.849	1,373	0.823
NV	565	0.843	375	0.814	558	0.844
OH	–	–	1,881	0.827	–	–
OK	520	0.781	–	–	534	0.850
RI	–	–	694	0.863	–	–
SD	734	0.809	489	0.815	733	0.851
WA	2,538	0.848	2,337	0.843	2,245	0.877
WI	514	0.858	1,249	0.838	560	0.863

Table C.11. Test-Retest with Alternate Forms Reliability by State—Science 3–5

	Fall 2016–Winter 2017		Spring 2017–Fall 2017		Winter 2017–Spring 2017	
State	N	Reliability	N	Reliability	N	Reliability
AR	3,744	0.843	2,106	0.817	3,941	0.857
CA	3,617	0.802	406	0.790	3,328	0.807
CO	1,639	0.761	691	0.799	1,682	0.811
CT	378	0.829	405	0.755	517	0.802
IA	–	–	662	0.819	–	–
IL	6,973	0.856	3,861	0.853	8,488	0.856
KS	387	0.831	–	–	320	0.829
KY	1,302	0.846	1,400	0.827	1,526	0.836
MA	719	0.799	–	–	489	0.798
MI	29,685	0.830	15,606	0.825	23,910	0.838
NJ	668	0.800	–	–	638	0.775
OH	–	–	640	0.782	–	–
WA	469	0.854	618	0.835	713	0.852
WI	–	–	309	0.804	–	–

Table C.12. Test-Retest with Alternate Forms Reliability by State—Science 6+

State	Fall 2016–Winter 2017		Spring 2017–Fall 2017		Winter 2017–Spring 2017	
	N	Reliability	N	Reliability	N	Reliability
AR	4,608	0.836	3,247	0.828	5,021	0.844
CA	4,933	0.823	4,097	0.834	5,674	0.838
CO	6,244	0.839	4,397	0.823	6,161	0.843
CT	2,190	0.861	2,154	0.851	2,548	0.861
IA	871	0.803	1,676	0.833	607	0.824
IL	8,829	0.851	5,975	0.855	9,120	0.861
KS	1,795	0.850	1,605	0.853	823	0.867
KY	2,632	0.819	1,528	0.835	3,039	0.837
MA	341	0.867	–	–	–	–
MD	–	–	354	0.875	–	–
MI	35,756	0.835	24,239	0.838	32,389	0.842
MO	1,211	0.841	–	–	1,160	0.838
NJ	802	0.806	524	0.813	734	0.798
NV	348	0.825	–	–	333	0.817
OH	–	–	833	0.796	–	–
OK	369	0.796	–	–	377	0.850
SD	731	0.809	488	0.815	732	0.852
WA	2,065	0.832	1,242	0.802	1,531	0.844
WI	368	0.829	660	0.835	396	0.833

Table C.13. Test-Retest with Alternate Forms Reliability by State and Grade—Reading, Spring 2017–Fall 2017

Reading, Spring 2017–Fall 2017													
State		Grade											
		K	1	2	3	4	5	6	7	8	9	10	11
AK	Reliability	–	–	–	–	–	0.869	0.857	0.848	0.659	–	–	–
	N	–	–	–	–	–	2,967	2,969	2,850	383	–	–	–
AZ	Reliability	0.700	0.692	0.808	0.808	0.820	0.842	0.864	0.847	–	–	–	–
	N	375	395	422	506	466	431	386	397	–	–	–	–
CA	Reliability	0.817	0.817	0.876	0.877	0.882	0.875	0.860	0.865	0.807	0.830	0.827	0.783
	N	9,327	11,606	14,223	12,323	12,741	12,156	10,385	10,433	5,855	6,011	2,855	783
CT	Reliability	0.801	0.810	0.832	0.842	0.846	0.845	0.841	0.846	0.832	0.857	–	–
	N	3,751	4,639	5,647	5,244	6,305	5,595	5,986	5,141	2,525	2,085	–	–
DC	Reliability	0.753	0.787	0.770	0.819	0.801	0.781	0.787	0.798	0.758	0.770	–	–
	N	1,738	1,680	1,611	1,354	1,267	734	889	800	515	337	–	–
DE	Reliability	0.834	0.797	0.833	0.832	0.858	0.842	0.829	0.826	–	0.814	0.836	
	N	565	1,555	1,382	1,210	1,118	1,353	545	584	–	486	340	
HI	Reliability	–	–	–	–	0.818	0.867	0.771	0.744	0.844	0.828	–	–
	N	–	–	–	–	334	316	435	631	590	340	–	–
ID	Reliability	0.779	0.813	0.832	0.844	0.845	0.872	0.863	0.843	0.855	0.791	0.728	–
	N	754	897	938	1,103	1,192	1,007	1,107	1,177	458	567	466	–
IL	Reliability	0.822	0.804	0.867	0.873	0.872	0.864	0.863	0.867	0.843	0.847	0.860	0.831
	N	31,988	40,681	62,579	66,132	67,276	68,904	65,782	68,266	18,278	13,601	5,753	1,849
KY	Reliability	0.789	0.768	0.850	0.841	0.848	0.835	0.847	0.843	0.848	0.841	0.814	
	N	20,446	22,349	25,697	27,594	27,912	26,756	22,550	23,315	9,946	7,370	1,262	–
ME	Reliability	0.755	0.808	0.823	0.871	0.870	0.870	0.860	0.865	0.841	0.830	0.858	0.836
	N	2,325	3,239	5,163	6,000	6,115	5,666	6,561	6,569	3,393	1,976	613	309
MI	Reliability	0.777	0.783	0.819	0.850	0.850	0.840	0.837	0.829	0.822	0.829	0.805	0.793
	N	45,084	50,888	56,382	59,667	61,972	59,959	56,255	52,556	23,867	19,707	8,394	2,747
MT	Reliability	0.768	0.779	0.804	0.835	0.848	0.837	0.843	0.851	0.824	0.826	0.848	0.807
	N	2,189	2,542	3,431	5,097	4,962	5,044	3,983	4,028	1,756	1,836	837	304
NC	Reliability	0.827	0.803	0.875	0.879	0.879	0.873	0.881	0.869	0.878	0.885	0.891	–
	N	7,066	8,897	12,599	13,302	13,076	12,387	11,155	10,254	528	509	318	–

Reading, Spring 2017–Fall 2017													
State		Grade											
		K	1	2	3	4	5	6	7	8	9	10	11
NE	Reliability	–	–	–	–	0.888	–	–	–	–	–	–	–
	N	–	–	–	–	309	–	–	–	–	–	–	–
NH	Reliability	0.760	0.759	0.826	0.845	0.831	0.842	0.858	0.845	0.847	0.861	–	–
	N	1,291	2,047	3,025	2,664	2,425	2,550	2,061	2,071	403	378	–	–
NM	Reliability	0.741	0.793	0.808	0.850	0.862	0.845	0.871	0.855	0.810	0.823	0.827	0.785
	N	1,887	2,118	2,368	2,561	2,553	2,624	2,547	2,798	843	826	789	555
NV	Reliability	0.802	0.773	0.866	0.877	0.876	0.866	0.846	0.842	0.803	0.816	–	–
	N	4,434	7,942	8,356	9,285	8,904	7,576	5,572	3,643	1,412	543	–	–
OR	Reliability	0.714	0.762	0.857	0.858	0.849	0.844	0.858	0.837	0.821	0.839	0.840	
	N	881	1,165	1,811	1,646	1,766	1,468	1,757	1,747	906	932	327	
PA	Reliability	–	0.778	0.799	0.818	0.822	0.857	0.817	0.847	–	–	–	–
	N	–	303	300	306	339	340	356	355	–	–	–	–
RI	Reliability	0.779	0.743	0.789	0.796	0.841	0.837	0.862	0.817	–	0.872	–	–
	N	340	308	438	475	521	561	555	490	–	315	–	–
SD	Reliability	0.790	0.765	0.819	0.828	0.858	0.850	0.856	0.833	0.823	0.820	0.846	0.791
	N	2,666	2,753	2,840	3,121	3,162	4,259	2,533	2,427	1,893	1,680	1,332	526
TX	Reliability	–	–	0.888	–	–	–	–	–	–	–	–	–
	N	–	–	324	–	–	–	–	–	–	–	–	–
UT	Reliability	0.817	0.738	0.841	0.845	0.832	0.828	0.847	0.851	0.839	0.862	0.836	–
	N	886	819	827	695	738	654	701	724	565	563	481	–
VT	Reliability	–	–	0.814	0.844	0.826	0.846	0.848	0.865	0.837	0.836	–	–
	N	–	–	400	571	563	629	553	609	343	440	–	–
WA	Reliability	0.815	0.808	0.844	0.861	0.863	0.864	0.860	0.861	0.860	0.869	0.869	0.851
	N	6,043	8,596	11,378	12,166	12,182	10,842	9,530	9,909	3,761	1,908	721	380
WI	Reliability	0.778	0.779	0.842	0.858	0.860	0.850	0.860	0.855	0.843	0.837	0.861	0.836
	N	7,454	12,510	17,702	22,220	22,903	22,176	22,208	21,605	6,595	4,260	829	379
WY	Reliability	0.801	0.731	0.832	0.842	0.861	0.843	0.851	0.852	0.843	0.791	–	–
	N	1,424	1,492	1,431	1,694	1,817	1,574	1,152	1,039	513	463	–	–

Table C.14. Test-Retest with Alternate Forms Reliability by State and Grade—Reading, Winter 2017–Spring 2017

State		Reading, Winter 2017–Spring 2017												
		Grade												
		K	1	2	3	4	5	6	7	8	9	10	11	12
AK	Reliability	–	–	–	–	–	–	0.882	0.850	0.848	–	–	–	–
	N	–	–	–	–	–	–	950	2,829	2,746	–	–	–	–
AZ	Reliability	0.679	–	0.786	0.807	0.831	0.849	0.854	0.843	0.848	–	–	–	–
	N	364	–	448	485	439	448	426	337	313	–	–	–	–
CA	Reliability	0.775	0.869	0.888	0.885	0.883	0.883	0.862	0.865	0.846	0.830	0.825	0.794	0.745
	N	10,306	12,376	14,787	12,394	12,812	12,831	10,017	9,954	8,593	7,948	6,675	2,488	566
CO	Reliability	–	0.819	0.852	0.851	0.837	0.845	0.869	0.846	0.859	–	–	–	–
	N	–	302	986	1,041	1,072	1,043	781	621	570	–	–	–	–
CT	Reliability	0.780	0.859	0.876	0.853	0.859	0.866	0.865	0.855	0.859	0.851	0.836	0.806	–
	N	4,375	6,366	7,608	7,541	8,568	8,687	8,898	8,332	8,442	4,900	3,826	839	–
DC	Reliability	0.683	0.827	0.827	0.798	0.816	0.826	0.834	0.824	0.819	0.791	–	–	–
	N	2,135	1,965	1,884	1,625	1,405	1,195	1,353	1,209	1,025	543	–	–	–
DE	Reliability	0.737	0.872	0.855	0.867	0.864	0.864	0.784	0.778	0.833	0.827	0.805	–	–
	N	662	1,614	1,584	1,536	1,453	1,496	498	392	371	418	400	–	–
FL	Reliability	0.742	0.851	0.850	0.824	0.802	0.794	0.800	0.767	0.741	0.789	0.781	–	–
	N	5,223	5,197	5,172	5,209	4,723	4,660	5,047	4,261	3,890	718	656	–	–
HI	Reliability	–	–	–	–	–	–	0.732	0.751	0.860	0.841	–	–	–
	N	–	–	–	–	–	–	396	597	577	304	–	–	–
ID	Reliability	0.753	0.834	0.854	0.821	0.855	0.846	0.838	0.845	0.860	0.859	0.833	–	–
	N	772	1,084	992	907	1,008	998	1,089	1,132	1,152	496	399	–	–
IL	Reliability	0.778	0.866	0.872	0.869	0.866	0.865	0.861	0.862	0.853	0.842	0.829	0.814	0.814
	N	33,644	43,931	72,448	82,553	83,494	82,250	78,547	78,033	73,165	14,943	10,610	4,404	1,325
KY	Reliability	0.767	0.857	0.870	0.858	0.864	0.861	0.849	0.852	0.855	0.850	0.830	0.761	–
	N	24,269	26,358	28,729	30,483	29,501	28,032	24,267	25,379	24,036	9,098	5,771	1,694	–
LA	Reliability	0.734	0.845	0.858	0.832	0.826	0.816	0.810	0.785	0.798	0.792	0.721	0.664	–
	N	5,579	6,024	6,097	5,025	4,548	4,131	3,868	3,550	3,280	2,614	1,838	327	–
ME	Reliability	0.737	0.849	0.868	0.869	0.873	0.869	0.857	0.864	0.860	0.841	0.849	–	–
	N	1,736	2,865	3,992	4,333	4,167	3,769	3,123	2,896	2,739	601	326	–	–

Reading, Winter 2017–Spring 2017														
State		Grade												
		K	1	2	3	4	5	6	7	8	9	10	11	12
MI	Reliability	0.733	0.849	0.861	0.853	0.856	0.858	0.847	0.840	0.837	0.837	0.813	0.777	0.763
	N	48,042	52,961	55,993	52,430	54,356	53,992	47,572	42,479	40,492	18,587	17,312	8,000	1,733
MO	Reliability	0.776	0.859	0.870	0.854	0.865	0.865	0.854	0.861	0.845	0.830	0.839	0.673	–
	N	3,350	4,075	5,502	4,851	5,221	4,295	3,906	3,095	3,179	986	800	370	–
MS	Reliability	0.792	0.860	0.850	0.835	0.837	0.821	0.840	0.826	0.834	0.809	0.804	0.765	–
	N	7,069	8,494	8,532	5,554	5,786	5,087	5,661	6,148	5,808	3,117	2,588	728	–
MT	Reliability	0.765	0.859	0.844	0.844	0.847	0.855	0.846	0.844	0.823	0.823	0.796	–	–
	N	2,298	2,517	3,170	4,627	4,557	4,351	3,968	3,052	2,938	679	1,736	–	–
NC	Reliability	0.810	0.883	0.884	0.883	0.883	0.882	0.880	0.871	0.874	0.856	0.867	0.869	–
	N	10,364	14,241	14,834	15,772	15,325	15,002	12,146	11,622	11,733	718	516	404	–
NE	Reliability	–	–	–	–	0.862	0.845	–	–	–	–	–	–	–
	N	–	–	–	–	317	361	–	–	–	–	–	–	–
NH	Reliability	0.757	0.833	0.868	0.854	0.829	0.839	0.855	0.836	0.842	–	–	–	–
	N	940	2,509	2,685	2,787	2,389	2,478	1,883	1,591	1,293	–	–	–	–
NJ	Reliability	0.726	0.839	0.866	0.851	0.849	0.851	0.827	0.839	0.837	0.805	0.807	0.734	–
	N	5,431	7,017	8,345	7,427	7,447	7,416	7,040	4,943	4,209	705	565	330	–
NM	Reliability	0.718	0.814	0.858	0.859	0.848	0.854	0.849	0.854	0.838	0.801	0.764	0.819	0.833
	N	1,274	1,518	2,734	2,921	3,024	2,964	3,148	2,236	2,015	1,234	986	740	365
NV	Reliability	0.765	0.850	0.868	0.878	0.878	0.872	0.867	0.843	0.836	0.805	0.807	0.782	–
	N	4,580	7,860	8,301	9,531	8,930	8,136	5,820	3,408	2,875	495	378	303	–
OR	Reliability	0.696	0.825	0.852	0.855	0.857	0.874	0.866	0.838	0.850	0.858	0.840	–	–
	N	682	1,128	1,807	1,615	1,771	1,431	1,694	1,713	1,453	734	637	–	–
PA	Reliability	–	0.860	0.831	0.811	0.837	0.850	0.869	0.817	0.849	–	–	–	–
	N	–	407	358	362	383	364	471	445	340	–	–	–	–
RI	Reliability	0.784	0.837	0.845	0.840	0.818	0.817	0.844	0.811	0.765	0.777	–	–	–
	N	387	389	504	489	414	501	489	602	353	425	–	–	–
SD	Reliability	0.755	0.844	0.872	0.848	0.852	0.855	0.847	0.845	0.841	0.803	0.832	0.837	–
	N	2,877	3,046	3,024	3,351	3,354	4,557	2,836	2,636	2,411	1,599	1,439	1,114	–

Reading, Winter 2017–Spring 2017														
State		Grade												
		K	1	2	3	4	5	6	7	8	9	10	11	12
TN	Reliability	0.670	0.815	0.810	0.833	0.846	0.850	0.856	0.862	0.858	0.860	0.854	0.762	0.648
	N	11,164	10,597	10,579	10,803	9,951	10,807	9,175	9,092	8,809	6,362	5,811	2,720	493
TX	Reliability	–	–	–	–	–	0.801	–	–	–	–	–	–	–
	N	–	–	–	–	–	349	–	–	–	–	–	–	–
UT	Reliability	0.769	0.849	0.860	0.870	0.848	0.874	0.857	0.847	0.866	0.861	0.818	–	–
	N	932	943	978	712	736	642	791	821	699	583	556	–	–
VT	Reliability	0.685	0.849	0.865	0.875	0.854	0.854	0.834	0.823	0.855	–	0.847	–	–
	N	374	384	484	636	550	628	613	509	497	–	310	–	–
WA	Reliability	0.803	0.858	0.869	0.863	0.872	0.871	0.868	0.862	0.859	0.856	0.829	0.820	–
	N	6,601	8,448	12,657	13,942	13,140	13,137	8,263	7,787	7,612	1,953	910	468	–
WI	Reliability	0.762	0.849	0.868	0.863	0.859	0.859	0.863	0.861	0.856	0.833	0.829	0.838	–
	N	8,674	11,904	18,222	23,250	24,027	23,561	23,220	22,491	21,432	4,944	3,362	823	–
WY	Reliability	0.760	0.843	0.846	0.842	0.853	0.861	0.845	0.855	0.833	0.847	0.792	–	–
	N	4,238	5,795	6,088	6,048	5,787	5,699	3,746	2,983	2,906	556	343	–	–

Table C.15. Test-Retest with Alternate Forms Reliability by State and Grade—Reading, Fall 2016–Winter 2017

Reading, Fall 2016–Winter 2017														
State		Grade												
		K	1	2	3	4	5	6	7	8	9	10	11	12
AK	Reliability	–	–	–	–	–	–	0.898	0.864	0.858	–	–	–	–
	N	–	–	–	–	–	–	920	2,759	2,828	–	–	–	–
AZ	Reliability	–	–	0.780	0.795	0.820	0.777	0.811	0.834	0.842	–	–	–	–
	N	–	–	398	444	396	392	409	342	324	–	–	–	–
CA	Reliability	0.675	0.841	0.866	0.874	0.878	0.879	0.874	0.870	0.864	0.842	0.819	0.812	0.762
	N	8,863	12,336	14,839	15,907	16,133	16,531	15,244	15,196	14,705	9,415	6,410	2,846	828
CO	Reliability	–	–	0.816	0.843	0.837	0.858	0.849	0.885	0.842	0.835	0.817	–	–
	N	–	–	1,064	1,119	1,138	1,100	983	804	816	673	588	–	–
CT	Reliability	0.684	0.823	0.844	0.845	0.854	0.859	0.856	0.829	0.850	0.835	0.811	0.825	–
	N	2,604	6,111	6,535	6,884	7,728	7,564	7,795	7,218	7,389	3,608	2,832	773	–

Reading, Fall 2016–Winter 2017														
State		**Grade**												
		K	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
DC	**Reliability**	0.666	0.808	0.816	0.800	0.811	0.788	0.808	0.803	0.816	0.773	0.723	–	–
	N	2,146	1,926	1,876	1,714	1,507	1,340	1,125	1,007	769	539	385	–	–
DE	**Reliability**	0.731	0.783	0.860	0.859	0.857	0.857	0.777	0.703	0.800	0.787	0.717	–	–
	N	613	1,543	1,503	1,447	1,420	1,539	594	514	447	406	406	–	–
FL	**Reliability**	0.676	0.804	0.853	0.826	0.794	0.802	0.785	0.785	0.789	0.770	0.770	–	–
	N	5,199	5,218	5,200	5,249	4,830	4,745	5,143	4,435	4,031	759	731	–	–
HI	**Reliability**	–	–	–	–	0.839	0.874	0.811	0.734	0.840	–	–	–	–
	N	–	–	–	–	395	430	438	593	579	–	–	–	–
ID	**Reliability**	0.697	0.773	0.831	0.813	0.841	0.862	0.851	0.832	0.851	0.866	0.821	–	–
	N	429	627	889	1,028	1,104	1,168	1,210	1,118	1,197	592	484	–	–
IL	**Reliability**	0.711	0.830	0.867	0.870	0.873	0.875	0.869	0.868	0.865	0.833	0.831	0.835	0.849
	N	27,356	39,683	59,605	65,087	66,042	64,271	62,584	61,199	59,485	16,281	11,738	6,691	1,958
KY	**Reliability**	0.692	0.836	0.859	0.856	0.861	0.856	0.852	0.843	0.846	0.849	0.844	0.792	–
	N	21,706	25,906	28,823	30,027	28,915	27,643	24,250	24,773	24,124	9,407	6,409	1,950	–
LA	**Reliability**	0.649	0.803	0.831	0.813	0.812	0.810	0.790	0.765	0.798	0.742	0.737	0.766	–
	N	5,559	5,954	6,076	4,647	4,321	4,183	4,107	3,844	3,593	2,706	2,029	363	–
ME	**Reliability**	0.614	0.796	0.838	0.853	0.874	0.873	0.861	0.857	0.859	0.846	0.838	–	–
	N	905	2,357	3,405	4,249	4,165	3,771	2,950	2,952	2,885	475	360	–	–
MI	**Reliability**	0.666	0.814	0.848	0.847	0.853	0.852	0.841	0.837	0.830	0.830	0.813	0.777	0.751
	N	43,148	51,866	55,491	54,337	56,562	55,846	50,632	47,092	45,207	22,303	20,971	9,895	2,790
MO	**Reliability**	0.701	0.827	0.851	0.848	0.856	0.836	0.841	0.861	0.834	0.808	0.794	0.796	–
	N	2,877	3,962	5,358	5,132	5,528	4,604	4,033	3,355	3,271	1,186	1,102	617	–
MS	**Reliability**	0.654	0.801	0.818	0.813	0.806	0.807	0.833	0.814	0.819	0.791	0.795	0.741	–
	N	7,006	8,524	8,530	7,097	7,371	6,475	7,371	7,928	7,627	3,293	3,299	739	–
MT	**Reliability**	0.651	0.822	0.826	0.829	0.839	0.853	0.844	0.854	0.833	0.836	0.795	–	–
	N	1,847	2,385	2,965	4,535	4,548	4,318	3,992	3,108	3,031	624	1,703	–	–
NC	**Reliability**	0.712	0.849	0.871	0.869	0.876	0.878	0.878	0.872	0.865	0.832	0.862	0.857	–
	N	8,095	13,941	14,765	15,763	15,528	15,139	13,048	12,674	12,243	627	506	427	–

Reading, Fall 2016–Winter 2017														
State		Grade												
		K	1	2	3	4	5	6	7	8	9	10	11	12
NE	**Reliability**	–	–	–	0.821	0.839	0.844	0.854	0.860	0.878	0.921	0.920	0.871	–
	N	–	–	–	781	702	710	706	651	742	585	540	499	–
NH	**Reliability**	0.649	0.788	0.846	0.849	0.841	0.849	0.859	0.846	0.832	–	0.821	–	–
	N	714	2,080	2,963	3,456	3,086	3,222	1,995	1,950	1,935	–	347	–	–
NJ	**Reliability**	0.660	0.802	0.848	0.834	0.852	0.855	0.839	0.844	0.853	0.786	0.777	0.731	0.690
	N	3,412	6,391	7,908	7,540	7,777	7,400	6,989	4,799	4,841	571	461	340	300
NM	**Reliability**	0.620	0.734	0.843	0.854	0.856	0.869	0.849	0.851	0.845	0.796	0.792	0.808	0.808
	N	1,214	1,563	2,777	3,179	3,239	3,205	3,571	2,666	2,560	1,587	1,245	931	463
NV	**Reliability**	0.680	0.806	0.854	0.865	0.870	0.879	0.862	0.866	0.856	0.815	0.751	0.765	0.703
	N	3,222	7,106	8,086	9,417	9,243	8,631	7,127	6,475	6,325	1,848	982	894	339
OR	**Reliability**	0.648	0.832	0.836	0.858	0.857	0.869	0.866	0.838	0.838	0.843	0.849	0.838	–
	N	436	1,084	1,338	1,396	1,916	1,627	1,977	1,991	1,960	1,139	915	473	–
PA	**Reliability**	–	0.766	0.806	0.823	0.783	0.850	0.863	0.859	0.832	–	–	–	–
	N	–	405	363	367	387	370	355	358	321	–	–	–	–
RI	**Reliability**	–	0.819	0.840	0.834	0.840	0.819	0.832	0.852	0.787	0.819	0.762	–	–
	N	–	362	410	465	398	490	467	544	377	441	313	–	–
SD	**Reliability**	0.703	0.803	0.830	0.824	0.847	0.848	0.835	0.845	0.839	0.811	0.843	0.855	0.751
	N	2,551	2,924	2,951	3,369	3,264	4,804	2,885	2,710	2,600	1,686	1,640	1,297	536
TN	**Reliability**	0.657	0.820	0.827	0.847	0.847	0.853	0.842	0.848	0.844	0.853	0.850	0.759	0.669
	N	11,011	10,738	10,755	11,006	10,082	10,984	9,485	9,070	9,025	6,520	5,916	2,978	1,526
TX	**Reliability**	–	–	–	–	–	0.844	–	–	–	–	–	–	–
	N	–	–	–	–	–	351	–	–	–	–	–	–	–
UT	**Reliability**	0.767	0.800	0.832	0.835	0.828	0.844	0.841	0.832	0.819	0.812	0.807	0.787	–
	N	897	930	949	848	923	802	890	874	783	577	539	517	–
VT	**Reliability**	–	0.763	0.833	0.848	0.860	0.853	0.798	0.848	0.840	–	–	–	–
	N	–	380	456	679	626	680	688	552	569	–	–	–	–
WA	**Reliability**	0.755	0.817	0.858	0.859	0.867	0.867	0.862	0.868	0.858	0.831	0.825	0.822	0.779
	N	3,530	7,785	12,152	15,735	14,711	14,848	10,276	10,247	10,174	2,250	1,347	527	340

Reading, Fall 2016–Winter 2017														
State		Grade												
		K	1	2	3	4	5	6	7	8	9	10	11	12
WI	Reliability	0.671	0.821	0.856	0.859	0.862	0.861	0.864	0.864	0.861	0.858	0.839	0.837	0.876
	N	7,031	10,209	17,341	22,752	23,469	23,104	23,203	22,701	21,371	5,076	3,780	1,090	530
WY	Reliability	0.700	0.814	0.828	0.832	0.849	0.852	0.842	0.843	0.837	0.850	0.786	–	–
	N	2,950	5,783	6,066	6,017	5,782	5,680	3,748	3,014	2,918	563	350	–	–

Table C.16. Test-Retest with Alternate Forms Reliability by State and Grade—Language Usage, Spring 2017–Fall 2017

Language Usage, Spring 2017–Fall 2017											
State		Grade									
		2	3	4	5	6	7	8	9	10	11
AZ	Reliability	–	0.816	0.823	–	–	–	–	–	–	–
	N	–	353	337	–	–	–	–	–	–	–
CA	Reliability	0.898	0.901	0.897	0.900	0.910	0.910	–	0.859	–	–
	N	6,408	5,420	6,093	3,413	2,589	2,221	–	723	–	–
CT	Reliability	0.853	0.869	0.871	0.858	0.879	0.866	0.855	0.881	–	–
	N	707	550	1,423	1,136	1,822	1,944	595	583	–	–
ID	Reliability	0.849	0.864	0.841	0.865	0.879	0.884	0.877	0.845	0.847	–
	N	591	948	993	898	871	892	451	743	455	–
IL	Reliability	0.862	0.867	0.865	0.876	0.877	0.891	0.847	0.864	0.878	0.856
	N	5,293	8,587	9,103	9,443	11,116	11,441	1,955	3,139	1,632	319
KY	Reliability	0.864	0.851	0.864	0.851	0.863	0.873	0.868	0.853	0.855	–
	N	4,978	7,970	9,379	7,291	7,345	7,149	1,003	1,151	551	–
ME	Reliability	0.809	0.841	0.851	0.845	0.847	0.879	0.869	0.840	–	–
	N	692	1,224	1,319	1,388	1,688	1,672	588	783	–	–
MI	Reliability	0.853	0.845	0.844	0.850	0.852	0.847	0.846	0.846	0.838	0.837
	N	8,921	17,953	19,380	18,491	20,848	20,635	8,363	9,466	4,031	907
MT	Reliability	0.814	0.840	0.855	0.862	0.867	0.872	0.858	0.870	0.875	–
	N	917	3,097	3,146	3,048	3,203	3,401	1,536	1,250	576	–
NC	Reliability	0.865	0.882	0.874	0.871	0.879	0.890	–	–	–	–
	N	340	429	402	411	500	338	–	–	–	–

Language Usage, Spring 2017–Fall 2017											
State		Grade									
		2	3	4	5	6	7	8	9	10	11
NH	Reliability	–	–	–	–	–	0.841	–	–	–	–
	N	–	–	–	–	–	315	–	–	–	–
NM	Reliability	0.837	0.838	0.823	0.820	0.865	0.843	0.826	0.833	–	–
	N	349	642	633	793	499	623	371	352	–	–
NV	Reliability	0.876	0.862	0.855	0.850	0.864	0.873	–	–	–	–
	N	1,020	1,074	931	580	410	428	–	–	–	–
OR	Reliability	0.834	0.867	0.884	0.900	0.857	0.802	–	0.889	–	–
	N	303	441	453	389	395	373	–	334	–	–
PA	Reliability	–	–	–	–	0.846	0.879	–	–	–	–
	N	–	–	–	–	336	328	–	–	–	–
SD	Reliability	0.896	0.861	0.879	0.864	0.872	0.886	0.881	0.853	0.886	0.844
	N	382	1,366	1,350	2,608	1,426	1,366	1,202	1,286	931	503
UT	Reliability	0.868	0.871	0.847	0.875	0.863	0.836	0.846	0.873	0.893	–
	N	656	603	739	574	616	566	420	441	395	–
VT	Reliability	–	0.887	–	0.867	0.819	0.892	–	0.865	–	–
	N	–	328	–	336	336	434	–	367	–	–
WA	Reliability	0.814	0.831	0.841	0.854	0.878	0.883	–	–	–	–
	N	1,408	2,027	1,891	1,804	2,081	2,059	–	–	–	–
WI	Reliability	0.830	0.829	0.840	0.845	0.870	0.879	0.836	0.860	0.845	–
	N	2,290	4,085	4,361	4,610	5,194	5,543	1,679	1,524	377	–
WY	Reliability	–	0.872	0.862	0.827	0.828	0.850	–	–	–	–
	N	–	519	732	670	571	518	–	–	–	–

Table C.17. Test-Retest with Alternate Forms Reliability by State and Grade—Language Usage, Winter 2017–Spring 2017

Language Usage, Winter 2017–Spring 2017												
State		Grade										
		2	3	4	5	6	7	8	9	10	11	12
AZ	Reliability	–	0.829	0.849	0.852	0.849	–	–	–	–	–	–
	N	–	336	314	324	302	–	–	–	–	–	–
CA	Reliability	0.902	0.897	0.896	0.898	0.894	0.916	0.871	0.868	0.839	–	–
	N	6,692	5,695	6,094	5,823	2,424	1,880	1,090	1,208	1,109	–	–
CT	Reliability	0.870	0.890	0.878	0.891	0.883	0.883	0.878	0.895	0.842	–	–
	N	1,439	1,201	2,118	2,111	2,560	2,531	2,847	581	625	–	–
ID	Reliability	0.873	0.851	0.861	0.885	0.865	0.864	0.878	0.875	0.896	–	–
	N	349	685	705	833	842	741	830	349	341	–	–
IL	Reliability	0.864	0.871	0.872	0.877	0.871	0.887	0.890	0.866	0.842	0.845	–
	N	4,461	6,884	7,213	8,164	9,231	9,365	8,633	3,668	3,044	1,390	–
KY	Reliability	0.883	0.874	0.878	0.873	0.874	0.869	0.871	0.859	0.869	0.853	–
	N	5,547	8,101	11,989	8,687	10,319	7,913	7,420	1,879	1,432	781	–
LA	Reliability	0.859	0.858	0.862	0.842	0.827	0.825	0.833	0.735	0.748	–	–
	N	2,330	2,740	2,557	2,468	2,215	1,890	1,837	1,441	1,149	–	–
ME	Reliability	–	0.826	0.859	0.845	0.858	0.863	0.867	–	–	–	–
	N	–	459	499	621	525	435	449	–	–	–	–
MI	Reliability	0.866	0.863	0.860	0.864	0.865	0.847	0.858	0.860	0.856	0.827	0.820
	N	12,066	19,604	21,101	21,069	21,390	20,161	19,568	10,194	9,515	5,598	697
MO	Reliability	0.873	0.854	0.868	0.836	0.849	0.848	0.835	0.869	0.830	0.776	–
	N	555	1,712	1,616	1,551	1,681	1,528	1,290	824	575	327	–
MS	Reliability	0.861	0.827	0.837	0.846	0.869	0.853	0.869	0.851	0.799	0.837	–
	N	2,643	2,073	2,338	2,267	3,138	2,819	2,635	902	1,084	617	–
MT	Reliability	0.854	0.853	0.847	0.885	0.879	0.862	0.859	0.853	0.829	–	–
	N	821	1,945	1,768	1,593	2,210	2,234	2,260	548	1,278	–	–
NC	Reliability	0.891	0.905	0.877	0.876	0.897	0.891	0.906	–	–	–	–
	N	795	675	689	643	496	407	398	–	–	–	–

Language Usage, Winter 2017–Spring 2017												
State		Grade										
		2	3	4	5	6	7	8	9	10	11	12
NJ	Reliability	0.865	0.872	0.852	0.843	0.844	0.823	0.836	–	–	–	–
	N	1,141	1,833	1,993	1,815	1,709	1,054	906	–	–	–	–
NM	Reliability	0.855	0.846	0.855	0.841	0.862	0.818	0.865	0.825	0.796	0.804	–
	N	1,132	1,828	1,901	1,991	1,704	807	780	619	516	367	–
NV	Reliability	0.883	0.869	0.864	0.863	0.865	0.869	0.877	–	–	–	–
	N	1,084	1,172	1,207	782	480	446	340	–	–	–	–
OR	Reliability	0.856	0.885	0.886	0.879	0.850	0.857	0.900	–	–	–	–
	N	310	404	408	420	416	462	403	–	–	–	–
PA	Reliability	–	–	–	–	0.859	0.888	–	–	–	–	–
	N	–	–	–	–	448	417	–	–	–	–	–
SD	Reliability	0.897	0.873	0.894	0.872	0.863	0.882	0.890	0.854	0.853	0.868	–
	N	403	1,414	1,395	2,998	1,294	1,245	1,220	1,497	1,260	831	–
TN	Reliability	0.871	0.869	0.871	0.861	0.877	0.886	0.886	0.788	0.729	0.747	–
	N	1,498	2,671	2,498	2,722	2,047	2,030	1,858	318	321	319	–
UT	Reliability	0.885	0.894	0.872	0.884	0.865	0.876	0.864	0.899	0.874	–	–
	N	749	608	749	662	642	605	553	491	433	–	–
VT	Reliability	–	0.882	–	0.869	0.857	0.837	0.856	–	–	–	–
	N	–	370	–	309	354	402	366	–	–	–	–
WA	Reliability	0.845	0.850	0.842	0.849	0.872	0.884	0.901	–	–	–	–
	N	839	1,238	1,297	1,238	1,413	1,241	1,013	–	–	–	–
WI	Reliability	0.862	0.854	0.859	0.848	0.864	0.870	0.873	0.834	0.856	0.826	–
	N	1,760	3,177	3,552	3,662	4,820	4,617	4,709	1,741	1,001	339	–
WY	Reliability	0.852	0.865	0.864	0.863	0.850	0.879	0.881	–	–	–	–
	N	1,109	1,297	1,242	1,284	1,278	527	513	–	–	–	–

Table C.18. Test-Retest with Alternate Forms Reliability by State and Grade—Language Usage, Fall 2016–Winter 2017

Language Usage, Fall 2016–Winter 2017												
State		Grade										
		2	3	4	5	6	7	8	9	10	11	12
CA	Reliability	0.884	0.884	0.887	0.892	0.900	0.910	0.904	0.863	0.858	0.852	–
	N	7,173	7,810	8,207	8,171	5,630	5,175	5,352	1,842	1,680	320	–
CT	Reliability	0.849	0.870	0.865	0.881	0.870	0.865	0.877	0.850	0.823	–	–
	N	1,429	1,473	2,412	2,066	2,576	2,439	2,417	570	477	–	–
ID	Reliability	0.837	0.822	0.854	0.861	0.839	0.858	0.876	0.906	0.861	–	–
	N	381	735	752	871	805	854	865	501	381	–	–
IL	Reliability	0.833	0.852	0.855	0.870	0.869	0.876	0.879	0.858	0.840	0.852	–
	N	4,408	6,922	7,211	8,029	9,072	9,436	8,796	3,112	2,596	1,665	–
KY	Reliability	0.865	0.866	0.863	0.869	0.861	0.871	0.868	0.867	0.858	0.858	–
	N	6,266	8,537	12,003	8,944	11,155	7,808	7,811	2,537	2,078	961	–
LA	Reliability	0.836	0.826	0.841	0.839	0.807	0.806	0.806	0.731	0.743	–	–
	N	2,447	2,641	2,449	2,427	2,237	2,041	1,941	1,870	1,610	–	–
ME	Reliability	–	0.798	0.844	0.855	0.847	0.860	0.871	–	–	–	–
	N	–	450	491	619	517	433	491	–	–	–	–
MI	Reliability	0.841	0.851	0.851	0.859	0.856	0.849	0.848	0.850	0.847	0.812	0.768
	N	12,611	22,452	23,670	22,781	22,922	23,657	23,005	12,689	12,138	6,876	1,041
MO	Reliability	0.852	0.844	0.856	0.842	0.839	0.858	0.845	0.844	0.847	0.797	–
	N	470	1,963	2,107	1,958	1,834	1,664	1,531	1,070	927	632	–
MS	Reliability	0.819	0.816	0.816	0.816	0.852	0.830	0.858	0.820	0.805	0.847	–
	N	3,036	3,120	3,352	3,273	4,043	3,981	3,820	1,555	1,586	624	–
MT	Reliability	0.834	0.830	0.843	0.868	0.869	0.864	0.860	0.866	0.830	–	–
	N	695	1,991	1,766	1,638	2,282	2,384	2,400	571	1,265	–	–
NC	Reliability	0.874	0.893	0.873	0.883	0.890	0.876	0.897	–	–	–	–
	N	804	800	754	717	561	501	468	–	–	–	–
NH	Reliability	–	0.831	–	0.831	–	–	–	–	–	–	–
	N	–	396	–	365	–	–	–	–	–	–	–

Language Usage, Fall 2016–Winter 2017												
State		Grade										
		2	3	4	5	6	7	8	9	10	11	12
NJ	Reliability	0.844	0.849	0.847	0.842	0.835	0.831	0.832	–	–	–	–
	N	1,072	2,027	2,288	2,165	1,816	1,306	1,174	–	–	–	–
NM	Reliability	0.845	0.845	0.852	0.853	0.864	0.855	0.849	0.854	0.834	0.828	–
	N	1,132	2,015	2,084	2,062	2,380	1,469	1,483	941	662	447	–
NV	Reliability	0.881	0.875	0.879	0.881	0.856	0.848	0.867	0.797	0.794	0.804	–
	N	853	1,145	1,261	849	777	572	433	336	410	403	–
OR	Reliability	–	0.857	0.858	0.884	0.862	0.818	0.805	–	–	–	–
	N	–	397	394	379	643	696	632	–	–	–	–
PA	Reliability	–	–	–	–	0.874	0.879	–	–	–	–	–
	N	–	–	–	–	324	324	–	–	–	–	–
SD	Reliability	0.870	0.850	0.880	0.878	0.859	0.877	0.881	0.852	0.870	0.873	0.772
	N	363	1,546	1,401	3,187	1,451	1,438	1,428	1,603	1,442	1,019	465
TN	Reliability	0.862	0.883	0.870	0.854	0.872	0.889	0.881	0.846	0.855	0.853	–
	N	1,696	2,698	2,405	2,780	2,570	2,433	2,284	495	397	391	–
UT	Reliability	0.863	0.834	0.864	0.860	0.866	0.880	0.863	0.886	0.826	0.844	–
	N	672	851	924	820	766	689	656	475	439	400	–
VT	Reliability	–	0.859	0.832	0.844	0.826	–	–	–	–	–	–
	N	–	408	326	353	309	–	–	–	–	–	–
WA	Reliability	0.802	0.847	0.851	0.845	0.888	0.888	0.895	–	–	–	–
	N	806	1,399	1,527	1,338	1,440	1,212	1,061	–	–	–	–
WI	Reliability	0.844	0.852	0.854	0.850	0.872	0.862	0.873	0.866	0.851	0.868	–
	N	1,606	3,206	3,542	3,668	4,427	4,447	4,478	1,818	1,050	405	–
WY	Reliability	0.817	0.848	0.831	0.844	0.837	0.855	0.893	–	–	–	–
	N	1,081	1,290	1,242	1,266	1,169	522	520	–	–	–	–

Table C.19. Test-Retest with Alternate Forms Reliability by State and Grade—Mathematics, Spring 2017–Fall 2017

Mathematics, Spring 2017–Fall 2017													
State		Grade											
		K	1	2	3	4	5	6	7	8	9	10	11
AK	Reliability	–	–	–	–	–	0.902	0.913	0.925	0.870	–	–	–
	N	–	–	–	–	–	2,939	3,015	2,836	555	–	–	–
AZ	Reliability	0.840	0.709	0.800	0.822	0.899	0.881	0.909	0.922	–	–	–	–
	N	375	391	417	511	466	433	392	383	–	–	–	–
CA	Reliability	0.829	0.835	0.872	0.908	0.926	0.925	0.920	0.924	0.910	0.914	0.904	0.904
	N	9,653	11,859	14,328	13,012	13,658	12,580	10,971	10,493	5,856	5,893	2,848	1,042
CT	Reliability	0.807	0.816	0.783	0.865	0.896	0.891	0.913	0.913	0.913	0.922	0.932	–
	N	4,234	5,502	5,372	6,489	6,680	5,808	6,281	5,644	2,707	2,482	792	–
DC	Reliability	0.772	0.759	0.766	0.858	0.855	0.860	0.895	0.893	0.863	0.865	0.832	–
	N	1,783	1,730	1,649	1,395	1,310	761	832	755	752	1,488	984	–
DE	Reliability	0.819	0.812	0.821	0.869	0.907	0.901	0.905	0.909	–	0.919	0.913	–
	N	906	1,730	1,386	1,208	1,185	1,355	560	591	–	457	332	–
HI	Reliability	–	–	–	–	0.889	0.911	0.898	0.871	0.903	0.888	–	–
	N	–	–	–	–	344	315	434	629	582	336	–	–
ID	Reliability	0.837	0.846	0.774	0.861	0.890	0.899	0.907	0.925	0.920	0.899	0.872	–
	N	749	980	1,002	1,089	1,178	1,084	1,208	1,214	652	729	475	–
IL	Reliability	0.833	0.813	0.831	0.890	0.905	0.902	0.922	0.932	0.918	0.919	0.914	0.909
	N	35,241	45,087	62,081	65,311	67,037	71,639	66,084	67,877	15,625	12,095	5,501	1,708
KY	Reliability	0.820	0.770	0.831	0.854	0.882	0.878	0.905	0.912	0.919	0.922	0.875	–
	N	20,965	22,740	25,823	27,584	27,974	26,840	23,298	24,041	9,859	6,643	1,446	–
ME	Reliability	0.774	0.804	0.780	0.868	0.887	0.899	0.908	0.929	0.923	0.916	0.931	0.887
	N	2,098	3,267	5,250	6,275	6,485	5,907	6,695	6,425	3,388	2,058	817	364
MI	Reliability	0.799	0.787	0.772	0.862	0.890	0.889	0.906	0.913	0.906	0.906	0.893	0.877
	N	45,136	50,811	59,354	59,499	62,022	60,418	57,090	53,722	22,015	18,385	8,885	2,755
MT	Reliability	0.800	0.768	0.759	0.855	0.892	0.895	0.917	0.926	0.923	0.924	0.936	–
	N	2,127	2,423	3,437	5,099	4,889	4,945	4,170	4,144	1,933	1,839	792	–

Mathematics, Spring 2017–Fall 2017													
State		Grade											
		K	1	2	3	4	5	6	7	8	9	10	11
NC	Reliability	0.843	0.827	0.845	0.889	0.904	0.907	0.924	0.936	0.909	0.945	–	–
	N	12,258	12,265	13,603	13,241	12,976	11,935	11,399	9,993	509	455	–	–
NE	Reliability	–	–	–	–	0.887	–	–	–	–	–	–	–
	N	–	–	–	–	310	–	–	–	–	–	–	–
NH	Reliability	0.777	0.740	0.749	0.837	0.859	0.873	0.909	0.910	0.928	0.900	–	–
	N	1,344	2,148	3,046	2,639	2,484	2,571	2,437	2,435	411	385	–	–
NM	Reliability	0.759	0.788	0.783	0.850	0.883	0.884	0.914	0.907	0.863	0.875	0.901	0.887
	N	2,006	2,275	2,618	2,611	2,586	2,697	2,741	2,674	704	795	718	482
NV	Reliability	0.824	0.806	0.858	0.893	0.909	0.904	0.914	0.915	0.904	0.914	–	–
	N	4,214	8,955	8,916	9,181	8,836	7,729	6,141	4,095	906	304	–	–
NY	Reliability	0.804	0.779	–	–	–	–	–	–	–	–	–	–
	N	475	531	–	–	–	–	–	–	–	–	–	–
OR	Reliability	0.791	0.782	0.802	0.863	0.895	0.867	0.899	0.909	0.904	0.926	0.901	–
	N	1,141	1,318	1,736	1,569	1,686	1,493	1,742	1,669	895	908	583	–
PA	Reliability	–	0.693	0.793	0.858	0.877	0.904	0.916	0.932	–	–	–	–
	N	–	304	300	307	340	338	371	371	–	–	–	–
RI	Reliability	0.817	0.785	0.704	0.802	0.866	0.894	0.880	0.925	–	0.881	–	–
	N	380	366	468	491	524	545	455	502	–	329	–	–
SD	Reliability	0.817	0.760	0.788	0.864	0.904	0.906	0.913	0.919	0.916	0.907	0.916	0.926
	N	2,662	2,740	2,883	3,137	3,160	4,233	2,627	2,480	2,001	2,010	1,433	562
TX	Reliability	–	–	0.889	–	–	–	–	–	–	–	–	–
	N	–	–	302	–	–	–	–	–	–	–	–	–
UT	Reliability	0.822	0.778	0.757	0.889	0.901	0.903	0.896	0.921	0.922	0.926	0.906	–
	N	907	883	813	705	721	630	715	738	531	476	504	–
VT	Reliability	0.757	0.746	0.736	0.845	0.875	0.903	0.913	0.909	0.896	0.921	–	–
	N	348	307	465	643	619	736	567	623	338	389	–	–
WA	Reliability	0.826	0.819	0.779	0.878	0.894	0.895	0.912	0.922	0.915	0.922	0.904	0.869
	N	6,421	9,167	11,847	12,105	12,277	10,802	9,573	8,257	2,668	2,102	1,034	449

Mathematics, Spring 2017–Fall 2017													
State		Grade											
		K	1	2	3	4	5	6	7	8	9	10	11
WI	Reliability	0.804	0.786	0.791	0.878	0.896	0.893	0.923	0.934	0.925	0.918	0.923	–
	N	9,433	13,678	18,720	23,175	23,640	22,642	22,213	21,579	6,059	3,990	913	–
WY	Reliability	0.827	0.758	0.806	0.853	0.892	0.888	0.900	0.913	0.914	0.902	–	–
	N	1,353	1,474	1,375	1,693	1,812	1,550	1,282	1,132	542	457	–	–

Table C.20. Test-Retest with Alternate Forms Reliability by State and Grade—Mathematics, Winter 2017–Spring 2017

Mathematics, Winter 2017–Spring 2017														
State		Grade												
		K	1	2	3	4	5	6	7	8	9	10	11	12
AK	Reliability	–	–	–	–	–	–	0.921	0.914	0.926	–	–	–	–
	N	–	–	–	–	–	–	973	2,793	2,584	–	–	–	–
AZ	Reliability	0.781	0.859	0.780	0.858	0.883	0.888	0.905	–	–	–	–	–	–
	N	453	433	446	485	455	482	450	–	–	–	–	–	–
CA	Reliability	0.809	0.873	0.889	0.899	0.916	0.930	0.912	0.928	0.920	0.895	0.889	0.891	0.859
	N	10,275	12,352	14,769	12,663	13,288	13,227	10,625	10,049	8,712	7,784	6,361	2,821	767
CO	Reliability	–	0.859	0.868	0.860	0.885	0.919	0.910	0.900	0.903	–	–	–	–
	N	–	302	984	1,042	1,080	1,043	912	760	877	–	–	–	–
CT	Reliability	0.779	0.852	0.855	0.855	0.879	0.912	0.917	0.920	0.926	0.919	0.917	0.912	–
	N	5,134	7,206	8,397	9,006	9,380	9,489	9,437	9,103	9,337	5,244	4,092	1,059	–
DC	Reliability	0.740	0.801	0.856	0.844	0.867	0.884	0.900	0.899	0.925	0.855	0.826	0.757	–
	N	2,156	2,013	1,965	1,649	1,398	1,238	1,343	1,246	1,055	1,394	1,074	502	–
DE	Reliability	0.824	0.874	0.803	0.876	0.912	0.915	0.914	0.906	0.903	0.911	0.900	–	–
	N	850	1,873	1,816	1,629	1,513	1,586	516	429	375	407	381	–	–
FL	Reliability	0.790	0.847	0.860	0.840	0.860	0.867	0.862	0.856	0.809	0.783	0.804	–	–
	N	5,190	5,152	5,125	5,138	4,726	4,697	5,048	4,263	3,757	612	569	–	–
GA	Reliability	–	–	–	–	–	–	0.904	0.928	0.914	–	–	–	–
	N	–	–	–	–	–	–	524	602	480	–	–	–	–
HI	Reliability	–	–	–	–	–	–	0.856	0.854	0.910	–	–	–	–
	N	–	–	–	–	–	–	396	601	580	–	–	–	–

Mathematics, Winter 2017–Spring 2017														
State		Grade												
		K	1	2	3	4	5	6	7	8	9	10	11	12
ID	Reliability	0.819	0.840	0.875	0.848	0.893	0.912	0.899	0.904	0.919	0.933	0.912	–	–
	N	774	1,088	1,042	939	1,026	1,039	1,232	1,491	1,558	554	424	–	–
IL	Reliability	0.799	0.858	0.857	0.872	0.886	0.905	0.909	0.919	0.918	0.911	0.906	0.893	0.843
	N	37,061	49,153	72,338	82,099	83,209	81,509	79,144	78,350	74,574	13,940	9,591	4,602	1,092
KY	Reliability	0.807	0.861	0.859	0.864	0.887	0.903	0.905	0.914	0.924	0.914	0.901	0.845	–
	N	23,940	26,758	29,023	29,865	29,498	28,443	25,132	25,859	25,223	8,545	5,361	1,480	–
LA	Reliability	0.786	0.858	0.859	0.849	0.867	0.877	0.861	0.864	0.878	0.858	0.842	–	–
	N	5,571	6,010	6,112	5,035	4,587	4,134	3,916	3,614	3,277	2,345	1,619	–	–
ME	Reliability	0.760	0.837	0.860	0.855	0.883	0.913	0.897	0.917	0.922	0.927	0.911	–	–
	N	1,447	2,665	3,760	4,255	4,331	3,847	3,502	3,215	2,948	751	669	–	–
MI	Reliability	0.777	0.851	0.845	0.861	0.883	0.907	0.902	0.910	0.913	0.905	0.897	0.874	0.823
	N	48,442	53,075	55,834	52,660	54,567	54,436	47,589	43,035	41,088	18,885	17,760	9,182	1,732
MO	Reliability	0.801	0.867	0.844	0.863	0.894	0.907	0.896	0.915	0.901	0.889	0.876	0.846	–
	N	3,297	4,165	5,612	4,908	5,023	4,081	3,615	3,524	3,147	1,023	826	374	–
MS	Reliability	0.832	0.862	0.870	0.858	0.871	0.897	0.902	0.907	0.902	0.871	0.889	0.851	–
	N	7,111	8,554	8,820	5,623	5,810	5,039	5,736	6,349	5,913	2,951	1,479	620	–
MT	Reliability	0.811	0.863	0.828	0.859	0.884	0.913	0.907	0.915	0.927	0.901	0.914	–	–
	N	2,163	2,384	3,157	4,588	4,635	4,468	4,265	3,307	3,227	896	1,771	–	–
NC	Reliability	0.836	0.886	0.872	0.891	0.901	0.918	0.919	0.936	0.942	0.926	0.905	0.922	–
	N	14,501	15,465	16,333	16,815	15,506	14,187	13,058	11,652	11,540	662	481	355	–
NE	Reliability	–	–	–	–	0.884	–	–	–	–	–	–	–	–
	N	–	–	–	–	316	–	–	–	–	–	–	–	–
NH	Reliability	0.784	0.841	0.844	0.840	0.859	0.885	0.900	0.909	0.911	0.863	0.857	–	–
	N	1,003	2,522	3,084	2,857	2,451	2,596	1,895	1,577	1,268	405	305	–	–
NJ	Reliability	0.752	0.826	0.844	0.868	0.892	0.886	0.887	0.888	0.889	0.894	0.914	0.886	–
	N	5,142	7,296	9,054	7,931	7,877	9,333	9,460	7,338	5,625	1,058	865	516	–
NM	Reliability	0.761	0.827	0.850	0.820	0.869	0.889	0.906	0.902	0.904	0.852	0.896	0.904	–
	N	1,486	1,784	2,781	2,748	2,877	2,932	3,386	2,443	2,234	1,187	914	697	–

Mathematics, Winter 2017–Spring 2017														
State		Grade												
		K	1	2	3	4	5	6	7	8	9	10	11	12
NV	Reliability	0.808	0.860	0.871	0.887	0.901	0.909	0.908	0.910	0.918	0.932	0.890	0.885	–
	N	4,120	9,009	8,831	9,099	8,736	8,002	6,309	3,832	2,948	372	343	310	–
NY	Reliability	0.755	0.818	0.801	–	–	–	–	–	–	–	–	–	–
	N	424	468	468	–	–	–	–	–	–	–	–	–	–
OK	Reliability	–	–	–	–	–	–	–	–	0.907	–	–	–	–
	N	–	–	–	–	–	–	–	–	401	–	–	–	–
OR	Reliability	0.786	0.834	0.826	0.861	0.893	0.897	0.904	0.895	0.919	0.928	0.886	0.863	–
	N	1,112	1,288	1,812	1,686	1,864	1,759	1,729	1,635	1,639	778	666	369	–
PA	Reliability	–	0.878	0.802	0.856	0.878	0.909	0.913	0.913	0.882	–	–	–	–
	N	–	405	360	362	383	362	475	420	404	–	–	–	–
RI	Reliability	0.834	0.841	0.830	0.807	0.865	0.877	0.890	0.908	0.875	0.808	–	–	–
	N	469	475	596	490	401	510	409	513	346	355	–	–	–
SD	Reliability	0.803	0.846	0.861	0.866	0.895	0.905	0.908	0.918	0.919	0.892	0.899	0.917	–
	N	2,862	3,039	3,045	3,367	3,361	4,448	2,904	2,688	2,571	2,026	1,821	1,126	–
TN	Reliability	0.724	0.795	0.815	0.848	0.866	0.886	0.894	0.903	0.915	0.899	0.902	0.834	0.802
	N	11,121	10,624	10,682	10,873	9,949	11,221	9,452	9,255	8,933	6,321	5,572	3,179	753
UT	Reliability	0.802	0.851	0.841	0.890	0.903	0.923	0.899	0.926	0.912	0.906	0.897	–	–
	N	929	940	980	717	741	666	739	807	675	643	608	–	–
VT	Reliability	0.727	0.820	0.843	0.846	0.865	0.902	0.911	0.905	0.933	0.913	0.919	–	–
	N	419	416	525	658	583	679	679	528	515	303	301	–	–
WA	Reliability	0.823	0.862	0.843	0.876	0.891	0.905	0.910	0.919	0.924	0.915	0.893	0.842	–
	N	7,144	8,884	12,910	13,810	13,308	13,288	8,995	7,448	6,463	1,781	1,186	570	–
WI	Reliability	0.811	0.861	0.851	0.878	0.892	0.907	0.916	0.929	0.932	0.920	0.899	0.886	–
	N	9,662	12,850	18,770	23,321	23,872	22,891	22,871	21,791	21,063	5,350	3,590	784	–
WY	Reliability	0.815	0.849	0.826	0.845	0.879	0.896	0.903	0.913	0.912	0.917	0.893	–	–
	N	4,248	5,816	6,010	6,108	5,852	5,920	3,839	2,953	2,615	598	413	–	–

Table C.21. Test-Retest with Alternate Forms Reliability by State and Grade—Mathematics, Fall 2016–Winter 2017

Mathematics, Fall 2016–Winter 2017														
		Grade												
State		K	1	2	3	4	5	6	7	8	9	10	11	12
AK	Reliability	–	–	–	–	–	–	0.925	0.917	0.931	–	–	–	–
	N	–	–	–	–	–	–	852	2,826	2,816	–	–	–	–
AZ	Reliability	0.701	0.732	0.800	0.821	0.857	0.853	0.866	–	–	–	–	–	–
	N	389	357	409	444	411	428	436	–	–	–	–	–	–
CA	Reliability	0.741	0.846	0.871	0.888	0.906	0.920	0.916	0.925	0.922	0.903	0.896	0.902	0.876
	N	8,821	12,323	14,844	15,904	16,262	16,595	16,045	15,161	14,412	8,724	6,157	2,944	1,022
CO	Reliability	–	–	0.838	0.848	0.870	0.904	0.907	0.901	0.917	0.892	0.914	–	–
	N	–	–	1,050	1,116	1,139	1,116	1,139	1,136	1,164	581	543	–	–
CT	Reliability	0.751	0.832	0.842	0.847	0.877	0.905	0.903	0.900	0.924	0.915	0.906	0.930	–
	N	3,589	6,921	7,624	8,511	8,675	8,436	8,309	7,676	7,910	4,054	3,183	931	–
DC	Reliability	0.694	0.818	0.852	0.825	0.858	0.876	0.877	0.897	0.909	0.826	0.826	0.807	–
	N	2,176	1,968	1,934	1,731	1,462	1,321	1,211	1,057	889	1,608	1,267	717	–
DE	Reliability	0.807	0.812	0.845	0.865	0.894	0.914	0.870	0.799	0.877	0.888	0.885	–	–
	N	769	1,749	1,725	1,540	1,488	1,599	603	545	447	407	380	–	–
FL	Reliability	0.712	0.806	0.843	0.839	0.848	0.863	0.844	0.856	0.854	0.872	0.886	–	–
	N	5,149	5,184	5,170	5,230	4,814	4,755	5,130	4,421	3,939	712	719	–	–
GA	Reliability	–	–	–	–	–	–	–	0.929	–	–	–	–	–
	N	–	–	–	–	–	–	–	382	–	–	–	–	–
HI	Reliability	–	–	–	0.888	0.891	0.901	0.839	0.846	0.908	–	–	–	–
	N	–	–	–	401	443	457	442	600	581	–	–	–	–
ID	Reliability	0.749	0.799	0.820	0.795	0.866	0.890	0.892	0.894	0.915	0.916	0.916	–	–
	N	432	572	881	1,036	1,110	1,169	1,300	1,502	1,556	582	464	–	–
IL	Reliability	0.767	0.845	0.858	0.875	0.894	0.913	0.915	0.925	0.929	0.909	0.897	0.907	0.880
	N	31,067	43,896	60,588	64,270	66,019	64,314	65,755	61,964	62,192	15,484	11,156	6,798	1,691
KY	Reliability	0.774	0.846	0.845	0.856	0.879	0.896	0.900	0.910	0.917	0.915	0.919	0.889	–
	N	21,569	26,474	28,725	29,312	28,905	28,019	25,088	25,534	25,214	8,872	5,949	2,004	–

Mathematics, Fall 2016–Winter 2017														
State		Grade												
		K	1	2	3	4	5	6	7	8	9	10	11	12
LA	Reliability	0.711	0.832	0.844	0.821	0.838	0.852	0.834	0.860	0.850	0.851	0.822	–	–
	N	5,500	5,996	6,079	4,690	4,348	4,220	4,120	3,953	3,601	2,612	1,797	–	–
ME	Reliability	0.725	0.825	0.837	0.830	0.864	0.909	0.892	0.911	0.919	0.912	0.900	–	–
	N	851	2,197	3,346	4,263	4,265	3,843	3,332	3,199	3,076	617	542	–	–
MI	Reliability	0.733	0.827	0.846	0.850	0.873	0.900	0.897	0.906	0.907	0.906	0.894	0.878	0.826
	N	43,575	52,317	55,507	54,625	56,782	56,157	50,422	47,153	45,113	22,545	21,601	10,776	2,777
MO	Reliability	0.752	0.843	0.836	0.843	0.881	0.887	0.882	0.909	0.895	0.881	0.899	0.891	–
	N	2,813	4,074	5,498	5,225	5,348	4,331	3,671	3,577	3,292	1,089	898	648	–
MS	Reliability	0.741	0.821	0.841	0.832	0.850	0.873	0.885	0.899	0.899	0.889	0.868	0.859	–
	N	7,074	8,622	8,681	7,269	7,315	6,524	7,274	7,960	7,597	3,657	2,172	705	–
MT	Reliability	0.709	0.822	0.794	0.825	0.861	0.899	0.898	0.914	0.921	0.922	0.904	–	–
	N	1,782	2,300	3,002	4,639	4,649	4,520	4,302	3,355	3,331	784	1,763	–	–
NC	Reliability	0.783	0.852	0.856	0.874	0.886	0.909	0.909	0.924	0.933	0.908	0.891	0.896	–
	N	12,637	15,333	16,428	16,954	15,557	14,362	14,058	12,827	12,886	596	406	359	–
NE	Reliability	–	–	–	0.869	0.871	0.874	0.905	0.903	0.919	0.927	0.946	0.931	–
	N	–	–	–	778	702	711	709	655	741	586	534	521	–
NH	Reliability	0.701	0.762	0.797	0.793	0.859	0.881	0.876	0.905	0.916	0.935	0.898	–	–
	N	711	2,067	3,008	3,469	3,124	3,297	2,320	2,243	2,183	498	441	–	–
NJ	Reliability	0.706	0.797	0.834	0.851	0.882	0.882	0.882	0.882	0.862	0.912	0.865	0.867	0.780
	N	3,574	6,690	8,715	7,911	8,399	9,455	9,906	7,798	6,339	841	797	576	319
NM	Reliability	0.712	0.794	0.819	0.816	0.856	0.893	0.898	0.910	0.914	0.869	0.890	0.893	0.894
	N	1,446	1,898	2,956	3,035	3,074	3,175	3,655	2,910	2,866	1,639	1,230	922	393
NV	Reliability	0.742	0.812	0.856	0.874	0.894	0.907	0.910	0.922	0.929	0.904	0.882	0.897	0.863
	N	2,794	8,838	8,706	9,061	9,051	8,557	7,263	6,443	6,393	1,413	735	688	475
NY	Reliability	0.688	0.819	0.840	–	–	–	–	–	–	–	–	–	–
	N	427	464	464	–	–	–	–	–	–	–	–	–	–
OK	Reliability	–	–	–	–	–	–	–	–	0.832	–	–	–	–
	N	–	–	–	–	–	–	–	–	383	–	–	–	–

Mathematics, Fall 2016–Winter 2017														
State		Grade												
		K	1	2	3	4	5	6	7	8	9	10	11	12
OR	Reliability	0.785	0.822	0.789	0.863	0.881	0.906	0.907	0.893	0.913	0.904	0.886	0.877	–
	N	758	1,236	1,334	1,454	1,953	1,905	2,005	1,956	1,953	1,049	858	628	–
PA	Reliability	–	0.769	0.810	0.822	0.869	0.903	0.917	0.896	0.885	–	–	–	–
	N	–	399	362	365	385	367	351	329	398	–	–	–	–
RI	Reliability	0.786	0.829	0.850	0.760	0.856	0.892	0.897	0.901	0.902	0.830	–	–	–
	N	324	447	569	482	395	502	392	486	361	363	–	–	–
SD	Reliability	0.768	0.816	0.839	0.838	0.887	0.898	0.891	0.899	0.912	0.895	0.914	0.918	0.876
	N	2,550	2,917	2,956	3,447	3,280	4,786	3,011	2,816	2,683	2,083	1,932	1,289	534
TN	Reliability	0.737	0.834	0.834	0.859	0.874	0.895	0.892	0.903	0.911	0.904	0.892	0.851	0.787
	N	10,971	10,789	10,910	11,135	10,107	11,494	9,660	9,076	8,792	6,588	5,716	3,615	2,250
UT	Reliability	0.812	0.839	0.840	0.831	0.874	0.873	0.890	0.913	0.909	0.892	0.847	0.871	–
	N	907	928	973	873	925	799	832	879	780	624	596	496	–
VT	Reliability	–	0.790	0.840	0.836	0.860	0.892	0.873	0.909	0.926	0.883	0.922	–	–
	N	–	406	514	698	683	739	754	587	600	328	321	–	–
WA	Reliability	0.784	0.822	0.840	0.860	0.881	0.901	0.900	0.912	0.916	0.915	0.888	0.884	0.871
	N	3,954	8,278	12,493	15,927	14,958	15,166	11,180	9,838	9,219	2,016	1,463	669	358
WI	Reliability	0.751	0.833	0.841	0.860	0.881	0.898	0.909	0.927	0.933	0.922	0.906	0.911	–
	N	7,139	11,536	18,013	22,801	23,317	22,915	22,922	21,764	20,993	5,659	4,065	1,047	–
WY	Reliability	0.748	0.821	0.791	0.830	0.867	0.884	0.889	0.903	0.906	0.920	0.906	–	–
	N	3,029	5,791	5,973	6,076	5,875	5,902	3,837	2,962	2,638	682	481	–	–

Table C.22. Test-Retest with Alternate Forms Reliability by State and Grade—Science, Spring 2017–Fall 2017

Science, Spring 2017–Fall 2017									
State		Grade							
		3	4	5	6	7	8	9	10
AR	Reliability	0.759	0.824	0.828	0.822	0.835	0.849	–	–
	N	893	1,199	1,268	1,239	1,345	511	–	–
CA	Reliability	–	–	0.744	0.815	0.842	–	–	–
	N	–	–	415	1,583	1,873	–	–	–
CO	Reliability	–	0.799	0.809	0.817	0.812	0.765	0.814	–
	N	–	690	701	1,516	1,471	601	545	–
CT	Reliability	–	0.760	0.796	0.796	0.804	0.814	0.864	–
	N	–	338	513	595	581	312	319	–
IA	Reliability	–	0.811	–	0.796	0.829	0.819	–	–
	N	–	377	–	377	495	378	–	–
IL	Reliability	0.863	0.832	0.861	0.847	0.856	–	–	–
	N	1,720	2,104	2,189	2,840	2,880	–	–	–
KS	Reliability	–	–	0.791	0.848	0.841	–	–	–
	N	–	–	337	602	727	–	–	–
KY	Reliability	0.813	0.782	0.805	0.817	0.870	–	–	–
	N	803	453	444	709	549	–	–	–
MI	Reliability	0.799	0.821	0.805	0.810	0.838	0.832	0.862	0.825
	N	7,058	8,321	8,543	9,673	10,496	1,942	1,380	508
OH	Reliability	–	0.765	0.738	0.774	0.796	–	–	–
	N	–	364	407	419	413	–	–	–
WA	Reliability	0.830	–	0.765	0.798	0.797	–	–	–
	N	324	–	475	555	561	–	–	–
WI	Reliability	–	–	–	0.836	0.823	–	–	–
	N	–	–	–	343	316	–	–	–

Table C.23. Test-Retest with Alternate Forms Reliability by State and Grade—Science, Winter 2017–Spring 2017

Science, Winter 2017–Spring 2017										
		Grade								
State		3	4	5	6	7	8	9	10	11
AR	Reliability	0.805	0.828	0.842	0.837	0.840	0.847	0.856	–	–
	N	1,077	1,419	1,446	1,536	1,470	1,512	362	–	–
CA	Reliability	–	–	0.806	0.839	0.835	0.828	0.867	–	–
	N	–	–	3,031	882	880	3,338	344	–	–
CO	Reliability	–	0.797	0.816	0.819	0.812	0.836	0.829	0.836	–
	N	–	716	943	1,606	1,528	1,688	596	614	–
CT	Reliability	–	–	0.775	0.797	0.835	0.830	0.843	0.896	–
	N	–	–	538	548	523	555	328	336	–
IL	Reliability	0.855	0.821	0.843	0.840	0.863	0.860	–	–	–
	N	2,339	2,929	3,232	3,171	3,218	2,628	–	–	–
KY	Reliability	0.755	0.794	0.836	0.839	0.836	0.821	0.826	–	–
	N	448	674	313	731	1,187	714	410	–	–
MA	Reliability	–	–	0.793	–	–	–	–	–	–
	N	–	–	491	–	–	–	–	–	–
MI	Reliability	0.797	0.804	0.835	0.829	0.841	0.845	0.846	0.827	0.832
	N	6,359	9,227	8,281	9,972	8,886	8,906	2,194	1,979	391
MO	Reliability	–	–	–	0.826	0.854	0.820	–	–	–
	N	–	–	–	405	402	354	–	–	–
WA	Reliability	–	–	0.852	0.799	0.829	0.865	–	–	–
	N	–	–	415	386	587	400	–	–	–

Table C.24. Test-Retest with Alternate Forms Reliability by State and Grade—Science, Fall 2016–Winter 2017

Science, Fall 2016–Winter 2017										
State		Grade								
		3	4	5	6	7	8	9	10	11
AR	Reliability	0.792	0.796	0.827	0.818	0.825	0.842	0.829	–	–
	N	990	1,237	1,520	1,544	1,408	1,354	353	–	–
CA	Reliability	–	–	0.800	0.802	0.827	0.804	0.869	–	–
	N	–	–	3,214	690	653	3,116	325	–	–
CO	Reliability	–	0.706	0.789	0.826	0.835	0.813	0.787	0.809	–
	N	–	709	906	1,622	1,516	1,699	656	620	–
CT	Reliability	–	–	0.814	0.811	0.799	0.783	0.872	0.884	–
	N	–	–	346	387	393	473	330	326	–
IL	Reliability	0.843	0.829	0.832	0.832	0.846	0.842	–	–	–
	N	1,919	2,271	2,790	3,010	2,925	2,751	–	–	–
KS	Reliability	–	–	–	0.828	0.854	0.871	–	–	–
	N	–	–	–	355	426	426	–	–	–
KY	Reliability	0.814	0.791	–	0.808	0.803	0.831	0.812	–	–
	N	358	658	–	763	1,073	484	315	–	–
MA	Reliability	–	–	0.765	–	–	0.867	–	–	–
	N	–	–	571	–	–	341	–	–	–
MI	Reliability	0.777	0.794	0.811	0.810	0.828	0.835	0.840	0.851	0.814
	N	8,601	11,026	9,989	11,117	9,540	9,661	2,408	2,347	647
MO	Reliability	–	–	–	0.822	0.840	0.841	–	–	–
	N	–	–	–	418	409	384	–	–	–
NJ	Reliability	–	–	–	–	–	0.798	–	–	–
	N	–	–	–	–	–	326	–	–	–
WA	Reliability	–	–	0.852	0.820	0.801	0.851	–	–	–
	N	–	–	343	524	811	555	–	–	–

Appendix D: Marginal Reliability by State

Table D.1. Marginal Reliability of Overall RIT Scores by State

	Reading		Language Usage		Mathematics		Science	
State	N	Reliability	N	Reliability	N	Reliability	N	Reliability
AK	51,421	0.970	1,639	0.922	51,386	0.981	–	–
AL	6,334	0.984	4,646	0.974	6,385	0.989	–	–
AR	–	–	–	–	–	–	45,034	0.946
AZ	27,535	0.984	12,344	0.976	27,465	0.990	–	–
CA	638,279	0.985	216,595	0.979	650,575	0.990	62,513	0.945
CO	31,188	0.977	2,671	0.978	33,409	0.985	36,749	0.940
CT	329,546	0.984	73,710	0.976	360,844	0.990	19,086	0.941
DC	69,591	0.985	1,412	0.974	89,412	0.990	1,372	0.913
DE	53,312	0.986	1,785	0.971	55,039	0.990	1,354	0.917
FL	147,409	0.985	3,814	0.976	146,590	0.990	336	0.905
GA	3,876	0.988	1,953	0.973	8,353	0.988	43,593	0.954
HI	20,329	0.980	3,387	0.979	21,034	0.989	438	0.958
IA	–	–	–	–	–	–	47,217	0.937
ID	57,322	0.985	36,846	0.976	62,264	0.991	1,121	0.938
IL	2,821,453	0.984	362,387	0.976	2,853,668	0.990	115,402	0.945
IN	4,816	0.978	1,471	0.967	6,291	0.983	617	0.900
KS	735	0.967	351	0.962	686	0.979	22,705	0.934
KY	1,175,059	0.986	348,865	0.975	1,178,738	0.990	31,761	0.944
LA	160,949	0.986	64,842	0.978	159,730	0.990	–	–
MA	6964	0.985	–	–	8,442	0.990	5,437	0.949
MD	6594	0.986	3,289	0.957	7,231	0.990	3,085	0.953
ME	232,454	0.983	53,701	0.973	235,269	0.988	424	0.932
MI	2,544,070	0.986	907,503	0.977	2,551,396	0.990	371,595	0.951
MN	850	0.981	482	0.981	1,447	0.984	455	0.904
MO	143,505	0.985	47,645	0.976	144,391	0.990	5,656	0.935
MS	235,119	0.984	93,389	0.975	234,424	0.990	–	–
MT	181,739	0.983	105,068	0.974	182,937	0.989	5,369	0.942
NC	524,790	0.985	25,245	0.979	564,309	0.991	663	0.935

State	Reading		Language Usage		Mathematics		Science	
	N	Reliability	N	Reliability	N	Reliability	N	Reliability
ND	–	–	–	–	–	–	657	0.900
NE	19,747	0.972	–	–	19,310	0.982	–	–
NH	138,381	0.982	20,672	0.976	143,572	0.988	1,047	0.936
NJ	288,428	0.984	70,346	0.971	340,094	0.989	9,369	0.941
NM	158,036	0.983	66,615	0.976	159,968	0.989	–	–
NV	403,279	0.985	41,736	0.979	394,368	0.990	9,453	0.940
NY	10,202	0.987	309	0.976	13,513	0.990	2,624	0.934
OH	–	–	–	–	–	–	5,867	0.921
OK	5,167	0.982	852	0.957	6,915	0.987	1,919	0.937
OR	83,745	0.984	23,182	0.977	88,787	0.990	2,669	0.940
PA	17,023	0.982	7,805	0.970	17,248	0.988	368	0.932
RI	25,422	0.981	4,498	0.970	25,665	0.989	2,865	0.944
SC	536	0.975	393	0.945	421	0.982	–	–
SD	168,811	0.986	77,268	0.977	171,907	0.991	4,168	0.936
TN	368,439	0.986	73,084	0.979	369,337	0.990	–	–
TX	11,063	0.987	2,719	0.966	11,285	0.991	725	0.955
UT	44,550	0.987	30,801	0.980	44,654	0.992	–	–
VA	2,104	0.976	1,837	0.970	2,205	0.983	755	0.955
VT	29,078	0.983	14,661	0.977	31,257	0.989	–	–
WA	552,106	0.984	68,459	0.973	557,851	0.989	23,053	0.937
WI	874,358	0.982	172,180	0.972	892,911	0.989	6,203	0.922
WV	1,684	0.983	579	0.968	1,660	0.986	–	–
WY	202,384	0.984	66,309	0.971	203,971	0.989	–	–

Table D.2. Marginal Reliability of Overall RIT Scores by State and Grade—Reading

State		Reading												
		Grade												
		K	1	2	3	4	5	6	7	8	9	10	11	12
AK	Reliability	–	0.974	0.976	0.963	0.961	0.958	0.959	0.955	0.955	0.954	0.955	0.958	0.955
	N	–	343	359	3,904	3,833	6,944	8,655	12,495	12,200	862	566	513	451
AL	Reliability	0.952	0.957	0.952	0.960	0.957	0.956	0.955	0.963	0.962	0.954	0.969	–	–
	N	341	660	686	573	648	674	702	619	601	336	306	–	–
AZ	Reliability	0.931	0.953	0.949	0.953	0.955	0.954	0.953	0.956	0.952	0.952	0.955	0.949	0.948
	N	2,117	2,481	2,753	3,242	3,020	2,969	2,893	2,615	2,507	962	732	636	608
CA	Reliability	0.958	0.970	0.967	0.965	0.965	0.964	0.962	0.963	0.960	0.959	0.960	0.964	0.968
	N	41,086	52,598	63,656	65,176	67,247	68,155	64,557	63,036	60,510	38,187	30,818	15,575	6,988
CO	Reliability	0.963	0.961	0.963	0.956	0.955	0.952	0.954	0.952	0.958	0.958	0.961	0.969	0.969
	N	412	864	3,485	3,749	3,777	3,629	3,171	2,946	2,913	2,702	2,399	638	503
CT	Reliability	0.957	0.969	0.966	0.960	0.956	0.956	0.957	0.956	0.956	0.964	0.966	0.971	0.972
	N	14,839	26,571	30,511	32,697	35,833	36,269	37,622	36,128	35,517	22,123	16,253	3,860	1,323
DC	Reliability	0.955	0.963	0.961	0.956	0.957	0.955	0.959	0.960	0.958	0.960	0.960	0.959	0.971
	N	8,825	8,265	7,871	7,272	6,417	6,015	6,008	5,525	4,857	3,584	2,513	1,505	832
DE	Reliability	0.949	0.968	0.965	0.960	0.955	0.952	0.957	0.954	0.952	0.955	0.964	0.965	0.948
	N	3,054	7,199	7,011	6,385	6,045	6,485	4,044	3,516	3,185	2,453	2,175	1,219	541
FL	Reliability	0.957	0.965	0.961	0.957	0.947	0.948	0.947	0.948	0.950	0.957	0.959	0.958	0.974
	N	16,611	16,533	16,626	16,769	15,414	15,114	16,382	14,174	12,728	2,819	2,703	1,160	376
GA	Reliability	0.961	0.968	0.969	0.968	–	–	0.950	0.960	–	–	–	–	–
	N	637	670	573	328	–	–	417	417	–	–	–	–	–
HI	Reliability	0.960	0.969	0.964	0.955	0.956	0.956	0.929	0.899	0.909	0.919	0.928	0.934	0.966
	N	639	967	1,034	1,453	1,808	1,850	2,011	2,701	2,627	2,872	1,292	606	467
ID	Reliability	0.945	0.967	0.966	0.960	0.956	0.956	0.952	0.949	0.949	0.958	0.956	0.960	–
	N	3,363	4,731	5,888	5,861	6,226	6,193	6,065	5,917	5,744	3,308	2,639	1,212	–
IL	Reliability	0.957	0.968	0.966	0.963	0.960	0.958	0.954	0.954	0.952	0.962	0.964	0.968	0.976
	N	144,003	190,274	303,992	332,108	335,970	333,372	331,355	328,623	323,368	90,022	65,527	31,344	10,655
IN	Reliability	–	–	–	–	–	–	–	0.959	0.962	0.969	0.969	0.971	–
	N	–	–	–	–	–	–	–	853	763	719	666	594	–

Reading														
State		Grade												
		K	1	2	3	4	5	6	7	8	9	10	11	12
KY	Reliability	0.950	0.962	0.963	0.959	0.957	0.954	0.952	0.953	0.953	0.963	0.962	0.966	0.971
	N	102,672	117,157	126,429	131,838	129,857	126,711	114,563	116,372	114,004	51,333	33,069	9,603	834
LA	Reliability	0.954	0.967	0.964	0.962	0.961	0.962	0.961	0.961	0.961	0.969	0.969	0.968	0.969
	N	18,473	19,837	20,026	16,343	15,130	13,994	13,490	12,652	11,537	10,302	6,884	1,516	761
MA	Reliability	0.861	0.942	0.945	0.957	0.963	0.967	0.964	0.971	0.972	–	–	–	–
	N	816	763	917	857	904	810	580	564	592	–	–	–	–
MD	Reliability	0.950	0.965	0.964	0.958	0.964	0.964	0.960	0.951	0.956	0.958	0.966	0.962	–
	N	455	588	429	360	480	588	615	756	593	762	402	358	–
ME	Reliability	0.946	0.964	0.965	0.963	0.960	0.958	0.954	0.954	0.953	0.953	0.957	0.968	0.973
	N	8,661	14,715	20,873	26,145	26,531	25,934	26,922	27,699	26,790	14,650	9,045	2,828	1,641
MI	Reliability	0.954	0.966	0.966	0.963	0.962	0.960	0.959	0.959	0.960	0.966	0.966	0.968	0.970
	N	212,760	237,535	252,885	256,231	266,775	271,411	256,731	244,711	233,181	124,304	112,171	54,742	19,047
MO	Reliability	0.954	0.967	0.966	0.963	0.961	0.961	0.959	0.961	0.963	0.961	0.961	0.958	0.969
	N	11,327	13,640	19,462	16,439	18,880	15,380	13,834	11,925	11,878	4,627	3,394	1,829	888
MS	Reliability	0.955	0.962	0.957	0.950	0.949	0.944	0.950	0.953	0.954	0.959	0.958	0.963	0.974
	N	22,356	26,687	27,059	21,085	21,502	19,682	22,213	24,138	23,176	12,271	11,106	3,146	379
MT	Reliability	0.951	0.963	0.963	0.959	0.956	0.955	0.953	0.951	0.949	0.957	0.955	0.962	0.965
	N	9,905	11,414	14,658	21,841	21,943	22,029	21,062	17,609	17,222	8,267	11,391	3,156	1,140
NC	Reliability	0.957	0.969	0.964	0.960	0.957	0.957	0.956	0.960	0.961	0.961	0.961	0.972	0.982
	N	40,352	55,442	58,029	65,457	64,837	63,710	58,536	54,941	54,054	4,096	2,723	1,895	705
NE	Reliability	–	–	–	0.957	0.952	0.955	0.957	0.962	0.960	0.975	0.975	0.969	–
	N	–	–	–	2,682	2,552	2,544	2,295	2,002	2,336	1,924	1,796	1,616	–
NH	Reliability	0.951	0.963	0.963	0.957	0.949	0.945	0.944	0.944	0.944	0.955	0.957	0.961	0.970
	N	4,698	11,318	15,519	16,813	17,111	17,379	15,713	14,668	13,758	5,417	4,126	1,199	653
NJ	Reliability	0.953	0.968	0.965	0.960	0.957	0.957	0.956	0.958	0.957	0.958	0.961	0.963	0.970
	N	19,093	27,577	34,994	34,160	35,505	34,145	33,519	26,977	25,344	6,263	5,267	3,542	1,784
NM	Reliability	0.935	0.953	0.959	0.960	0.960	0.959	0.959	0.960	0.958	0.957	0.959	0.954	0.952
	N	8,672	9,725	14,045	16,979	17,159	17,229	18,538	15,511	15,158	8,702	7,128	5,730	3,448

Reading														
State		Grade												
		K	1	2	3	4	5	6	7	8	9	10	11	12
NV	**Reliability**	0.948	0.960	0.961	0.961	0.959	0.957	0.953	0.951	0.950	0.952	0.958	0.965	0.970
	N	20,743	59,903	61,780	65,875	42,335	40,669	32,885	28,571	27,563	10,099	5,675	4,372	2,794
NY	**Reliability**	0.943	0.959	0.953	0.951	0.941	0.945	0.944	0.945	0.945	–	–	–	–
	N	1,352	1,323	1,404	1,106	1,009	953	992	1,016	808	–	–	–	–
OK	**Reliability**	0.933	–	–	–	0.952	0.959	0.951	0.947	–	0.940	–	–	–
	N	301	–	–	–	550	747	1,102	629	–	345	–	–	–
OR	**Reliability**	0.957	0.969	0.969	0.965	0.961	0.959	0.961	0.957	0.956	0.960	0.960	0.962	0.974
	N	3,360	5,449	7,860	8,327	9,030	8,347	9,432	9,086	8,789	5,734	5,250	2,203	875
PA	**Reliability**	0.953	0.966	0.965	0.962	0.955	0.961	0.960	0.959	0.957	0.973	0.973	0.978	–
	N	629	1,774	1,675	1,962	1,882	1,852	2,100	2,061	1,781	534	394	302	–
RI	**Reliability**	0.951	0.964	0.962	0.951	0.942	0.951	0.961	0.960	0.960	0.971	0.971	0.965	–
	N	1,430	1,578	2,017	2,049	2,075	2,521	2,693	2,887	2,597	2,613	1,893	835	–
SD	**Reliability**	0.948	0.964	0.964	0.961	0.960	0.958	0.957	0.958	0.958	0.962	0.960	0.962	0.963
	N	14,026	15,468	15,534	16,936	16,873	21,059	15,187	12,943	12,306	9,929	8,979	6,553	3,018
TN	**Reliability**	0.959	0.967	0.964	0.964	0.964	0.963	0.964	0.966	0.965	0.970	0.968	0.966	0.971
	N	36,043	35,032	35,159	35,793	32,582	36,454	32,203	31,064	30,091	22,470	20,220	13,533	7,703
TX	**Reliability**	0.955	0.967	0.966	0.962	0.950	0.965	0.958	0.950	0.950	0.902	0.892	–	–
	N	1,301	982	990	1,140	822	1,878	1,149	897	1,218	338	322	–	–
UT	**Reliability**	0.950	0.966	0.967	0.963	0.962	0.960	0.959	0.958	0.956	0.960	0.966	0.969	0.978
	N	3,762	4,591	4,860	3,654	3,868	3,583	3,808	3,932	3,608	3,138	3,018	2,397	331
VT	**Reliability**	0.945	0.963	0.965	0.966	0.962	0.960	0.956	0.957	0.959	0.959	0.962	0.970	0.968
	N	1,331	1,771	2,184	3,073	2,942	3,124	3,193	3,042	3,089	2,474	1,877	590	388
WA	**Reliability**	0.958	0.970	0.967	0.964	0.962	0.959	0.957	0.957	0.955	0.960	0.966	0.969	0.971
	N	26,414	43,070	62,844	69,895	68,801	67,763	57,735	57,709	57,391	21,262	10,736	5,221	3,121
WI	**Reliability**	0.955	0.966	0.964	0.959	0.956	0.952	0.950	0.949	0.947	0.954	0.958	0.965	0.972
	N	37,504	52,662	82,226	104,532	108,002	108,603	108,703	106,972	103,085	31,557	21,484	5,858	2,457
WY	**Reliability**	0.954	0.962	0.960	0.952	0.948	0.945	0.944	0.947	0.945	0.949	0.947	0.960	0.965
	N	15,408	21,988	22,496	22,729	22,789	22,422	19,801	17,915	17,801	9,047	6,989	2,317	666

Table D.3. Marginal Reliability of Overall RIT Scores by State and Grade—Language Usage

Language Usage												
		Grade										
State		2	3	4	5	6	7	8	9	10	11	12
AK	Reliability	–	–	–	–	–	–	–	0.914	0.893	0.900	0.915
	N	–	–	–	–	–	–	–	438	401	411	389
AL	Reliability	–	0.966	0.965	0.958	0.962	0.966	0.960	0.960	0.963	–	–
	N	–	573	638	655	671	590	581	308	300	–	–
AZ	Reliability	0.952	0.955	0.959	0.959	0.958	0.960	0.950	0.955	0.950	0.939	0.948
	N	1,199	1,632	1,572	1,598	1,459	1,242	1,116	840	658	559	469
CA	Reliability	0.972	0.969	0.967	0.965	0.965	0.966	0.965	0.963	0.964	0.971	0.975
	N	30,453	31,960	34,319	33,917	24,329	22,179	21,357	7,414	6,880	2,104	1,683
CO	Reliability	0.969	0.956	0.968	0.946	–	–	–	–	–	–	–
	N	396	532	501	467	–	–	–	–	–	–	–
CT	Reliability	0.966	0.964	0.960	0.963	0.963	0.962	0.960	0.965	0.963	0.973	0.977
	N	5,185	5,240	9,045	8,618	12,025	12,421	12,322	4,127	3,813	506	408
DE	Reliability	–	–	–	–	–	–	–	–	0.971	–	–
	N	–	–	–	–	–	–	–	–	371	–	–
FL	Reliability	0.960	0.960	0.952	0.955	0.959	0.955	0.962	0.963	–	–	–
	N	363	451	536	505	424	407	366	319	–	–	–
GA	Reliability	–	0.970	0.954	–	0.952	0.969	–	–	–	–	–
	N	–	321	303	–	408	417	–	–	–	–	–
HI	Reliability	–			–	–	–	–	0.950	0.936	0.928	0.963
	N	–			–	–	–	–	628	814	453	453
ID	Reliability	0.969	0.966	0.961	0.960	0.957	0.955	0.952	0.957	0.956	0.964	–
	N	2,488	4,366	4,501	4,812	4,622	4,344	4,236	3,340	2,970	964	–
IL	Reliability	0.969	0.966	0.962	0.959	0.961	0.960	0.960	0.967	0.966	0.972	0.982
	N	24,995	40,075	41,090	45,189	53,038	54,293	53,924	20,748	17,314	9,512	2,209
IN	Reliability	–	–	–	–	–	0.946	0.963	–	–	–	–
	N	–	–	–	–	–	489	493	–	–	–	–
KY	Reliability	0.967	0.963	0.960	0.956	0.955	0.956	0.957	0.967	0.966	0.968	–
	N	30,737	45,199	60,637	49,440	54,217	41,487	41,020	12,133	9,708	4,091	–

Language Usage												
		Grade										
State		2	3	4	5	6	7	8	9	10	11	12
LA	Reliability	0.969	0.967	0.966	0.966	0.967	0.966	0.965	0.970	0.970	–	–
	N	7,596	9,017	8,344	8,048	7,364	6,539	6,194	6,344	5,040	–	–
MD	Reliability	–	–	–	–	0.929	0.898	0.911	0.951	0.966	0.964	–
	N	–	–	–	–	320	319	333	719	387	347	–
ME	Reliability	0.964	0.964	0.959	0.954	0.951	0.951	0.952	0.955	0.960	0.968	0.969
	N	2,786	5,249	5,824	6,191	8,033	7,930	7,866	4,294	3,360	1,307	861
MI	Reliability	0.968	0.967	0.964	0.963	0.962	0.961	0.961	0.967	0.966	0.968	0.972
	N	58,348	104,048	109,915	110,979	117,329	118,678	116,178	69,621	61,266	33,420	7,721
MO	Reliability	0.967	0.965	0.963	0.958	0.960	0.954	0.957	0.959	0.956	0.955	0.966
	N	1,973	6,457	6,385	6,308	6,261	5,902	5,242	3,932	2,806	1,756	623
MS	Reliability	0.962	0.956	0.952	0.948	0.957	0.956	0.958	0.962	0.957	0.966	–
	N	10,179	9,907	10,555	10,810	13,006	13,062	12,302	5,163	5,674	2,452	–
MT	Reliability	0.966	0.965	0.961	0.959	0.958	0.954	0.950	0.957	0.955	0.960	0.965
	N	3,671	12,719	12,906	13,461	14,329	14,713	14,751	6,487	8,707	2,545	779
NC	Reliability	0.969	0.964	0.962	0.956	0.959	0.960	0.961	0.972	0.971	0.975	0.983
	N	3,362	3,437	3,527	3,312	2,941	2,971	2,503	1,067	888	705	532
NH	Reliability	0.968	0.961	0.958	0.951	0.948	0.955	0.952	0.964	0.960	0.966	–
	N	1,299	2,536	2,311	2,814	2,388	2,686	2,782	1,709	1,522	439	–
NJ	Reliability	0.968	0.965	0.959	0.955	0.955	0.958	0.956	0.962	0.962	0.963	0.971
	N	4,795	10,457	11,639	10,771	10,000	8,020	7,335	2,928	2,197	1,191	1,013
NM	Reliability	0.959	0.963	0.962	0.960	0.960	0.960	0.958	0.959	0.962	0.950	0.957
	N	4,794	8,434	8,628	8,728	9,496	6,808	6,589	4,956	3,826	2,792	1,564
NV	Reliability	0.970	0.967	0.964	0.964	0.957	0.956	0.956	0.951	0.953	0.962	0.962
	N	5,356	6,407	6,150	5,296	4,322	2,829	2,455	2,253	2,540	2,278	1,850
OR	Reliability	0.970	0.971	0.967	0.964	0.964	0.960	0.957	0.965	0.962	0.966	0.977
	N	1,498	2,300	2,329	2,319	3,103	3,096	3,084	1,962	1,929	1,065	497
PA	Reliability	0.970	0.961	0.950	0.944	0.956	0.951	0.952	–	–	–	–
	N	322	682	986	694	1,761	1,735	1,381	–	–	–	–

Language Usage												
State		Grade										
		2	3	4	5	6	7	8	9	10	11	12
RI	Reliability	–	0.967	0.957	0.957	0.943	0.951	0.955	0.961	0.953	0.956	–
	N	–	527	484	506	476	564	579	465	443	404	–
SD	Reliability	0.971	0.967	0.965	0.962	0.961	0.962	0.964	0.965	0.964	0.965	0.961
	N	1,907	8,817	8,330	14,062	8,580	7,484	7,080	7,536	6,636	4,669	2,167
TN	Reliability	0.969	0.970	0.971	0.968	0.968	0.971	0.967	0.971	0.970	0.967	0.974
	N	6,980	10,792	9,904	10,766	9,355	9,353	8,667	2,284	2,170	1,952	861
TX	Reliability	–	0.924	0.938	0.939	–	0.937	0.935	–	–	–	–
	N	–	483	451	415	–	340	354	–	–	–	–
UT	Reliability	0.969	0.967	0.963	0.962	0.961	0.962	0.959	0.964	0.968	0.969	0.979
	N	3,386	3,502	3,816	3,560	3,318	3,293	3,061	2,411	2,304	1,845	305
VT	Reliability	0.969	0.969	0.964	0.961	0.959	0.957	0.960	0.959	0.963	–	–
	N	836	1,625	1,491	1,512	1,775	1,926	1,962	1,658	1,483	–	–
WA	Reliability	0.965	0.960	0.952	0.949	0.956	0.958	0.958	0.968	0.970	0.971	0.973
	N	6,102	9,284	9,663	9,188	10,056	9,613	8,723	2,150	1,854	1,154	672
WI	Reliability	0.967	0.960	0.954	0.950	0.950	0.948	0.946	0.954	0.955	0.959	0.971
	N	9,845	19,563	20,911	22,257	27,092	27,120	26,919	9,607	6,109	2,051	706
WY	Reliability	0.967	0.959	0.951	0.947	0.945	0.948	0.947	0.953	0.950	0.962	0.963
	N	5,605	6,444	7,045	7,858	10,315	9,607	8,638	4,831	3,997	1,437	532

Table D.4. Marginal Reliability of Overall RIT Scores by State and Grade—Mathematics

Mathematics														
State		Grade												
		K	1	2	3	4	5	6	7	8	9	10	11	12
AK	Reliability	–	0.981	0.980	0.957	0.962	0.969	0.972	0.972	0.975	0.969	0.975	0.965	0.964
	N	–	350	351	3,891	3,829	6,926	8,607	12,582	12,028	1,195	495	434	402
AL	Reliability	0.965	0.959	0.963	0.948	0.954	0.961	0.962	0.970	0.969	0.967	0.978	–	–
	N	334	659	685	565	655	677	693	621	588	320	366	–	–
AZ	Reliability	0.957	0.968	0.956	0.957	0.960	0.964	0.965	0.971	0.970	0.971	0.970	0.970	0.975
	N	2,191	2,662	2,750	3,156	3,018	2,940	2,873	2,594	2,432	959	688	597	605

Mathematics														
		Grade												
State		K	1	2	3	4	5	6	7	8	9	10	11	12
CA	Reliability	0.970	0.975	0.969	0.967	0.970	0.975	0.973	0.976	0.977	0.976	0.978	0.981	0.982
	N	41,032	52,921	65,035	67,279	69,929	70,770	68,842	63,735	60,095	36,949	29,601	15,745	7,965
CO	Reliability	0.970	0.962	0.960	0.955	0.963	0.967	0.969	0.973	0.977	0.975	0.975	0.985	0.988
	N	403	863	3,465	3,743	3,786	3,647	3,893	3,821	3,890	2,542	2,262	746	347
CT	Reliability	0.966	0.971	0.969	0.957	0.961	0.968	0.969	0.973	0.976	0.979	0.980	0.982	0.981
	N	17,932	30,244	34,422	38,213	39,152	38,569	38,918	37,907	37,667	22,851	18,225	5,512	1,231
DC	Reliability	0.968	0.971	0.968	0.958	0.964	0.965	0.970	0.974	0.976	0.981	0.979	0.978	0.979
	N	9,134	8,532	8,208	7,432	6,455	6,102	6,089	5,594	5,160	11,526	8,574	5,354	1,152
DE	Reliability	0.968	0.971	0.965	0.959	0.963	0.968	0.969	0.970	0.973	0.977	0.978	0.981	0.973
	N	3,823	7,619	7,562	6,479	6,072	6,674	4,108	3,683	3,196	2,200	2,040	1,164	419
FL	Reliability	0.968	0.968	0.952	0.953	0.955	0.964	0.962	0.968	0.971	0.975	0.975	0.977	–
	N	16,542	16,464	16,561	16,674	15,431	15,137	16,374	14,249	12,631	2,591	2,525	1,125	–
GA	Reliability	0.969	0.973	0.973	0.973	–	–	0.969	0.972	0.978	–	–	–	–
	N	636	667	588	326	–	–	1,849	2,078	1,617	–	–	–	–
HI	Reliability	0.964	0.969	0.958	0.954	0.959	0.968	0.954	0.938	0.950	0.953	0.960	0.969	0.979
	N	919	1,242	1,197	1,665	1,876	1,885	2,016	2,731	2,610	2,700	1,196	533	462
ID	Reliability	0.959	0.972	0.969	0.961	0.964	0.970	0.968	0.970	0.973	0.975	0.973	0.979	0.971
	N	3,321	4,860	5,957	5,945	6,200	6,197	6,583	7,285	7,113	4,036	3,148	1,301	317
IL	Reliability	0.969	0.973	0.965	0.962	0.965	0.970	0.970	0.974	0.976	0.978	0.980	0.983	0.986
	N	160,071	211,693	306,580	329,942	335,258	332,835	338,729	330,412	326,860	81,035	59,039	31,290	9,472
IN	Reliability	–	–	–	–	0.936	0.965	0.957	0.968	0.978	0.977	0.974	0.972	–
	N	–	–	–	–	330	473	531	1,023	1,196	717	659	612	–
KY	Reliability	0.966	0.968	0.959	0.956	0.959	0.965	0.965	0.971	0.974	0.979	0.979	0.979	0.980
	N	102,530	119,042	126,819	130,406	129,867	127,215	117,161	118,577	116,433	48,497	30,425	9,953	1,199
LA	Reliability	0.968	0.971	0.965	0.960	0.964	0.970	0.968	0.973	0.976	0.978	0.978	0.978	–
	N	18,439	19,839	20,066	16,414	15,219	14,154	13,896	13,056	11,589	9,806	6,156	853	–
MA	Reliability	0.894	0.948	0.947	0.952	0.960	0.970	0.969	0.972	0.975	–	–	–	–
	N	810	763	920	853	911	809	968	974	1,265	–	–	–	–

Mathematics														
State		Grade												
		K	1	2	3	4	5	6	7	8	9	10	11	12
MD	Reliability	0.959	0.967	0.969	0.949	0.956	0.970	0.964	0.962	0.972	0.968	0.977	0.976	–
	N	526	614	447	534	625	879	829	655	528	628	392	359	–
ME	Reliability	0.960	0.969	0.965	0.956	0.959	0.966	0.965	0.970	0.974	0.974	0.977	0.981	0.983
	N	7,933	14,463	20,656	26,288	27,250	26,592	27,722	27,952	26,885	14,386	9,431	3,939	1,751
MI	Reliability	0.967	0.973	0.969	0.963	0.966	0.971	0.970	0.974	0.976	0.979	0.980	0.981	0.981
	N	211,302	237,434	252,702	260,010	267,238	272,418	258,802	247,069	234,210	121,549	111,023	58,029	18,076
MO	Reliability	0.968	0.973	0.967	0.961	0.965	0.971	0.970	0.973	0.977	0.970	0.976	0.975	–
	N	11,427	14,008	19,888	16,677	18,931	15,354	13,834	12,763	11,966	4,424	3,074	1,845	–
MS	Reliability	0.967	0.963	0.956	0.946	0.952	0.960	0.963	0.969	0.972	0.974	0.976	0.975	0.980
	N	22,645	26,971	28,022	21,773	21,863	20,046	22,314	24,379	23,293	12,397	7,302	2,655	447
MT	Reliability	0.965	0.967	0.962	0.956	0.959	0.966	0.966	0.969	0.972	0.975	0.977	0.978	0.980
	N	9,600	10,992	14,658	21,807	21,949	21,974	21,603	18,131	17,653	8,613	11,336	3,392	1,127
NC	Reliability	0.966	0.971	0.959	0.957	0.961	0.969	0.969	0.976	0.980	0.981	0.982	0.985	0.991
	N	58,406	64,717	66,748	69,952	64,997	61,517	60,102	55,490	53,966	3,457	2,484	1,765	695
NE	Reliability	–	–	–	0.953	0.960	0.964	0.966	0.969	0.972	0.982	0.983	0.982	–
	N	–	–	–	2,663	2,551	2,472	2,112	1,999	2,201	1,922	1,768	1,622	–
NH	Reliability	0.962	0.966	0.959	0.948	0.951	0.959	0.960	0.965	0.968	0.977	0.978	0.981	0.983
	N	4,722	11,292	15,993	17,096	17,257	17,597	16,589	15,931	14,215	6,174	4,542	1,520	635
NJ	Reliability	0.965	0.971	0.967	0.961	0.965	0.970	0.972	0.976	0.979	0.977	0.979	0.980	0.979
	N	19,250	30,748	40,603	37,978	39,372	42,105	42,809	36,181	29,094	8,394	6,816	4,669	2,056
NM	Reliability	0.958	0.962	0.962	0.952	0.957	0.964	0.966	0.971	0.972	0.972	0.974	0.971	0.969
	N	10,254	11,545	15,467	16,592	16,615	17,079	18,975	15,856	14,969	7,934	6,559	5,243	2,880
NV	Reliability	0.964	0.968	0.962	0.961	0.962	0.967	0.965	0.969	0.972	0.971	0.976	0.979	0.981
	N	19,321	61,466	60,810	62,443	41,995	40,623	33,567	29,208	27,480	7,458	4,021	3,222	2,750
NY	Reliability	0.965	0.965	0.964	0.948	0.947	0.960	0.958	0.965	0.967	–	–	–	–
	N	2,260	2,463	2,425	1,137	1,009	929	1,065	1,077	892	–	–	–	–
OK	Reliability	0.952	–	–	0.931	0.954	0.961	0.961	0.974	0.980	–	–	–	–
	N	301	–	–	307	545	763	1,409	1,039	1,533	–	–	–	–

Mathematics														
		Grade												
State		K	1	2	3	4	5	6	7	8	9	10	11	12
OR	**Reliability**	0.965	0.974	0.968	0.963	0.965	0.969	0.971	0.974	0.976	0.976	0.975	0.976	0.980
	N	4,740	6,138	8,345	8,557	9,213	8,876	9,268	9,048	9,195	5,673	5,098	3,286	1,349
PA	**Reliability**	0.961	0.970	0.969	0.964	0.961	0.972	0.972	0.976	0.977	0.982	0.981	–	–
	N	629	1,755	1,664	1,994	1,909	1,801	2,111	2,036	2,282	431	346	–	–
RI	**Reliability**	0.963	0.963	0.962	0.945	0.944	0.960	0.961	0.972	0.978	0.977	0.978	0.979	–
	N	1,774	1,897	2,408	2,188	2,165	2,456	2,401	2,529	2,505	2,444	1,778	878	–
SD	**Reliability**	0.963	0.969	0.969	0.962	0.965	0.969	0.969	0.973	0.976	0.978	0.979	0.981	0.981
	N	13,991	15,475	15,534	17,080	16,941	20,977	15,560	13,310	12,694	10,892	9,816	6,599	3,038
TN	**Reliability**	0.969	0.971	0.960	0.961	0.966	0.970	0.971	0.976	0.978	0.980	0.981	0.978	0.980
	N	35,967	35,066	35,348	35,821	32,601	36,991	32,202	30,929	29,724	22,474	19,340	14,031	8,754
TX	**Reliability**	0.967	0.973	0.963	0.960	0.948	0.969	0.966	0.970	0.970	0.974	0.973	–	–
	N	1,283	972	992	1,113	827	1,807	1,177	951	1,293	425	372	–	–
UT	**Reliability**	0.965	0.972	0.969	0.962	0.963	0.969	0.967	0.976	0.975	0.978	0.981	0.980	–
	N	3,816	4,738	5,103	3,718	3,895	3,562	3,752	3,969	3,629	3,148	2,876	2,218	–
VT	**Reliability**	0.957	0.966	0.964	0.959	0.959	0.965	0.964	0.969	0.976	0.976	0.979	0.981	0.982
	N	1,479	1,925	2,391	3,335	3,214	3,389	3,533	3,094	3,184	2,493	2,001	832	387
WA	**Reliability**	0.970	0.974	0.967	0.961	0.964	0.969	0.968	0.972	0.975	0.975	0.978	0.976	0.978
	N	28,103	45,298	65,371	71,340	69,805	69,311	60,233	57,271	50,942	18,334	11,954	6,356	3,264
WI	**Reliability**	0.968	0.970	0.963	0.959	0.962	0.967	0.967	0.972	0.974	0.976	0.977	0.980	0.984
	N	41,481	59,507	86,262	106,899	109,522	109,188	110,028	106,208	103,034	31,391	21,649	5,783	1,296
WY	**Reliability**	0.967	0.967	0.951	0.950	0.954	0.962	0.960	0.966	0.968	0.971	0.973	0.976	0.982
	N	15,424	21,916	22,403	22,729	22,862	22,672	19,913	18,075	17,395	9,678	6,999	2,951	875

Table D.5. Marginal Reliability of Overall RIT Scores by State and Grade—Science

State		Science									
		Grade									
		3	4	5	6	7	8	9	10	11	12
AR	**Reliability**	0.917	0.918	0.924	0.922	0.924	0.936	0.934	0.944	0.931	–
	N	5,227	6,398	7,475	7,475	7,597	7,447	1,947	923	466	–
CA	**Reliability**	0.924	0.925	0.918	0.930	0.936	0.934	0.939	0.944	0.932	0.925
	N	1,475	1,736	15,237	8,507	8,754	19,599	3,214	2,388	1,002	547
CO	**Reliability**	–	0.893	0.904	0.925	0.927	0.936	0.922	0.926	0.947	–
	N	–	3,678	4,688	7,335	7,113	7,684	2,763	2,605	661	–
CT	**Reliability**	–	0.896	0.905	0.907	0.928	0.929	0.932	0.938	0.936	–
	N	–	496	3,083	3,430	3,662	3,833	1,634	1,530	1,170	–
DC	**Reliability**	–	–	–	0.883	0.923	0.915	–	–	–	–
	N	–	–	–	446	459	454	–	–	–	–
DE	**Reliability**	–	–	–	–	–	–	0.907	–	–	–
	N	–	–	–	–	–	–	346	–	–	–
GA	**Reliability**	0.932	0.933	0.939	0.941	0.943	0.951	–	–	–	–
	N	8,108	7,425	7,791	6,892	6,684	6,693	–	–	–	–
IA	**Reliability**	0.891	0.890	0.896	0.899	0.905	0.912	0.926	0.934	0.933	0.947
	N	2,603	3,524	5,134	6,301	8,227	8,540	4,438	4,444	3,407	577
IL	**Reliability**	0.930	0.921	0.928	0.928	0.932	0.933	0.920	0.940	0.940	–
	N	12,796	15,088	18,895	21,916	22,866	21,846	902	504	360	–
KS	**Reliability**	0.909	0.906	0.913	0.913	0.916	0.921	0.920	0.930	0.932	0.936
	N	507	972	2,576	4,313	4,843	4,820	1,611	1,400	1,145	498
KY	**Reliability**	0.910	0.904	0.908	0.910	0.920	0.919	0.945	–	–	–
	N	3,665	6,274	3,270	4,972	7,245	4,393	1,501	–	–	–
MA	**Reliability**	–	0.921	0.931	–	–	0.944	–	–	–	–
	N	–	312	2,775	–	–	1,704	–	–	–	–
MD	**Reliability**	–	–	0.923	0.936	0.936	0.951	0.909	–	–	–
	N	–	–	349	646	650	633	440	–	–	–
MI	**Reliability**	0.926	0.923	0.928	0.927	0.936	0.941	0.948	0.954	0.954	0.954
	N	45,092	55,427	54,543	65,537	60,461	58,554	13,932	11,876	4,466	1,059

Science											
		Grade									
State		3	4	5	6	7	8	9	10	11	12
MO	Reliability	–	–	0.907	0.930	0.935	0.935	–	–	–	–
	N	–	–	1,450	1,327	1,288	1,238	–	–	–	–
MT	Reliability	0.906	0.896	0.916	0.912	0.910	0.912	0.927	0.924	–	–
	N	583	737	702	703	808	988	363	417	–	–
NC	Reliability	–	–	–	0.904	–	–	–	–	–	–
	N	–	–	–	311	–	–	–	–	–	–
NJ	Reliability	0.899	0.907	0.914	0.914	0.931	0.927	–	–	–	–
	N	1,091	1,134	1,053	1,657	1,860	1,946	–	–	–	–
NV	Reliability	0.926	0.915	0.916	0.914	0.922	0.930	0.913	–	–	–
	N	674	926	1,440	1,694	1,879	1,813	581	–	–	–
NY	Reliability	–	–	–	0.902	0.920	0.926	–	–	–	–
	N	–	–	–	634	981	430	–	–	–	–
OH	Reliability	0.873	0.876	0.887	0.871	0.878	0.878	–	–	–	–
	N	747	938	1,036	1,129	1,083	910	–	–	–	–
OK	Reliability	–	–	0.917	0.920	0.938	0.925	–	–	–	–
	N	–	–	485	393	442	362	–	–	–	–
OR	Reliability	–	0.909	–	0.910	0.927	0.922	0.938	0.924	–	–
	N	–	312	–	373	354	401	355	357	–	–
RI	Reliability	0.924	0.911	0.924	0.892	0.917	0.927	–	–	–	–
	N	442	465	495	552	483	428	–	–	–	–
SD	Reliability	–	–	–	0.919	0.903	0.928	–	–	–	–
	N	–	–	–	1,274	1,284	1,172	–	–	–	–
WA	Reliability	0.925	0.916	0.916	0.910	0.921	0.931	0.933	0.932	–	–
	N	1,427	1,927	3,924	4,008	5,673	4,312	696	622	–	–
WI	Reliability	–	0.893	0.892	0.901	0.890	0.883	–	–	–	–
	N	–	1,037	1,121	1,295	1,219	1,319	–	–	–	–

Table D.6. Marginal Reliability of Overall RIT Scores by Instructional Area and State—Reading K–2

State	N	Reliability by Instructional Area: Foundational Skills	Language & Writing	Literature & Informational	Vocabulary Use & Functions
AK	881	0.927	0.923	0.919	0.917
AL	1,268	0.887	0.866	0.863	0.874
AZ	5,381	0.883	0.860	0.856	0.842
CA	101,748	0.922	0.904	0.899	0.901
CO	1,105	0.912	0.898	0.894	0.896
CT	56,055	0.920	0.908	0.911	0.910
DC	21,603	0.910	0.903	0.907	0.905
DE	12,356	0.915	0.901	0.901	0.899
FL	33,489	0.907	0.892	0.895	0.891
GA	1,720	0.914	0.897	0.902	0.895
HI	1,823	0.907	0.904	0.904	0.902
ID	10,714	0.924	0.908	0.905	0.909
IL	389,466	0.915	0.903	0.902	0.901
KY	237,151	0.913	0.885	0.882	0.883
LA	46,144	0.917	0.901	0.903	0.902
MA	1,675	0.848	0.817	0.815	0.843
MD	1,193	0.920	0.903	0.904	0.910
ME	36,033	0.911	0.899	0.901	0.903
MI	578,405	0.918	0.905	0.905	0.905
MO	34,071	0.920	0.909	0.910	0.908
MS	53,774	0.924	0.904	0.898	0.896
MT	26,139	0.917	0.897	0.893	0.896
NC	98,358	0.912	0.895	0.903	0.898
NH	20,774	0.916	0.895	0.892	0.895
NJ	65,442	0.925	0.916	0.915	0.912
NM	24,877	0.910	0.894	0.890	0.888
NV	84,378	0.891	0.867	0.870	0.873
NY	3,093	0.895	0.887	0.891	0.884
OK	645	0.902	0.878	0.879	0.883
OR	10,492	0.910	0.901	0.899	0.904
PA	3,467	0.918	0.907	0.907	0.907
RI	3,815	0.923	0.915	0.911	0.910
SD	40,173	0.921	0.903	0.899	0.899
TN	73,141	0.914	0.894	0.892	0.892
TX	2,465	0.914	0.899	0.903	0.906
UT	10,602	0.920	0.901	0.894	0.898
VT	4,366	0.907	0.899	0.896	0.899
WA	88,500	0.915	0.903	0.904	0.906
WI	110,067	0.914	0.901	0.900	0.899
WV	584	0.903	0.885	0.894	0.892
WY	38,418	0.916	0.887	0.886	0.880

Table D.7. Marginal Reliability of Overall RIT Scores by Instructional Area and State—Reading 2–12

State	N	Reliability by Instructional Area		
		Literary Text	Informational Text	Vocabulary
AK	50,540	0.874	0.876	0.871
AL	5,066	0.885	0.889	0.891
AZ	22,154	0.886	0.890	0.891
CA	536,531	0.912	0.914	0.916
CO	30,083	0.913	0.915	0.914
CT	273,491	0.905	0.907	0.907
DC	47,988	0.896	0.898	0.897
DE	40,956	0.900	0.902	0.901
FL	113,920	0.914	0.914	0.911
GA	2,156	0.915	0.916	0.912
HI	18,506	0.879	0.880	0.882
ID	46,608	0.901	0.901	0.903
IL	2,431,987	0.913	0.914	0.914
IN	4,554	0.912	0.911	0.906
KS	735	0.873	0.873	0.882
KY	937,908	0.906	0.908	0.908
LA	114,805	0.923	0.924	0.924
MA	5,289	0.868	0.875	0.888
MD	5,401	0.907	0.908	0.908
ME	196,421	0.900	0.902	0.903
MI	1,965,665	0.903	0.905	0.907
MN	756	0.921	0.922	0.924
MO	109,434	0.921	0.921	0.921
MS	181,345	0.912	0.911	0.909
MT	155,600	0.899	0.900	0.902
NC	426,432	0.908	0.909	0.909
NE	19,747	0.898	0.896	0.897
NH	117,607	0.897	0.899	0.900
NJ	222,986	0.914	0.913	0.910
NM	133,159	0.905	0.907	0.908
NV	318,901	0.907	0.911	0.913
NY	7,109	0.903	0.907	0.910
OK	4,522	0.871	0.871	0.875
OR	73,253	0.909	0.910	0.912
PA	13,556	0.900	0.900	0.898
RI	21,607	0.889	0.889	0.891
SC	489	0.831	0.818	0.835
SD	128,638	0.898	0.900	0.901
TN	295,298	0.928	0.928	0.929
TX	8,598	0.908	0.911	0.912
UT	33,948	0.916	0.916	0.918
VA	1,978	0.916	0.913	0.911
VT	24,712	0.903	0.904	0.907
WA	463,606	0.907	0.910	0.910

State	N	Reliability by Instructional Area		
		Literary Text	Informational Text	Vocabulary
WI	764,291	0.900	0.902	0.902
WV	1,100	0.860	0.868	0.867
WY	163,966	0.909	0.909	0.910

Table D.8. Marginal Reliability of Overall RIT Scores by Instructional Area and State—Language Usage 2–12

State	N	Reliability by Instructional Area		
		Writing	Language: Understand, Edit for Grammar, Usage	Language: Understand, Edit for Mechanics
AK	1,639	0.824	0.763	0.791
AL	4,646	0.924	0.921	0.924
AZ	12,344	0.925	0.930	0.934
CA	216,595	0.938	0.937	0.940
CO	2,671	0.936	0.935	0.936
CT	73,710	0.935	0.925	0.930
DC	1,412	0.926	0.922	0.920
DE	1,785	0.926	0.905	0.912
FL	3,814	0.930	0.928	0.929
GA	1,953	0.923	0.919	0.917
HI	3,387	0.938	0.934	0.934
ID	36,846	0.932	0.925	0.929
IL	362,387	0.930	0.924	0.928
IN	1,471	0.909	0.901	0.904
KS	351	0.887	0.887	0.901
KY	348,865	0.929	0.925	0.927
LA	64,842	0.933	0.933	0.937
MD	3,289	0.897	0.864	0.872
ME	53,701	0.926	0.913	0.922
MI	907,503	0.934	0.928	0.933
MN	482	0.948	0.943	0.940
MO	47,645	0.932	0.924	0.930
MS	93,389	0.924	0.926	0.925
MT	105,068	0.926	0.919	0.923
NC	25,245	0.940	0.935	0.935
NH	20,672	0.932	0.922	0.930
NJ	70,346	0.921	0.910	0.916
NM	66,615	0.932	0.928	0.931
NV	41,736	0.938	0.935	0.940
NY	309	0.939	0.924	0.920
OK	852	0.887	0.872	0.878
OR	23,182	0.935	0.928	0.933
PA	7,805	0.919	0.912	0.911
RI	4,498	0.919	0.903	0.911
SC	393	0.868	0.830	0.846
SD	77,268	0.932	0.928	0.932

State	N	Reliability by Instructional Area		
		Writing	Language: Understand, Edit for Grammar, Usage	Language: Understand, Edit for Mechanics
TN	73,084	0.936	0.939	0.937
TX	2,719	0.911	0.891	0.902
UT	30,801	0.942	0.938	0.940
VA	1,837	0.921	0.904	0.909
VT	14,661	0.935	0.928	0.933
WA	68,459	0.924	0.915	0.922
WI	172,180	0.921	0.912	0.918
WV	579	0.913	0.908	0.901
WY	66,309	0.922	0.910	0.916

Table D.9. Marginal Reliability of Overall RIT Scores by Instructional Area and State—Mathematics K–2

State	N	Reliability by Instructional Area			
		Operations & Algebraic Thinking	Number & Operations	Measurement & Data	Geometry
AK	876	0.944	0.944	0.941	0.942
AL	1,549	0.918	0.922	0.907	0.921
AZ	5,706	0.915	0.912	0.898	0.908
CA	102,663	0.929	0.930	0.920	0.930
CO	1,065	0.928	0.929	0.921	0.931
CT	67,879	0.931	0.934	0.928	0.935
DC	22,167	0.931	0.931	0.920	0.934
DE	13,952	0.923	0.926	0.914	0.928
FL	33,340	0.917	0.916	0.906	0.921
GA	1,755	0.920	0.923	0.913	0.913
HI	2,324	0.916	0.907	0.896	0.919
ID	11,223	0.928	0.933	0.921	0.931
IL	428,375	0.926	0.927	0.918	0.929
KY	237,379	0.920	0.920	0.902	0.914
LA	45,868	0.929	0.931	0.918	0.927
MA	1,674	0.883	0.874	0.864	0.869
MD	1,395	0.935	0.939	0.933	0.938
ME	34,643	0.922	0.925	0.916	0.926
MI	574,980	0.931	0.934	0.924	0.933
MO	34,156	0.932	0.933	0.924	0.933
MS	54,682	0.926	0.926	0.914	0.924
MT	24,679	0.922	0.923	0.908	0.918
NC	130,912	0.922	0.921	0.911	0.922
NH	21,028	0.917	0.919	0.906	0.914
NJ	70,747	0.929	0.934	0.928	0.936
NM	29,310	0.925	0.928	0.914	0.921
NV	83,830	0.902	0.906	0.891	0.908
NY	6,170	0.927	0.930	0.923	0.932
OK	763	0.900	0.901	0.878	0.884
OR	12,344	0.923	0.922	0.913	0.925

State	N	Reliability by Instructional Area			
		Operations & Algebraic Thinking	Number & Operations	Measurement & Data	Geometry
PA	3,447	0.917	0.925	0.916	0.925
RI	5,032	0.933	0.936	0.932	0.935
SD	40,352	0.927	0.927	0.921	0.930
TN	72,976	0.924	0.921	0.910	0.920
TX	2,359	0.924	0.924	0.915	0.919
UT	10,999	0.926	0.928	0.919	0.927
VT	4,711	0.918	0.919	0.905	0.916
WA	94,429	0.926	0.931	0.922	0.930
WI	121,971	0.924	0.924	0.916	0.926
WV	583	0.890	0.910	0.898	0.896
WY	38,174	0.917	0.915	0.899	0.915

Table D.10. Marginal Reliability of Overall RIT Scores by Instructional Area and State—Mathematics 2–12

State	N	Reliability by Instructional Area					
		Algebraic Thinking	Number & Operations	Measurement & Data	Geometry	The Real & Complex Number Systems	Statistics & Probability
AK	50,510	0.922	0.907	0.901	0.916	0.899	0.907
AL	4,836	0.922	0.877	0.883	0.917	0.894	0.902
AZ	21,759	0.929	0.890	0.887	0.926	0.890	0.897
CA	547,912	0.937	0.919	0.921	0.933	0.908	0.915
CO	32,344	0.933	0.913	0.911	0.930	0.895	0.909
CT	292,965	0.933	0.906	0.906	0.928	0.907	0.915
DC	67,245	0.930	0.899	0.897	0.923	0.907	0.916
DE	41,087	0.931	0.913	0.915	0.925	0.901	0.916
FL	113,250	0.924	0.904	0.904	0.918	0.885	0.896
GA	6,598	0.906	0.917	0.918	0.906	0.901	0.910
HI	18,710	0.928	0.906	0.908	0.926	0.850	0.869
ID	51,041	0.933	0.911	0.911	0.931	0.897	0.905
IL	2,425,293	0.934	0.911	0.912	0.930	0.906	0.911
IN	6,032	0.913	0.900	0.899	0.906	0.893	0.903
KS	686	0.917	0.890	0.896	0.908	0.823	0.833
KY	941,359	0.933	0.901	0.905	0.928	0.900	0.906
LA	113,862	0.933	0.902	0.901	0.927	0.904	0.912
MA	6,768	0.926	0.908	0.901	0.931	0.901	0.906
MD	5,836	0.915	0.899	0.898	0.909	0.893	0.901
ME	200,626	0.928	0.899	0.901	0.923	0.898	0.907
MI	1,976,416	0.932	0.906	0.908	0.927	0.906	0.913
MN	1,364	0.930	0.905	0.916	0.926	0.930	0.936
MO	110,235	0.932	0.901	0.905	0.925	0.904	0.910
MS	179,742	0.929	0.887	0.888	0.919	0.889	0.898
MT	158,258	0.933	0.899	0.900	0.929	0.899	0.905
NC	433,397	0.936	0.916	0.916	0.932	0.911	0.919
NE	19,310	0.931	0.874	0.893	0.928	0.909	0.925
NH	122,544	0.929	0.895	0.896	0.924	0.890	0.896

State	N	Reliability by Instructional Area					
		Algebraic Thinking	Number & Operations	Measurement & Data	Geometry	The Real & Complex Number Systems	Statistics & Probability
NJ	269,347	0.928	0.913	0.914	0.924	0.907	0.915
NM	130,658	0.926	0.896	0.894	0.922	0.892	0.900
NV	310,538	0.938	0.916	0.915	0.936	0.891	0.898
NY	7,343	0.926	0.894	0.896	0.923	0.893	0.896
OK	6,152	0.922	0.860	0.864	0.915	0.914	0.926
OR	76,443	0.939	0.913	0.915	0.936	0.902	0.911
PA	13,801	0.923	0.905	0.907	0.919	0.908	0.917
RI	20,633	0.922	0.889	0.885	0.917	0.899	0.912
SC	365	0.861	0.848	0.859	0.853	0.754	0.811
SD	131,555	0.936	0.906	0.907	0.932	0.911	0.918
TN	296,361	0.938	0.905	0.901	0.928	0.915	0.916
TX	8,926	0.932	0.905	0.912	0.929	0.886	0.899
UT	33,655	0.942	0.912	0.914	0.940	0.915	0.924
VA	2,081	0.924	0.895	0.902	0.925	0.893	0.905
VT	26,546	0.933	0.895	0.898	0.930	0.903	0.910
WA	463,422	0.930	0.908	0.910	0.927	0.895	0.905
WI	770,940	0.931	0.905	0.906	0.928	0.896	0.907
WV	1,077	0.912	0.891	0.884	0.915	0.910	0.925
WY	165,797	0.929	0.903	0.904	0.922	0.883	0.891

Table D.11. Marginal Reliability of Overall RIT Scores by Instructional Area and State—Science 3–12

State	N	Reliability by Instructional Area		
		Life Science	Physical Science	Earth & Space Science
AR	45,034	0.856	0.848	0.834
CA	62,513	0.858	0.844	0.832
CO	36,749	0.840	0.834	0.819
CT	19,086	0.852	0.831	0.817
DC	1,372	0.797	0.764	0.752
DE	1,354	0.793	0.771	0.772
FL	336	0.757	0.754	0.743
GA	43,593	0.881	0.856	0.865
HI	438	0.880	0.873	0.880
IA	47,217	0.831	0.822	0.819
ID	1,121	0.832	0.823	0.826
IL	115,402	0.857	0.840	0.838
IN	617	0.715	0.771	0.729
KS	22,705	0.825	0.820	0.809
KY	31,761	0.842	0.847	0.834
MA	5,437	0.868	0.852	0.841
MD	3,085	0.874	0.857	0.863
ME	424	0.814	0.814	0.808
MI	371,595	0.867	0.857	0.854
MN	455	0.736	0.767	0.754

State	N	Reliability by Instructional Area		
		Life Science	Physical Science	Earth & Space Science
MO	5,656	0.824	0.823	0.817
MT	5,369	0.841	0.835	0.839
NC	663	0.833	0.803	0.822
ND	657	0.767	0.714	0.745
NH	1,047	0.829	0.820	0.818
NJ	9,369	0.849	0.831	0.820
NV	9,453	0.841	0.835	0.823
NY	2,624	0.830	0.827	0.793
OH	5,867	0.800	0.785	0.780
OK	1,919	0.823	0.837	0.816
OR	2,669	0.842	0.831	0.823
PA	368	0.825	0.790	0.812
RI	2,865	0.836	0.851	0.838
SD	4,168	0.832	0.816	0.819
TX	725	0.870	0.887	0.852
VA	755	0.885	0.859	0.863
WA	23,053	0.832	0.826	0.822
WI	6,203	0.798	0.787	0.786

Appendix E: Concurrent Validity by State

Table E.1. Concurrent Validity of MAP Growth Tests as Measured by Pearson Product-Moment Correlations between RIT Scores and State Summative Test Scores

State	State Test	Admin.*		Grade								
				3	4	5	6	7	8	9**	10**	11**
Reading												
AK	AMP ELA	Spring 2015	*r*	0.82	0.83	0.85	0.84	0.83	0.83	0.80	0.81	–
			N	1,748	1,639	1,764	1,599	1,633	1,673	980	780	–
AR	ACTAAP Reading	Spring 2009*	*r*	0.77	0.79	0.83	0.82	0.80	0.78	–	–	–
			N	1,868	1,743	1,307	1,056	1,164	1,144	–	–	–
AZ	AzMERIT ELA/ Reading	Spring 2015	*r*	0.83	0.84	0.83	0.82	0.81	0.82	–	–	–
			N	1,779	1,572	1,651	1,501	1,493	1,602	–	–	–
FL	FSA ELA	Spring 2016	*r*	0.80	0.82	0.81	0.79	0.76	0.76	–	–	–
			N	5,824	5,479	5,293	4,784	3,905	3,710	–	–	–
GA	Milestones ELA/ Reading	Spring 2015	*r*	0.83	0.81	0.83	0.81	0.80	0.79	–	–	–
			N	1,615	1,521	1,514	1,497	1,505	1,407	–	–	–
IA	ITBS Reading	Fall 2007–2009	*r*	0.68	0.74	0.75	0.77	0.76	0.75	0.69	0.71	0.68
			N	1,104	1,017	1,074	861	993	1,019	1,651	1,196	968
IN	ISTEP+ Reading	Spring 2016	*r*	0.85	0.82	0.81	0.8	0.80	0.79	–	–	–
			N	8,969	8,684	15,069	8,797	7,877	7,251	–	–	–
KS	KAP ELA	Spring 2015	*r*	0.85	0.84	0.84	0.83	0.83	0.84	–	0.83	–
			N	3,339	3,099	3,156	2,979	2,415	2,413	–	815	–
KY	K-PREP Reading	Spring 2015	*r*	0.73	0.72	0.70	0.74	0.74	0.74	–	–	–
			N	9,619	10,165	10,013	10,440	10,283	10,038	–	–	–
LA	LEAP ELA	Spring 2016	*r*	0.76	0.79	0.75	0.73	0.75	0.76	–	–	–
			N	2,756	2,756	2,605	2,632	2,461	2,501	–	–	–
MA	MCAS ELA/Reading	Spring 2018	*r*	0.78	0.79	0.78	0.77	0.78	0.77	–	–	–
			N	2,389	2,650	2,516	2,045	1,414	1,218	–	–	–
MI	M-STEP ELA/ Reading	Spring 2016	*r*	0.80	0.81	0.82	0.81	0.80	0.80	–	–	–
			N	4,824	4,599	4,613	4,732	4,571	4,530	–	–	–

State	State Test	Admin.*		Grade 3	4	5	6	7	8	9**	10**	11**
MN	MCA-III Reading	Spring 2015	*r*	0.86	0.85	0.85	0.85	0.86	0.85	–	–	–
			N	6,706	6,460	6,513	5,964	5,886	5,315	–	–	–
MS	Mississippi Assessment Program ELA	Spring 2016	*r*	0.80	0.78	0.82	0.82	0.80	0.78	–	–	–
			N	2,567	2,277	2,285	2,323	2,088	2,032	–	–	–
NC	EOG ELA/Reading	Spring 2013	*r*	0.82	0.79	0.80	0.78	0.77	0.78	–	–	–
			N	6,503	7,115	6,898	4,623	4,495	4,395	–	–	–
NE	NeSA Reading	Spring 2015	*r*	0.81	0.80	0.81	0.81	0.82	0.79	–	–	–
			N	1,675	1,635	1,698	1,617	1,815	1,333	–	–	–
NY	NYSTP ELA/Reading	Spring 2013	*r*	0.73	0.74	0.72	0.70	0.70	0.71	–	–	–
			N	1,027	1,070	1,047	1,026	1,028	958	–	–	–
OH	OST ELA	Spring 2016	*r*	0.73	0.77	0.76	0.76	0.77	0.74	–	–	–
			N	5,421	4,991	4,642	4,636	4,450	4,573	–	–	–
PA	PSSA ELA/Reading	Spring 2015	*r*	0.80	0.77	0.78	0.78	0.72	0.75	–	–	–
			N	1,207	1,262	1,262	846	854	821	–	–	–
SC	SC READY ELA/Reading	Spring 2017	*r*	0.85	0.84	0.82	0.83	0.82	0.83	–	–	–
			N	15,018	16,203	15,783	15,333	14,928	14,245	–	–	–
TX	STAAR Reading	Spring 2017	*r*	0.78	0.83	0.84	0.80	0.80	0.73	–	–	–
			N	21,354	22,182	21,296	20,301	17,464	9,725	–	–	–
VA	SOL Reading	Spring 2014	*r*	0.76	0.76	0.75	0.77	0.75	0.81	–	–	–
			N	1,573	1,573	1,556	1,249	1,179	258	–	–	–
WI	Forward ELA	Spring 2016	*r*	0.79	0.79	0.78	0.81	0.81	0.80	–	–	–
			N	4,282	4,127	4,616	4,686	4,697	4,377	–	–	–
WY	PAWS ELA	Spring 2016	*r*	0.81	0.81	0.82	0.83	0.81	0.80	–	–	–
			N	2,740	2,542	2,597	2,406	2,497	2,362	–	–	–
Mathematics												
AK	AMP Mathematics	Spring 2015	*r*	0.81	0.87	0.84	0.8	0.82	0.81	0.71	0.70	–
			N	1,744	1,644	1,770	1,603	1,643	1677	1055	789	–
AR	ACTAAP Mathematics	Spring 2009*	*r*	0.80	0.82	0.87	0.85	0.87	0.87	–	–	–
			N	1,787	1,712	1,286	1,054	1,155	1,135	–	–	–

State	State Test	Admin.*		Grade 3	4	5	6	7	8	9**	10**	11**
AZ	AzMERIT Mathematics	Spring 2015	*r*	0.84	0.88	0.87	0.85	0.88	0.89	–	–	–
			N	1,776	1,573	1,652	1,503	1,559	1,855	–	–	–
FL	FSA Mathematics	Spring 2016	*r*	0.82	0.86	0.88	0.85	0.81	0.75	–	–	–
			N	5,806	5,516	5,267	4,677	3,491	2,352	–	–	–
GA	Milestones Mathematics	Spring 2015	*r*	0.84	0.86	0.87	0.85	0.85	0.83	–	–	–
			N	1,620	1,546	1,553	1,470	1,506	1,442	–	–	–
IA	ITBS Mathematics	Fall 2007–2009	*r*	0.76	0.81	0.80	0.80	0.84	0.83	0.73	0.76	0.73
			N	940	876	1,075	860	991	968	1651	1201	975
IN	ISTEP+ Mathematics	Spring 2016	*r*	0.89	0.89	0.90	0.89	0.87	0.88	–	–	–
			N	9,010	8,721	15,135	8,877	7,870	7,263	–	–	–
KS	KAP Mathematics	Spring 2015	*r*	0.85	0.87	0.88	0.84	0.83	0.79	–	0.79	–
			N	3,359	3,135	3,203	3,014	2,547	2,491	–	867	–
KY	K-PREP Mathematics	Spring 2015	*r*	0.78	0.80	0.81	0.80	0.81	0.80	–	–	–
			N	9,635	10,164	10,011	10,449	10,312	10,004	–	–	–
LA	LEAP Mathematics	Spring 2016	*r*	0.84	0.85	0.85	0.84	0.84	0.83	–	–	–
			N	2,743	2,772	2,635	2,656	2,468	2,444	–	–	–
MA	MCAS Mathematics	Spring 2018	*r*	0.82	0.85	0.86	0.86	0.85	0.83	–	–	–
			N	2,649	2,858	2,835	2,436	1,381	1,172	–	–	–
MI	M-STEP Mathematics	Spring 2016	*r*	0.82	0.85	0.86	0.89	0.87	0.87	–	–	–
			N	4,794	4,579	4,623	4,742	4,608	4,606	–	–	–
MN	MCA-III Mathematics	Spring 2015	*r*	0.90	0.90	0.90	0.92	0.91	0.89	–	–	–
			N	6,737	6,458	6,566	5,876	5,535	4,493	–	–	–
MS	Mississippi Assessment Program Mathematics	Spring 2016	*r*	0.85	0.88	0.86	0.87	0.85	0.82	–	–	–
			N	2,581	2,274	2,282	2,313	2,092	1,960	–	–	–
NC	EOG Mathematics	Spring 2013	*r*	0.82	0.84	0.85	0.85	0.86	0.85	–	–	–
			N	6,527	7,033	6,823	4,588	4,529	4,474	–	–	–
NE	NeSA Mathematics	Spring 2015	*r*	0.83	0.84	0.86	0.84	0.86	0.85	–	–	–
			N	1,674	1,635	1,700	1,618	1,821	1,365	–	–	–

State	State Test	Admin.*		Grade 3	4	5	6	7	8	9**	10**	11**
NY	NYSTP Mathematics	Spring 2013	*r*	0.75	0.76	0.76	0.74	0.76	0.77	–	–	–
			N	1,025	1,074	1,048	1,018	1,029	956	–	–	–
OH	OST Mathematics	Spring 2016	*r*	0.77	0.78	0.80	0.80	0.82	0.73	–	–	–
			N	5,189	5,035	4,388	4,418	4,376	3,804	–	–	–
PA	PSSA Mathematics	Spring 2015	*r*	0.85	0.87	0.88	0.86	0.87	0.85	–	–	–
			N	1,210	1,265	1,266	850	854	830	–	–	–
SC	SC READY Mathematics	Spring 2017	*r*	0.86	0.85	0.85	0.86	0.87	0.87	–	–	–
			N	15,037	16,285	15,796	15,366	14,953	14,118	–	–	–
TX	STAAR Mathematics	Spring 2017	*r*	0.77	0.8	0.77	0.77	0.76	0.73	–	–	–
			N	21,045	21,951	21,075	19,463	17,149	11,297	–	–	–
VA	SOL Mathematics	Spring 2014	*r*	0.79	0.81	0.79	0.76	0.77	0.79	–	–	–
			N	1,550	1,550	1,522	1,229	1,052	722	–	–	–
WI	Forward Mathematics	Spring 2016	*r*	0.86	0.85	0.86	0.89	0.88	0.85	–	–	–
			N	4,530	4,337	4,866	4,685	4,689	4,360	–	–	–
WY	PAWS Mathematics	Spring 2016	*r*	0.83	0.85	0.86	0.84	0.85	0.84	–	–	–
			N	2,744	2,544	2,602	2,402	2,496	2,367	–	–	–
Science												
TX	STAAR Science	Spring 2017	*r*	–	–	0.78	–	–	0.79	–	–	–
			N	–	–	13,454	–	–	4,220	–	–	–

*Dates reflect the most recent studies available in each state.
**Blank cells indicate that no data were available for that grade and test.

Table E.2. Concurrent Validity of MAP Growth Tests as Measured by Pearson Product-Moment Correlations between RIT Scores and ACT Aspire, PARCC, and SBAC Scores

States	State Test	Admin.		Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8
Reading									
SC	ACT Aspire Reading	Spring 2015	***r***	0.76	0.78	0.75	0.75	0.74	0.75
			N	2,804	2,780	2,645	2,577	2,698	2,801
CO, RI, NM, NJ, MD, Il, DC	PARCC ELA	Spring 2016	***r***	0.80	0.79	0.79	0.78	0.77	0.76
			N	47,463	45,045	44,093	46,123	44,179	40,387
CA, WA, ME	SBAC ELA	Spring 2015	***r***	0.81	0.82	0.83	0.81	0.80	0.80
			N	7,000	6,581	7,050	6,672	6,308	5,919
Mathematics									
SC	ACT Aspire Mathematics	Spring 2015	***r***	0.76	0.77	0.75	0.77	0.77	0.84
			N	2,781	2,704	2,658	2,685	2,658	2,783
CO, RI, NM, NJ, MD, IL, DC	PARCC Mathematics	Spring 2016	***r***	0.84	0.85	0.85	0.85	0.84	0.82
			N	47,534	45,129	44,138	46,184	43,899	37,699
CA, WA, ME	SBAC Mathematics	Spring 2015	***r***	0.86	0.88	0.88	0.89	0.87	0.85
			N	6,993	6,665	7,116	7,042	6,141	5,625

Appendix F: Classification Accuracy by State

Table F.1. Criterion-Related Validity of MAP Growth Tests as Measured by Classification Accuracy Between MAP Growth Predictions and Observed Proficiency Status on State Summative Assessments

				ELA/Reading**				Mathematics**				Science**			
State	State Test	Admin.*	Grade	N	Class. Accuracy	FP	FN	N	Class. Accuracy	FP	FN	N	Class. Accuracy	FP	FN
AK	AMP	Spring 2015	3	1,748	0.87	0.06	0.07	1,744	0.86	0.07	0.07	–	–	–	–
			4	1,639	0.87	0.07	0.06	1,644	0.87	0.07	0.06	–	–	–	–
			5	1,764	0.86	0.08	0.06	1,770	0.89	0.06	0.05	–	–	–	–
			6	1,599	0.86	0.07	0.07	1,603	0.90	0.05	0.05	–	–	–	–
			7	1,633	0.85	0.08	0.07	1,643	0.89	0.05	0.06	–	–	–	–
			8	1,673	0.87	0.07	0.06	1,677	0.90	0.04	0.06	–	–	–	–
			9	980	0.88	0.06	0.06	1,055	0.89	0.06	0.05	–	–	–	–
			10	780	0.88	0.05	0.07	789	0.91	0.03	0.06	–	–	–	–
AR	ACTAAP	Spring 2009*	3	1,868	0.81	0.09	0.10	1,787	0.89	0.05	0.06	–	–	–	–
			4	1,743	0.82	0.08	0.10	1,712	0.87	0.06	0.07	–	–	–	–
			5	1,307	0.83	0.08	0.10	1,286	0.87	0.06	0.07	–	–	–	–
			6	1,056	0.84	0.07	0.09	1,054	0.86	0.07	0.07	–	–	–	–
			7	1,164	0.82	0.09	0.09	1,155	0.86	0.07	0.07	–	–	–	–
			8	1,144	0.83	0.08	0.10	1,135	0.86	0.06	0.07	–	–	–	–
AZ	AzMERIT	Spring 2015	3	1,779	0.85	0.07	0.08	1,776	0.85	0.07	0.08	–	–	–	–
			4	1,572	0.81	0.10	0.09	1,573	0.87	0.05	0.08	–	–	–	–
			5	1,651	0.86	0.06	0.08	1,652	0.88	0.05	0.07	–	–	–	–
			6	1,501	0.87	0.06	0.07	1,503	0.90	0.05	0.05	–	–	–	–
			7	1,493	0.82	0.09	0.09	1,559	0.89	0.05	0.06	–	–	–	–
			8	1,602	0.85	0.07	0.08	1,855	0.88	0.06	0.06	–	–	–	–

State	State Test	Admin.*	Grade	ELA/Reading**				Mathematics**				Science**			
				N	Class. Accuracy	FP	FN	N	Class. Accuracy	FP	FN	N	Class. Accuracy	FP	FN
FL	FSA	Spring 2016	3	5,824	0.83	0.09	0.08	5,806	0.83	0.08	0.09	–	–	–	–
			4	5,479	0.83	0.09	0.08	5,516	0.86	0.08	0.06	–	–	–	–
			5	5,293	0.82	0.10	0.08	5,267	0.86	0.07	0.07	–	–	–	–
			6	4,784	0.82	0.10	0.08	4,677	0.84	0.09	0.07	–	–	–	–
			7	3,905	0.81	0.11	0.08	3,491	0.82	0.09	0.09	–	–	–	–
			8	3,710	0.80	0.11	0.09	2,352	0.79	0.13	0.09	–	–	–	–
GA	Milestones	Spring 2015	3	1,615	0.84	0.07	0.09	1,620	0.84	0.09	0.07	–	–	–	–
			4	1,521	0.84	0.08	0.08	1,546	0.87	0.07	0.06	–	–	–	–
			5	1,514	0.84	0.08	0.08	1,553	0.87	0.07	0.06	–	–	–	–
			6	1,497	0.85	0.08	0.07	1,470	0.87	0.07	0.06	–	–	–	–
			7	1,505	0.84	0.09	0.07	1,506	0.87	0.07	0.06	–	–	–	–
			8	1,407	0.85	0.06	0.09	1,442	0.88	0.06	0.06	–	–	–	–
IA	ITBS	Fall 2007–2009*	3	1,104	0.87	0.06	0.07	940	0.89	0.05	0.06	–	–	–	–
			4	1,017	0.88	0.06	0.06	876	0.91	0.05	0.05	–	–	–	–
			5	1,074	0.88	0.06	0.06	1,075	0.91	0.04	0.05	–	–	–	–
			6	861	0.82	0.09	0.09	860	0.89	0.05	0.05	–	–	–	–
			7	993	0.85	0.08	0.08	991	0.90	0.04	0.06	–	–	–	–
			8	1,019	0.87	0.06	0.07	968	0.87	0.06	0.07	–	–	–	–
			9	1,651	0.87	0.06	0.07	1,651	0.88	0.05	0.07	–	–	–	–
			10	1,196	0.87	0.06	0.07	1,201	0.87	0.06	0.07	–	–	–	–
			11	968	0.87	0.06	0.07	975	0.87	0.05	0.07	–	–	–	–
IN	ISTEP+	Spring 2016	3	8,969	0.87	0.08	0.05	9,010	0.89	0.08	0.03	–	–	–	–
			4	8,684	0.87	0.07	0.06	8,721	0.87	0.07	0.06	–	–	–	–
			5	15,069	0.87	0.07	0.06	15,135	0.89	0.06	0.05	–	–	–	–
			6	8,797	0.85	0.08	0.07	8,877	0.88	0.06	0.06	–	–	–	–
			7	7,877	0.86	0.08	0.06	7,870	0.87	0.07	0.06	–	–	–	–
			8	7,251	0.82	0.10	0.08	7,263	0.86	0.07	0.07	–	–	–	–

State	State Test	Admin.*	Grade	ELA/Reading** N	ELA/Reading** Class. Accuracy	ELA/Reading** FP	ELA/Reading** FN	Mathematics** N	Mathematics** Class. Accuracy	Mathematics** FP	Mathematics** FN	Science** N	Science** Class. Accuracy	Science** FP	Science** FN
KS	KAP	Spring 2015	3	3,339	0.85	0.08	0.07	3,359	0.86	0.08	0.06	–	–	–	–
			4	3,099	0.87	0.07	0.06	3,135	0.86	0.08	0.06	–	–	–	–
			5	3,156	0.83	0.08	0.09	3,203	0.88	0.07	0.05	–	–	–	–
			6	2,979	0.84	0.07	0.09	3,014	0.87	0.06	0.07	–	–	–	–
			7	2,415	0.82	0.07	0.11	2,547	0.90	0.05	0.05	–	–	–	–
			8	2,413	0.86	0.07	0.07	2,491	0.93	0.03	0.04	–	–	–	–
			10	815	0.86	0.10	0.04	867	0.92	0.03	0.05	–	–	–	–
KY	K-PREP	Spring 2015	3	9,619	0.82	0.09	0.09	9,635	0.82	0.08	0.10	–	–	–	–
			4	10,165	0.80	0.11	0.09	10,164	0.83	0.10	0.07	–	–	–	–
			5	10,013	0.80	0.10	0.10	10,011	0.84	0.08	0.08	–	–	–	–
			6	10,440	0.81	0.10	0.09	10,449	0.84	0.08	0.08	–	–	–	–
			7	10,283	0.81	0.09	0.10	10,312	0.85	0.07	0.08	–	–	–	–
			8	10,038	0.80	0.10	0.10	10,004	0.84	0.08	0.08	–	–	–	–
LA	LEAP	Spring 2016	3	2,756	0.83	0.09	0.08	2,743	0.85	0.07	0.08	–	–	–	–
			4	2,756	0.82	0.10	0.08	2,772	0.87	0.08	0.05	–	–	–	–
			5	2,605	0.82	0.09	0.09	2,635	0.87	0.06	0.07	–	–	–	–
			6	2,632	0.79	0.11	0.10	2,656	0.88	0.06	0.06	–	–	–	–
			7	2,461	0.80	0.11	0.09	2,468	0.90	0.05	0.05	–	–	–	–
			8	2,501	0.80	0.11	0.09	2,444	0.86	0.07	0.07	–	–	–	–
MA	MCAS	Spring 2018	3	2,389	0.81	0.16	0.25	2,649	0.84	0.16	0.17	–	–	–	–
			4	2,650	0.81	0.16	0.23	2,858	0.85	0.15	0.16	–	–	–	–
			5	2,516	0.82	0.16	0.20	2,835	0.86	0.14	0.13	–	–	–	–
			6	2,045	0.83	0.12	0.26	2,436	0.87	0.13	0.13	–	–	–	–
			7	1,414	0.83	0.13	0.24	1,381	0.90	0.11	0.10	–	–	–	–
			8	1,218	0.81	0.14	0.30	1,172	0.88	0.10	0.20	–	–	–	–

State	State Test	Admin.*	Grade	ELA/Reading**				Mathematics**				Science**			
				N	Class. Accuracy	FP	FN	N	Class. Accuracy	FP	FN	N	Class. Accuracy	FP	FN
MI	M-STEP	Spring 2016	3	4,824	0.84	0.08	0.08	4,794	0.86	0.07	0.07	–	–	–	–
			4	4,599	0.84	0.08	0.08	4,579	0.86	0.07	0.07	–	–	–	–
			5	4,613	0.85	0.08	0.07	4,623	0.89	0.05	0.06	–	–	–	–
			6	4,732	0.86	0.07	0.07	4,742	0.90	0.05	0.05	–	–	–	–
			7	4,571	0.84	0.08	0.08	4,608	0.91	0.04	0.05	–	–	–	–
			8	4,530	0.84	0.08	0.08	4,606	0.90	0.04	0.06	–	–	–	–
MN	MCA-III	Spring 2015	3	6,706	0.86	0.08	0.06	6,737	0.90	0.06	0.04	–	–	–	–
			4	6,460	0.85	0.07	0.08	6,458	0.90	0.06	0.04	–	–	–	–
			5	6,513	0.86	0.06	0.08	6,566	0.88	0.06	0.06	–	–	–	–
			6	5,964	0.86	0.08	0.06	5,876	0.89	0.05	0.06	–	–	–	–
			7	5,886	0.84	0.08	0.08	5,535	0.88	0.06	0.06	–	–	–	–
			8	5,315	0.85	0.07	0.08	4,493	0.86	0.07	0.07	–	–	–	–
MS	Mississippi Assessment Program	Spring 2016	3	2,567	0.83	0.09	0.08	2,581	0.85	0.08	0.07	–	–	–	–
			4	2,277	0.81	0.09	0.10	2,274	0.86	0.07	0.07	–	–	–	–
			5	2,285	0.86	0.07	0.07	2,282	0.86	0.07	0.07	–	–	–	–
			6	2,323	0.86	0.07	0.07	2,313	0.86	0.07	0.07	–	–	–	–
			7	2,088	0.84	0.09	0.07	2,092	0.83	0.08	0.09	–	–	–	–
			8	2,032	0.84	0.09	0.07	1,960	0.85	0.09	0.06	–	–	–	–
NC	EOG	Spring 2013	3	6,503	0.83	0.08	0.09	6,527	0.83	0.07	0.10	–	–	–	–
			4	7,115	0.82	0.09	0.09	7,033	0.86	0.07	0.07	–	–	–	–
			5	6,898	0.81	0.09	0.10	6,823	0.85	0.07	0.08	–	–	–	–
			6	4,623	0.82	0.09	0.09	4,588	0.85	0.06	0.09	–	–	–	–
			7	4,495	0.81	0.09	0.10	4,529	0.86	0.07	0.07	–	–	–	–
			8	4,395	0.82	0.09	0.09	4,474	0.86	0.06	0.08	–	–	–	–

State	State Test	Admin.*	Grade	ELA/Reading**				Mathematics**				Science**			
				N	Class. Accuracy	FP	FN	N	Class. Accuracy	FP	FN	N	Class. Accuracy	FP	FN
NE	NeSA	Spring 2015	3	1,675	0.89	0.06	0.05	1,674	0.88	0.07	0.05	–	–	–	–
			4	1,635	0.91	0.05	0.04	1,635	0.90	0.06	0.04	–	–	–	–
			5	1,698	0.91	0.04	0.05	1,700	0.90	0.06	0.04	–	–	–	–
			6	1,617	0.89	0.05	0.06	1,618	0.90	0.06	0.04	–	–	–	–
			7	1,815	0.91	0.04	0.05	1,821	0.88	0.06	0.06	–	–	–	–
			8	1,333	0.86	0.07	0.07	1,365	0.89	0.06	0.05	–	–	–	–
NY	NYSTP	Spring 2013	3	1,027	0.82	0.12	0.06	1,025	0.81	0.09	0.10	–	–	–	–
			4	1,070	0.83	0.08	0.09	1,074	0.80	0.10	0.10	–	–	–	–
			5	1,047	0.81	0.09	0.10	1,048	0.80	0.11	0.09	–	–	–	–
			6	1,026	0.81	0.10	0.09	1,018	0.77	0.12	0.11	–	–	–	–
			7	1,028	0.82	0.10	0.08	1,029	0.80	0.11	0.09	–	–	–	–
			8	958	0.79	0.08	0.13	956	0.82	0.08	0.10	–	–	–	–
OH	OST	Spring 2016	3	5,421	0.79	0.11	0.10	5,189	0.83	0.08	0.09	–	–	–	–
			4	4,991	0.81	0.10	0.09	5,035	0.82	0.09	0.09	–	–	–	–
			5	4,642	0.82	0.10	0.08	4,388	0.82	0.09	0.09	–	–	–	–
			6	4,636	0.83	0.11	0.06	4,418	0.85	0.08	0.07	–	–	–	–
			7	4,450	0.84	0.09	0.07	4,376	0.87	0.06	0.07	–	–	–	–
			8	4,573	0.83	0.09	0.08	3,804	0.80	0.10	0.10	–	–	–	–
PA	PSSA	Spring 2015	3	1,207	0.91	0.05	0.04	1,210	0.87	0.09	0.04	–	–	–	–
			4	1,262	0.88	0.06	0.06	1,265	0.87	0.08	0.05	–	–	–	–
			5	1,262	0.90	0.04	0.06	1,266	0.88	0.06	0.06	–	–	–	–
			6	846	0.87	0.06	0.07	850	0.86	0.08	0.06	–	–	–	–
			7	854	0.86	0.08	0.06	854	0.85	0.09	0.06	–	–	–	–
			8	821	0.86	0.07	0.07	830	0.84	0.06	0.10	–	–	–	–

State	State Test	Admin.*	Grade	ELA/Reading** N	ELA/Reading** Class. Accuracy	ELA/Reading** FP	ELA/Reading** FN	Mathematics** N	Mathematics** Class. Accuracy	Mathematics** FP	Mathematics** FN	Science** N	Science** Class. Accuracy	Science** FP	Science** FN
SC***	SC READY	Spring 2017	3	15,018	0.85	n/a	n/a	n/a	n/a	n/a	n/a	–	–	–	–
			4	16,203	0.85	n/a	n/a	n/a	n/a	n/a	n/a	–	–	–	–
			5	15,783	0.85	n/a	n/a	n/a	n/a	n/a	n/a	–	–	–	–
			6	15,333	0.85	n/a	n/a	n/a	n/a	n/a	n/a	–	–	–	–
			7	14,928	0.85	n/a	n/a	n/a	n/a	n/a	n/a	–	–	–	–
			8	14,245	0.84	n/a	n/a	n/a	n/a	n/a	n/a	–	–	–	–
TX	STAAR	Spring 2017	3	21,354	0.83	0.08	0.09	21,045	0.83	0.09	0.08	–	–	–	–
			4	22,182	0.84	0.07	0.09	21,951	0.86	0.07	0.07	–	–	–	–
			5	21,296	0.82	0.07	0.11	21,075	0.86	0.07	0.07	13,454	0.82	0.07	0.11
			6	20,301	0.85	0.07	0.08	19,463	0.88	0.07	0.05	–	–	–	–
			7	17,464	0.84	0.08	0.08	17,149	0.88	0.06	0.06	–	–	–	–
			8	9,725	0.83	0.07	0.10	11,297	0.83	0.08	0.09	4,220	0.86	0.06	0.08
VA	SOL	Spring 2014	3	1,573	0.84	0.08	0.08	1,550	0.83	0.09	0.08	–	–	–	–
			4	1,573	0.83	0.11	0.06	1,550	0.86	0.07	0.07	–	–	–	–
			5	1,556	0.83	0.08	0.09	1,522	0.84	0.08	0.08	–	–	–	–
			6	1,249	0.82	0.10	0.08	1,229	0.86	0.07	0.07	–	–	–	–
			7	1,179	0.84	0.08	0.08	1,052	0.82	0.09	0.09	–	–	–	–
			8	258	0.85	0.10	0.05	722	0.81	0.09	0.10	–	–	–	–
WI	Forward	Spring 2016	3	4,282	0.82	0.09	0.09	4,530	0.86	0.08	0.06	–	–	–	–
			4	4,127	0.82	0.10	0.08	4,337	0.87	0.08	0.05	–	–	–	–
			5	4,616	0.81	0.10	0.09	4,866	0.86	0.08	0.06	–	–	–	–
			6	4,686	0.82	0.10	0.08	4,685	0.87	0.06	0.07	–	–	–	–
			7	4,697	0.83	0.08	0.09	4,689	0.88	0.08	0.04	–	–	–	–
			8	4,377	0.82	0.09	0.09	4,360	0.87	0.08	0.05	–	–	–	–

State	State Test	Admin.*	Grade	ELA/Reading** N	Class. Accuracy	FP	FN	Mathematics** N	Class. Accuracy	FP	FN	Science** N	Class. Accuracy	FP	FN
WY	PAWS	Spring 2016	3	2,740	0.83	0.09	0.08	2,744	0.84	0.08	0.08	–	–	–	–
			4	2,542	0.83	0.08	0.09	2,544	0.87	0.08	0.07	–	–	–	–
			5	2,597	0.85	0.08	0.07	2,602	0.87	0.07	0.06	–	–	–	–
			6	2,406	0.84	0.09	0.07	2,402	0.84	0.09	0.07	–	–	–	–
			7	2,497	0.84	0.08	0.08	2,496	0.86	0.07	0.07	–	–	–	–
			8	2,362	0.80	0.09	0.11	2,367	0.85	0.08	0.07	–	–	–	–

*Dates reflect the most recent studies available in each state.
**N = number of students. FP = The proportion of below-proficient students who were incorrectly predicted by MAP Growth to be proficient. FN = The proportion of proficient students who were incorrectly predicted by MAP Growth to be below proficiency. Class. Accuracy = The proportion of students in the study sample whose proficiency classification on the state test was correctly predicted by MAP Growth cut scores. Due to rounding, proportions may not sum to 1.
***n/a = not available. For more details, see “2018 Linking Study: Predicting Performance on SC READY from NWEA MAP Growth” available online at https://www.nwea.org/resource/type/linking-studies/.

Table F.2. Criterion-Related Validity of MAP Growth Tests as Measured by Classification Accuracy Between MAP Growth Predictions and Observed Proficiency Status on ASPIRE, PARCC, and SBAC Summative Assessments

States	State Test	Admin.*	Grade	ELA/Reading** N	Class. Accuracy	FP	FN	Mathematics** N	Class. Accuracy	FP	FN
SC***	ACT Aspire	Spring 2015	3	2,804	0.84	n/a	n/a	2,781	0.77	n/a	n/a
			4	2,780	0.84	n/a	n/a	2,704	0.79	n/a	n/a
			5	2,645	0.81	n/a	n/a	2,658	0.77	n/a	n/a
			6	2,577	0.82	n/a	n/a	2,685	0.71	n/a	n/a
			7	2,698	0.83	n/a	n/a	2,658	0.84	n/a	n/a
			8	2,801	0.80	n/a	n/a	2,783	0.86	n/a	n/a
CO, RI, NM, NJ, MD, IL, DC	PARCC	Spring 2016	3	47,463	0.84	0.09	0.07	47,534	0.85	0.07	0.07
			4	45,045	0.83	0.09	0.08	45,129	0.88	0.05	0.07
			5	44,093	0.84	0.08	0.09	44,138	0.87	0.06	0.07
			6	46,123	0.83	0.09	0.08	46,184	0.89	0.05	0.06
			7	44,179	0.82	0.08	0.10	43,899	0.89	0.06	0.06
			8	40,387	0.81	0.09	0.10	37,699	0.88	0.05	0.07

States	State Test	Admin.*	Grade	ELA/Reading**				Mathematics**			
				N	Class. Accuracy	FP	FN	N	Class. Accuracy	FP	FN
CA, WA, ME	SBAC	Spring 2015	3	7,000	0.84	0.09	0.07	6,993	0.85	0.08	0.07
			4	6,581	0.84	0.08	0.08	6,665	0.87	0.06	0.07
			5	7,050	0.84	0.08	0.08	7,116	0.88	0.06	0.06
			6	6,672	0.83	0.09	0.08	7,042	0.88	0.06	0.06
			7	6,308	0.83	0.08	0.09	6,141	0.89	0.06	0.05
			8	5,919	0.83	0.09	0.08	5,625	0.89	0.05	0.06

*Dates reflect the most recent studies available in each state.
**N = number of students. FP = The proportion of below-proficient students who were incorrectly predicted by MAP Growth to be proficient. FN = The proportion of proficient students who were incorrectly predicted by MAP Growth to be below proficiency. Class. Accuracy = The proportion of students in the study sample whose proficiency classification on the state test was correctly predicted by MAP Growth cut scores. Due to rounding, proportions may not sum to 1.
***n/a = not available. For more details, see “Linking the ACT Aspire Assessments to NWEA MAP Growth Tests” available online at https://www.nwea.org/resource/type/linking-studies/.