i-Ready Inform and Growth Monitoring Assessments Technical Report

A Curriculum Associates Technical Report | March 2026

The *i-Ready Diagnostic* assessment is becoming

i-Ready Inform™

Starting in school year 2026–2027, the name of the *i-Ready Diagnostic* is changing to *i-Ready Inform* for Reading and for Mathematics.

This new name better reflects the power and the primary purpose of our adaptive assessment: to equip teachers, families, and students with the information they need to guide learning.

New Name, Same *i-Ready* Insights

While it may have a new name, the assessment will maintain the valid, reliable data, easy-to-use reports, and tailored support educators expect from *i-Ready*.

Although some screenshots and other content in this technical manual refers to *i-Ready*'s adaptive assessment as the *i-Ready Diagnostic*, the results featured in this document continue to hold for the newly renamed *i-Ready Inform* assessment.

A Shortened Version Available in Select States

Beginning with the 2026–2027 school year, a shortened version of *i-Ready*'s adaptive assessment will be available in select states. The validity evidence referenced in this document continues to hold for the shortened version of the assessment; however, in recognition of the fact that test reliability is a function of test length, the shortened version will have slightly lower—yet still exceedingly high—reliability estimates.

Reliability information for the shortened version of the assessment is available in the document *i-Ready Inform*: Modeled Reliability Estimates for the Shortened Assessment Experience.

Acknowledgements

The creation of this technical manual would not have been possible without the dedication and expertise of the following individuals: Nathan Minchen, Jeff Steedle, Kevin Cappaert, Laurie Davis, Dan Mix, Amanda Brice, Bill Spagnola, Logan Rome, Luciana Cançado, Margie McCaw, Yu Su, Natasha Williams, Kristin Morrison, Kristen Huff, Becky Bidwell-Hanson, Allison Johnson, Val Duval, Peter Swerdzewski, Ashleigh Crabtree, Jan Draime, Rhonda Collins, Sarah Salerno, Karla Egan (EdMetric), Annie Davidson (EdMetric)

Suggested Citation: Curriculum Associates, LLC. (2026). *i-Ready Inform & Growth Monitoring Assessments Technical Report*. North Billerica, MA: Author.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction and Foundations

1.1 Chapter Summary

In Chapter 1, we establish the purpose of this technical manual as providing an evidentiary basis for *i-Ready Inform*'s validity argument, which is organized around the assessment claims outlined in this chapter. The *i-Ready Growth Monitoring* assessment program is also discussed where relevant. The chapter starts with the conceptual foundations and the target student population of the *i-Ready Inform* assessment. Next, we discuss the structure of the assessment, including key features, subjects, and domains. Then we outline the validity framework of the program—grounded in our Theory of Action—which includes the intended purpose of the assessment, the interpretations and use of the scores, and the process we use for collecting and organizing evidence in support of our validity argument. We conclude the chapter with an orientation to the remainder of the manual.

1.2 Introduction

1.2.1 Foundations

This manual describes the technical characteristics of *i-Ready Inform* and the *i-Ready Growth Monitoring* assessments, which are components of a suite [1] of assessment and curriculum offerings that are designed to work together to support Curriculum Associates' mission to make classrooms better places for teachers and students. To accomplish this mission, we have designed our products to (1) enhance student knowledge of their performance and progress throughout the year, (2) equip teachers with data to support their decision making in the classroom, and (3) empower students, teachers, and administrators by creating a common lexicon around data that can be used to discuss student learning and resource planning. Taken together, we believe that these steps will result in increased student, teacher, and administrator knowledge, strategy, and engagement, which will ultimately lead to improved student learning outcomes. To ensure the best possible data from the *i-Ready Inform* and Growth Monitoring assessments, Curriculum Associates employs best practices in our test development procedures and psychometric methods according to the available industry-wide standards, which serve as the foundation of this technical manual.

[1] Other components of the *i-Ready Assessment* suite are (a) *i-Ready Standards Mastery* for Grades 2–8 covering reading and mathematics in a digital, fixed form; (b) *i-Ready Evaluación Diagnóstica de lectura en español* for Grades K–6 in a digital, adaptive format; and (c) *i-Ready Literacy Tasks*: Benchmarking and Progress Monitoring in English and Spanish with one-on-one administration, including Passage Reading Fluency (Grades 1–6) and additional Literacy Tasks (Grades K–3).

1.2.2 *i-Ready Inform*

1.2.2.1 About the Assessment

i-Ready Inform is an adaptive assessment built to evaluate student proficiency of College- and Career-Readiness Standards (CCRS) for reading and mathematics in kindergarten through Grade 12 (K–12) using an online interface. It is typically administered up to three times per school year. A transadapted Spanish language version of the mathematics assessment is also available.

1.2.2.2 Reported Scores

Both overall and domain-level scores are reported on a vertical scale that ranges from 100 to 800 with increments that can be as small as one point. Overall and domain placement levels are also reported; grade-level placements consist of Early On Grade, Mid On Grade, and Late On Grade, commonly referred to simply as Early, Mid, and Late. Students can also receive below- and above-grade level placements where applicable. Students in Grades 2 and above cannot place below K—Emerging K is reserved exclusively for kindergarten and Grade 1—and all students cannot place more than three grade levels above their chronological grade (i.e., the grade level in which the student is rostered, as reported by their school). Skill-based statements about what a student can likely do and what would be helpful to work on next—referred to as *Can Dos* and *Next Steps*—are also provided on score reports, which are presented in Chapter 11.

1.2.3 *i-Ready Growth Monitoring*: Background and End-of-Year Projected Scores

1.2.3.1 About the Assessment

The *i-Ready Growth Monitoring* assessments are short computer-adaptive assessments of reading and mathematics for students in Grades K–8. The Growth Monitoring assessments were designed with the intention of taking approximately 20 minutes or less per administration. Administration times vary, but average times for students in Grades 4 and under tend to be less than 15 minutes, and average times for students in Grades 5–8 tend to range between 15 and 20 minutes, although some students may take longer to complete an assessment in some circumstances. The optimal usage is approximately once per month in between *i-Ready Inform* administrations. While *i-Ready Inform* is the primary focus of this technical manual, Growth Monitoring assessments are discussed when relevant.

1.2.3.2 Projected End-of-Year Score

The purpose of Growth Monitoring is to monitor progress between *i-Ready Inform* administrations to determine whether students are making sufficient growth by using an end-of-year (EOY) projected score. This projection is compared to the Typical Growth and Stretch Growth® measures (see Chapter 9) assigned during the fall administration (or most recent if no fall test was administered) to determine their relative status. For students placing below grade level, Growth Monitoring can also be used as a tool for evaluating response to intervention programs.

1.2.4 Target Student Population

1.2.4.1 English Language Versions

i-Ready Inform is intended to be used by students for whom a multiple-choice assessment is deemed to be appropriate in Grades K–12. It is typically administered up to three times per year (i.e., fall, winter, and spring), and students may be routed into Personalized Instruction lessons based on their performance. The lesson sequence is designed to meet students at their current proficiency level[2] in each domain, prioritizing those domains with the lowest placement levels. Students who would benefit from this process are the target student population. Although Growth Monitoring can be used by all students in Grades K–8 who also take *i-Ready Inform*, it is specifically for students who have a need to show how they are tracking toward their growth goals, such as students receiving Tier 2 instruction through a multi-tiered system of supports (MTSS) or a response to intervention (RTI) model.

The degree to which students with significant cognitive disabilities (i.e., those students eligible for their state's alternative assessment of alternate achievement standards [AA-AAS]) may be adequately assessed by, or benefit from, *i-Ready Inform* depends greatly on an individual student's disability status and proficiencies, as well as the target use of *i-Ready Inform* for the student. For example, educators may find it useful to administer *i-Ready Inform* to some students with significant cognitive disabilities for the purpose of understanding these students' foundational skills. *i-Ready Inform* does not currently have an AA-AAS version built to alternate achievement standards, and it has not been evaluated for use with students with significant cognitive disabilities.

1.2.4.2 Spanish Language Versions

The goal of Spanish mathematics assessments is to match the content and rigor of their English counterparts, thereby providing students assessed in Spanish with the same experience as students assessed in English and providing teachers with the same reports they currently receive from the English versions. English language items are transadapted into Spanish for this assessment. The Spanish mathematics assessments are designed to be used as interim/benchmark assessments administered three times per year to help drive mathematics instruction in Spanish dual-language, Spanish immersion, and English Learner programs with the goal of measuring mathematics proficiency using a Spanish-language form.

Note that a separate assessment called the *i-Ready Evaluación Diagnóstica de lectura en español* is designed to assess Spanish-speaking students in reading. This assessment was developed from a set of original content developed and written by native Spanish-speaking experts in reading education and was not transadapted from an existing English assessment. As such, the assessment has a distinct item bank and scale. The technical characteristics of this assessment are discussed in a separate document.

1.2.4.3 Student Counts and Descriptive Statistics

Table 1.1 shows the number of *i-Ready Inform* administered by grade, subject, and language during the 2021–2022 school year for Mathematics (English) and Reading (English), and during the 2023–2024 school year for

[2]Note that *proficiency level*—sometimes referred to as *ability level* by other testing programs—refers to the level of proficiency demonstrated by a student either across all domains or in a single domain within a given grade and subject. Proficiency level should be distinguished from *placement level*, which refers to the criterion-referenced level in which the overall or domain score is situated.

the Mathematics (Spanish). Some data filtering rules were applied to best determine the number of valid tests for the purposes of this manual. In particular, the reported counts allow for at most one test per student, per testing window. Only tests that were administered fully in school are included. Table 1.2, Table 1.3, and Table 1.4 show the descriptive statistics for these data by grade. Table 1.5 to Table 1.13 show this same information, but disaggregated by *i-Ready*'s fall, winter, and spring testing windows.

Table 1.1. *i-Ready Inform* Counts by Subject, Language, and Grade after Filtering

Grade	Mathematics	Reading	Mathematics (Spanish)
K	2,281,818	1,878,648	79,291
1	2,600,489	2,175,773	79,865
2	2,774,663	2,328,438	75,411
3	2,870,693	2,486,736	48,678
4	2,826,144	2,452,357	37,293
5	2,854,279	2,469,573	32,963
6	2,352,756	2,023,140	22,866
7	2,081,161	1,839,402	19,585
8	1,961,463	1,808,160	18,753
9	283,093	313,476	5,137
10	177,740	214,492	3,165
11	118,950	146,513	1,300
12	68,945	98,845	518

Table 1.2. Descriptive Statistics for *i-Ready Inform* for Mathematics in English

Grade	Count	P5	P25	Median	P75	P95	Mean	SD
K	2,281,818	310	336	356	375	402	356	28.2
1	2,600,489	333	367	387	406	434	386	30.8
2	2,774,663	353	389	410	432	454	408	31.3
3	2,870,693	374	412	435	453	481	432	33.0
4	2,826,144	389	431	452	473	504	450	34.7
5	2,854,279	403	444	467	489	518	465	35.7
6	2,352,756	411	453	479	502	535	476	38.0
7	2,081,161	415	460	488	511	546	484	39.9
8	1,961,463	421	466	495	519	557	492	41.7
9	283,093	411	460	492	516	554	487	44.3
10	177,740	415	465	498	525	564	494	46.3
11	118,950	412	465	499	528	570	495	48.9
12	68,945	394	460	496	527	572	491	54.3

Table 1.3. Descriptive Statistics for *i-Ready Inform* for Mathematics in Spanish

Grade	Count	P5	P25	Median	P75	P95	Mean	SD
K	79,291	305	326	344	362	387	345	25.5
1	79,865	322	351	371	391	420	371	29.5
2	75,411	344	374	395	416	443	394	30.6
3	48,678	358	392	416	437	463	414	32.6
4	37,293	370	405	431	452	481	428	34.2
5	32,963	380	417	439	461	493	438	34.3
6	22,866	382	419	439	461	495	439	33.7
7	19,585	384	423	443	464	498	443	34.0
8	18,753	392	429	450	472	504	450	34.1
9	5,137	394	431	452	472	507	452	35.2
10	3,165	396	436	458	482	517	458	37.7
11	1,300	399	441	461	484	521	462	38.3
12	518	406	443	464	491	539	468	43.1

Table 1.4. Descriptive Statistics for *i-Ready Inform* for Reading

Grade	Count	P5	P25	Median	P75	P95	Mean	SD
K	1,878,648	307	341	367	400	440	371	41.7
1	2,175,773	340	385	416	455	514	421	52.1
2	2,328,438	368	421	471	512	562	467	59.9
3	2,486,736	396	463	512	547	592	503	61.2
4	2,452,357	416	496	540	576	618	532	61.4
5	2,469,573	439	520	562	597	637	554	60.8
6	2,023,140	453	534	577	611	652	568	62.0
7	1,839,402	461	547	591	626	665	581	63.6
8	1,808,160	474	561	606	637	676	594	63.7
9	313,476	459	556	604	640	688	592	70.7
10	214,492	471	570	616	652	699	605	70.1
11	146,513	466	572	620	658	704	608	73.4
12	98,845	454	573	624	664	708	609	79.1

Table 1.5. Descriptive Statistics for *i-Ready Inform* for Mathematics in English: Fall

Grade	Count	P5	P25	Median	P75	P95	Mean	SD
K	728,312	304	322	338	354	378	339	23.1
1	873,506	323	352	372	389	414	370	27.4
2	932,558	345	378	395	413	438	394	28.5
3	977,274	367	398	421	439	461	418	29.7

4	965,559	382	420	441	460	483	438	31.7
5	983,383	396	436	458	477	504	455	33.1
6	818,976	407	447	470	494	518	468	35.2
7	736,716	413	455	482	503	533	478	37.2
8	709,825	419	463	491	513	548	487	39.3
9	115,020	412	458	489	512	548	484	41.7
10	71,634	416	464	496	520	556	491	43.5
11	50,000	415	467	498	524	562	494	45.4
12	31,633	401	464	498	526	568	493	50.9

Table 1.6. Descriptive Statistics for *i-Ready Inform* for Mathematics in English: Winter

Grade	Count	P5	P25	Median	P75	P95	Mean	SD
K	764,592	314	340	358	374	396	357	25.3
1	847,187	338	371	389	405	433	388	28.6
2	906,862	358	393	412	431	451	410	29.2
3	946,757	378	416	437	453	474	433	30.1
4	932,059	393	433	454	472	498	451	32.4
5	943,359	406	446	468	489	515	466	34.0
6	773,759	413	454	480	502	532	477	37.0
7	687,247	416	461	489	511	545	485	39.3
8	650,767	422	469	496	521	556	494	41.3
9	86,978	410	460	493	518	556	488	45.4
10	55,484	415	466	499	528	569	496	47.4
11	37,786	412	467	501	531	575	498	50.0
12	22,094	393	460	497	529	577	492	56.0

Table 1.7. Descriptive Statistics for *i-Ready Inform* for Mathematics in English: Spring

Grade	Count	P5	P25	Median	P75	P95	Mean	SD
K	788,914	324	354	371	387	414	370	26.7
1	879,796	348	381	401	421	442	399	29.3
2	935,243	366	403	425	442	462	421	30.1
3	946,662	383	427	448	466	494	445	33.2
4	928,526	398	442	466	484	517	462	35.7
5	927,537	411	452	477	500	531	474	37.2
6	760,021	415	459	488	511	547	484	40.3
7	657,198	418	463	494	517	556	490	42.3
8	600,871	422	470	499	527	566	497	44.2
9	81,095	411	461	494	521	561	490	46.6

10	50,622	414	466	500	529	570	496	48.7
11	31,164	406	461	498	530	576	494	52.5
12	15,218	378	451	492	524	574	486	58.1

Table 1.8. Descriptive Statistics for *i-Ready Inform* for Mathematics in Spanish: Fall

Grade	Count	P5	P25	Median	P75	P95	Mean	SD
K	21,063	299	314	326	340	361	327	19.2
1	22,867	313	339	354	371	392	354	23.8
2	21,357	334	363	380	397	423	380	27.0
3	13,172	352	382	402	423	446	401	29.3
4	9,837	365	397	423	441	467	420	32.1
5	8,427	376	410	433	454	483	432	32.8
6	5,548	381	415	434	454	484	434	31.0
7	4,530	383	419	440	460	489	438	32.0
8	4,529	392	427	447	466	499	447	31.8
9	879	397	428	449	467	496	447	31.3
10	779	393	433	454	476	507	454	33.4
11	292	392	436	456	478	507	455	34.4
12	155	404	444	462	485	528	464	37.6

Table 1.9. Descriptive Statistics for *i-Ready Inform* for Mathematics in Spanish: Winter

Grade	Count	P5	P25	Median	P75	P95	Mean	SD
K	28,999	308	328	344	358	379	344	22.1
1	27,654	324	354	372	389	414	371	26.7
2	25,927	347	376	395	414	439	394	28.4
3	17,275	360	393	417	436	458	414	30.7
4	13,200	372	405	430	451	476	427	32.8
5	11,977	381	418	439	461	489	438	33.2
6	8,317	382	419	439	461	493	438	33.0
7	7,113	385	423	443	464	496	443	33.1
8	6,831	393	430	450	471	502	450	33.3
9	1,805	394	432	451	469	502	450	33.4
10	1,160	396	436	457	478	510	456	35.3
11	502	403	440	461	483	520	462	36.9
12	195	403	442	463	491	534	467	44.1

Table 1.10. Descriptive Statistics for *i-Ready Inform* for Mathematics in Spanish: Spring

Grade	Count	P5	P25	Median	P75	P95	Mean	SD
K	29,229	316	342	358	373	399	358	25.1
1	29,344	336	366	386	404	431	385	29.1
2	28,127	352	385	406	428	450	405	30.5
3	18,231	364	401	427	447	471	423	33.6
4	14,256	374	412	436	460	489	435	35.5
5	12,559	382	420	443	466	498	442	35.7
6	9,001	384	421	442	465	499	442	35.4
7	7,942	384	424	445	468	502	445	35.7
8	7,393	392	430	453	475	510	452	36.0
9	2,453	394	433	453	476	516	454	37.5
10	1,226	396	438	462	491	530	464	41.7
11	506	397	442	465	491	524	465	41.4
12	168	407	443	471	497	566	473	46.3

Table 1.11. Descriptive Statistics for *i-Ready Inform* for Reading: Fall

Grade	Count	P5	P25	Median	P75	P95	Mean	SD
K	586,487	296	326	341	361	402	345	31.8
1	725,879	330	367	397	419	474	396	43.5
2	776,899	357	407	443	485	536	445	55.2
3	839,411	388	446	490	529	574	486	58.5
4	836,199	408	482	526	562	605	519	60.1
5	850,212	430	511	550	585	628	543	59.8
6	701,407	448	527	569	603	643	560	60.7
7	649,043	459	541	584	619	658	575	62.3
8	646,725	472	555	598	632	670	588	62.3
9	129,822	464	555	602	637	683	591	67.8
10	91,286	476	571	616	650	695	605	67.5
11	64,736	476	575	621	658	702	610	69.9
12	46,976	463	577	626	665	708	612	75.6

Table 1.12. Descriptive Statistics for *i-Ready Inform* for Reading: Winter

Grade	Count	P5	P25	Median	P75	P95	Mean	SD
K	634,396	314	347	369	397	427	371	36.6
1	710,382	345	392	417	452	509	422	48.4
2	763,564	372	425	474	513	560	470	58.2
3	822,135	399	468	515	549	591	505	59.8

4	811,419	418	500	542	577	617	534	60.2
5	821,394	443	523	565	598	637	556	59.8
6	666,049	454	536	578	612	652	569	61.8
7	607,476	462	548	591	626	665	581	63.6
8	604,473	475	563	607	638	677	595	63.4
9	94,018	456	555	604	640	688	592	71.6
10	65,150	469	570	616	653	700	605	70.9
11	46,609	465	573	620	659	705	609	73.9
12	32,078	455	573	625	664	709	610	78.9

Table 1.13. Descriptive Statistics for *i-Ready Inform* for Reading: Spring

Grade	Count	P5	P25	Median	P75	P95	Mean	SD
K	657,765	330	366	395	416	464	393	40.7
1	739,512	358	409	442	481	530	444	52.5
2	787,975	384	448	493	529	574	487	58.8
3	825,190	405	481	526	562	603	518	61.1
4	804,739	425	512	553	587	628	544	61.5
5	797,967	448	530	573	606	644	563	61.0
6	655,684	457	542	585	618	658	575	62.7
7	582,883	464	554	598	631	670	586	64.6
8	556,962	474	567	612	642	681	598	65.0
9	89,636	454	559	608	644	694	595	73.9
10	58,056	466	570	618	655	704	606	73.1
11	35,168	451	565	617	657	705	603	78.9
12	19,791	424	559	617	661	708	600	86.7

Table 1.14 shows the number of Growth Monitoring assessments administered by grade, subject, and language during the 2021–2022 school year. As with *i-Ready Inform*, some data filtering rules were applied to best determine the number of valid tests for the purposes of this manual. Only tests that were administered fully in school are included.

Table 1.14. Growth Monitoring *N*-Counts by Subject, Language, and Grade after Filtering

Grade	Mathematics	Reading	Mathematics (Spanish)
K	462,524	408,542	5,067
1	586,812	520,870	5,921
2	627,361	561,076	5,088
3	650,589	593,470	2,307
4	603,873	553,011	1,697

5	589,155	533,357	1,020
6	313,211	269,006	529
7	230,392	204,402	448
8	200,795	192,397	414

1.3 Structure of *i-Ready Inform*

i-Ready Inform measures student performance in reading and mathematics by assessing student proficiency in specific content domains. The *test flow* refers to the number of items and order of domains administered within each subject and grade and can be affected by individual student performance. Test flows are discussed in detail in Chapter 4. The Growth Monitoring assessments provide only an overall score; they do not provide domain-specific information.

1.3.1 Key Components

To best achieve its purpose, *i-Ready Inform* is designed with several features that support the validity and usability of the scores. These key features, outlined below, are fundamental to the design and development of *i-Ready Inform* described in subsequent chapters.

- **Computer-Adaptive Design:** *i-Ready Inform* is a computerized-adaptive test (CAT) in which the item selection algorithm is applied after the completion of each item or testlet (in the case of reading comprehension) and is designed to match students to items with difficulty levels near their proficiency levels to best estimate proficiency. By adapting to student responses while also assessing a broad range of skills, *i-Ready Inform* is designed to pinpoint overall and domain proficiency levels with a high degree of precision. The item selection algorithm has minimal constraints, allowing for a high degree of adaptivity, which is made possible by a robust item bank.

- **Item Bank:** The *i-Ready Inform* item bank is composed of items that contain high-quality content aligned to meaningful content claims and placement levels based on both content and cognitive complexity, is usable and accessible to the full target student population, and is monitored to minimize bias or unfairness. The item bank is continuously updated through annual field testing of new items and is designed to meet the needs of students across all proficiency levels. The Growth Monitoring assessments use a subset of the *i-Ready Inform* item bank and do not include field test items.

- **Vertical Scale:** The vertical scales for reading and mathematics range from 100 to 800, allowing for scores within a given subject area to be compared both within and across grades in aggregate; interpretations at the individual level should be guided by the criterion-referenced placement levels, which are described below. Note that although the ranges of the scales for reading and mathematics are identical, they are unique scales and have different properties and interpretations (see Chapter 1 for descriptive statistics, Chapter 5 for scale transformation constants, and Chapter 6 for information on measurement precision). This design not only facilitates the tracking of growth, but it is also flexible in that it allows students to place below or above grade level.

- **Multiple Administrations:** *i-Ready Inform* is administered in the fall (i.e., beginning of school year through November 15), winter (i.e., November 16 through March 1), and spring (i.e., March 2 through end of school year). Seasonal administration allows for proficiency estimates to be updated throughout the school year. Growth Monitoring assessments can be administered once per month in between *i-Ready Inform* administrations to produce updated EOY projected scores.
- **Scores and Score Reports:** The scores and score reports from *i-Ready Inform* are available upon completion of the assessment and are designed to provide educators with detailed information about what a student can likely do, suggested next steps for skill building, and recommend instructional resources. Students receive an overall scale score, which represents performance in the content area, as well as domain-level scores, which represent performance in the domains (e.g., Algebra and Algebraic Thinking). Each domain is represented at different placement levels, with the aim of accurately measuring student understanding of the entire range of content for each domain.
- **Criterion-Referenced Placement Levels:** Scale scores are categorized into placement levels at both the overall and domain level within each grade. Students may place into one of the three on-grade placement levels for each grade (i.e., Early, Mid, and Late), or above or below grade level, with some restrictions, as described earlier. The Mid On Grade placement level is considered the minimum level of achievement needed to have met expectations for CCRS, and it serves as a benchmark for creating some Stretch Growth targets, which are described in Chapter 9. The standard-setting procedures used to derive the thresholds for the placement levels are described in Chapter 7. Note that the same scale score does not necessarily map onto the same placement level in different grades; criterion-referenced interpretations must be made relative to the expectations in the designated grade level.
- **Individualized Learning:** The domain placement levels resulting from an *i-Ready Inform* administration are used to place students into appropriate Personalized Instruction lessons. These lessons provide a concrete pathway for each student to improve their skills between *i-Ready Inform* administrations. While the Personalized Instruction program is a separate product, one of the primary uses of *i-Ready Inform* scores is to determine appropriate Personalized Instruction lessons.

1.3.2 Overview of Reading

The test flow for *i-Ready Inform* for Reading includes up to six domains:

1. Phonics
2. Phonological Awareness
3. High-Frequency Words
4. Vocabulary
5. Comprehension: Literature
6. Comprehension: Informational Text

i-Ready Inform for Reading test flow varies by chronological grade level. All students, regardless of grade level, are presented with items from Vocabulary, Comprehension: Literature, and Comprehension: Informational Text domains. Students in lower chronological grade levels or performing at lower placement levels may also be assigned Phonics, Phonological Awareness, and High-Frequency Words items. The Growth Monitoring

assessments administer items from all domains except High-Frequency Words and Phonological Awareness. See Chapter 4 for a detailed discussion of test flows.

The Foundational Reading Extension (FRE) is also available to schools and districts that wish to require all students up through Grade 6 to be administered items from the eligible foundational reading domains, which is composed of high-frequency words, phonics, and phonological awareness. Educators can also complement *i-Ready Inform* for Reading scores by selecting from *i-Ready Literacy Tasks*, a suite of companion tasks that teachers can print and administer offline (with digital data entry and digital scoring [3] available) to assess students' foundational and fluency skills in English or Spanish. *i-Ready Literacy Tasks* (Grades K–6) allow for more targeted understanding of reading skills of students who may need further evaluation, providing educators with tools to observe, quantify, and record students' reading behaviors.

1.3.3 Overview of Mathematics

The test flow for *i-Ready Inform* for Mathematics includes four domains:

1. Algebra and Algebraic Thinking
2. Number and Operations
3. Geometry
4. Measurement and Data

While each domain is typically administered at all grade levels, the percentage of items representing each of the domains for Grades K–8 is different than for high school. See Chapter 4 for a detailed discussion of test flows.

1.4 Validity Framework

As stated in the Standards for Educational and Psychological Testing (AERA et al., 2014), "Validity refers to the degree to which evidence and theory support the interpretations of test scores for proposed uses of tests." When documenting validity evidence for assessments, we do not find a single score or set of scores allowing us to say that the interpretation of test scores for a particular use is valid; rather, we build a validity argument that will either lend support for or discredit the intended uses for the assessment (Kane, 2006, 2013). Standard 1.0 (AERA et al., 2014) states, "Clear articulation of each intended test score interpretation for a specified use should be set forth, and appropriate validity evidence in support of each intended interpretation should be provided."

Therefore, building a validity argument requires documentation of evidence throughout the design, development, administration, and reporting phases of the assessment cycle. The validity framework, including internal, external, and procedural evidence (Kane, 2006, 2013), begins with our theory of action from which our assessment claims are derived and around which our validity argument is constructed.

Assessment design is the process of defining how to build an assessment so that it ultimately reports scores that can be used for a particular purpose and interpreted validly. The design process starts with the theory of

[3]Access to digital task scoring/administration is not available for Spelling and Encoding task types, which are generally group administered.

action—outlined in the next section—which is a robust description of the desired outcomes and major steps required to achieve those outcomes. Assessment design occurs upon inception but can be revisited frequently as adjustments and updates are made to the assessment. *Assessment development* follows from the design and is the subsequent process of constructing the assessment so that it meets the goals and objectives of the theory of action and design.

The assessment design and development processes involve many steps, including planning for the live administration and reporting, defining the target population (discussed earlier in this chapter), considering the resources and time available for the administration (see Chapter 3), and determining the desired level of granularity for reporting purposes (see Chapters 4 and 5 for technical details on measurement, Chapter 7 for details on standard setting, and Chapter 11 for details on score reporting). Decisions are informed by the theory of action and design specifications, which together provide the framework for decision making. The theory of action is ultimately used to evaluate the relative success of the assessment against the intended outcomes, and to identify where to make necessary adjustments to the assessment program over time.

1.4.1 Theory of Action

A theory of action can be thought of as a flow chart wherein the initial components relate to the administration of the assessment and the final component(s) relates to the long-term impact. Intermediate steps connect the beginning and end through a series of if–then links that are supported by reason and/or research.

Our complete theory of action is presented in Figure 1.1, which shows a series of interconnected components that start with an *i-Ready Inform* administration and result in improved learning outcomes for students, thereby showing how our program is intended to work. Various stakeholders (e.g., students, teachers, school and district leaders) are intended to play a role in the successful progression from left to right, and certain steps may be best completed with the use of our other product offerings. Thus, we acknowledge that the *i-Ready Inform* and Growth Monitoring assessments are components of a larger, more comprehensive theory of action that interacts with our other curriculum and instruction products as well as professional learning.

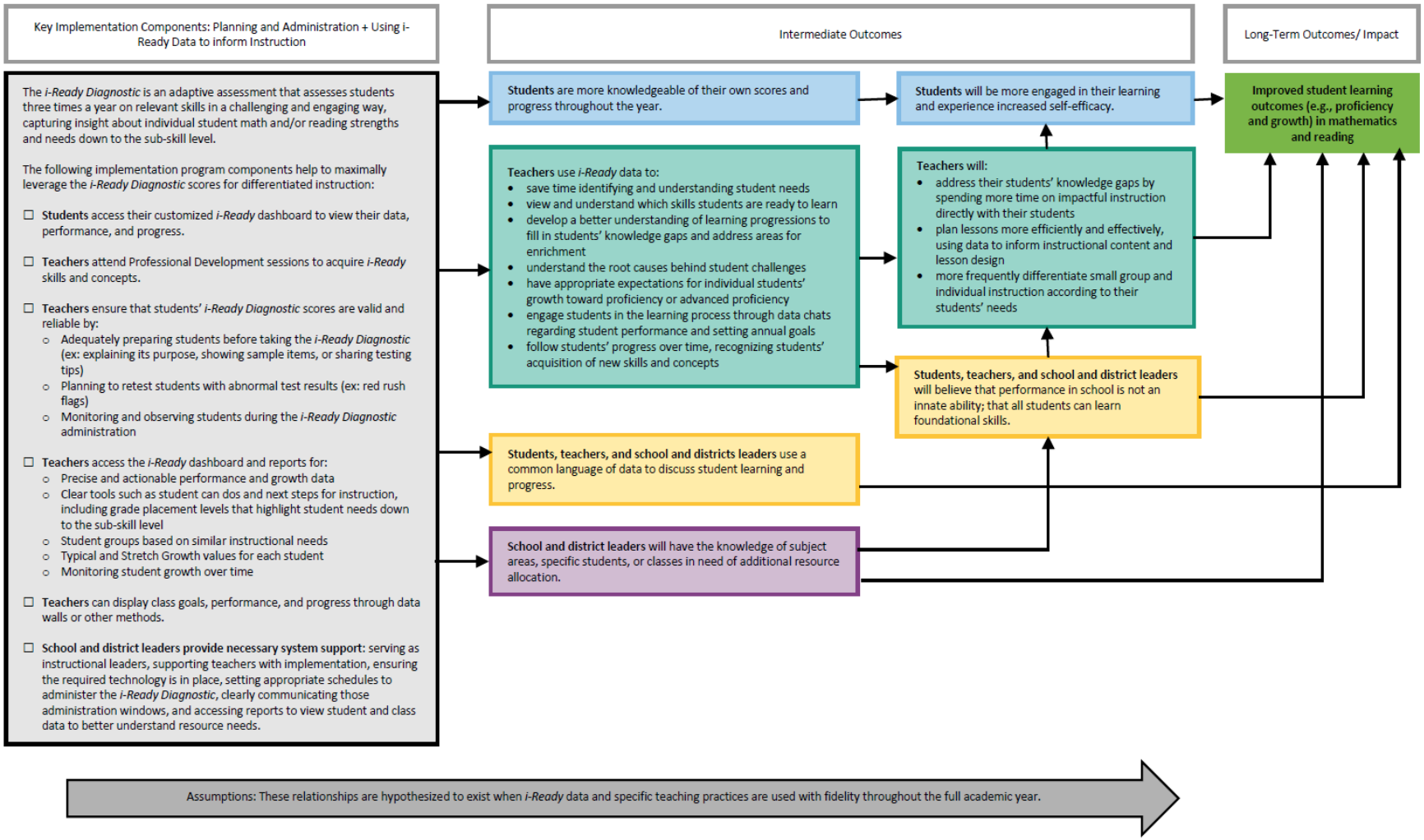


Figure 1.1. The i-Ready *Theory of Action*

1.4.2 The Purpose of Testing

The chief goal of *i-Ready Inform* is to improve classroom instruction and learning, making it an assessment *for* learning. Its purpose, derived from the theory of action, is to provide students, teachers, and administrators with precise and timely information that is both *descriptive* of what students know and can do and *prescriptive* by providing actionable information that can be used to improve student learning. This intended purpose guides assessment design, development, administrative guidance, and the interpretation of results.

The purpose of the Growth Monitoring assessments is to evaluate whether students are on track to achieve their growth measures (see Chapter 9) by the end of the year.

1.4.3 Assessment Claims

Assessment claims are broad inferences about the intended interpretation and use of test scores. In the case of *i-Ready Inform*, scores are intended to reflect meaningful information about student performance relative to expected knowledge and skills in Reading and Mathematics. In this section, we articulate the assessment claims at the heart of the validity framework.

These assessment claims are explicit statements that describe the intended meaning of scores, provided that *i-Ready Inform* is used as directed. Broadly, the *i-Ready Inform* assessment claims state (a) how scores are intended to reflect achievement with respect to the test constructs, (b) how assessment participants, including students and teachers, are intended to interact with the program, and (c) how assessment results can be interpreted and used validly for instructional purposes.

1.4.3.1 Claims Related to the Interpretation of Scores

As the flagship assessment in the *i-Ready* system, *i-Ready Inform*:

- *Measures* a series of subject-specific and grade-appropriate domains in Mathematics and Reading
- *Reports* the following estimates of student proficiency, which are reliable and can be compared across administrations and grades:
    - Overall scale scores
    - Overall placement levels
    - Domain scale scores
    - Domain placement levels
- *Contextualizes* student proficiency through criterion-referenced placement levels, applicable state standards, and norms
- *Provides* differentiated fall-to-spring Typical Growth and Stretch Growth measures and a Projected Proficiency status on state tests (select states only) based on these measures

1.4.3.2 Claims Related to the Intended Uses of the Scores

Validity evidence supports the following uses of scores and growth measures to:

- *Identify* and *monitor* students who are not meeting their growth measures, who need additional support and/or may benefit from intervention services, and who are ready for more challenging content, may no longer need intervention services, or may be eligible for gifted and talented services

- *Provide* students with actionable feedback about what they likely can do and what they should work on next
- *Group* students to provide differentiated instruction, for which teachers are provided appropriate instructional support tools
- *Route* students into Personalized Instruction modules[4], which provide additional support targeted to specific skills

To make inferences, we reason from these claims throughout the design, development, and evaluation process of *i-Ready Inform*. The intentionally designed layers of evidence, made explicit and transparent, result in increased clarity about the intended targets of measurement in *i-Ready Inform* and the constructs to which the inferences about students can be generalized. This results in greater coherence between what the assessment results say about what students know and can do and the role the assessment plays in a larger educational context. This clarity and coherence support the use of results for policymakers, educators, students, and parents.

1.4.4 The Validation Process

1.4.4.1 Validation

Validation is the process of developing and substantiating a validity argument. "Validation logically begins with an explicit statement of the proposed interpretation of the test scores, along with a rationale for the relevance of the interpretation to the proposed use" (AERA et al., 2014). This aspect of our validity argument is represented by our claims in the preceding section and the rationale detailed in the theory of action. Additionally, the testing standards state that "validation is the joint responsibility of the test developer and the test user" (AERA et al., 2014). To that end, Curriculum Associates is explicit about the intended interpretations and uses of the scores; it then becomes the responsibility of the users to adhere to administration and reporting protocols, as shown in the theory of action. The use of scores as intended, which extends beyond the formal engagement with *i-Ready Inform* itself, is imperative to realize the maximum impact on student learning.

Validation is not merely about the assessment itself but about supporting the inferences made from the assessment results. Defending an assessment's validity argument entails gathering ample evidence that supports the score interpretations. The validity principles are outlined in this chapter and the evidence will be presented and discussed throughout the document; we will conclude the manual with a narrative to summarize and organize the current state of the evidence.

1.4.4.2 Types of Validity Evidence

A theory of action frames the evidentiary and validation arguments for the assessment within the larger programmatic and educational context. Our theory of action details how *i-Ready Inform* and the resulting reports and resources will work in tandem to make classrooms better places for both teachers and students.

[4] *i-Ready Personalized Instruction* is an optional component within the *i-Ready* system which may be purchased in addition to *i-Ready Inform* and is intended to use results from *i-Ready Inform* as the key input to determining a student's personalized lesson path.

Because *i-Ready Inform* is an assessment *for* learning, the scores can be inputs into other features, reports, or products that support the theory of action.

Figure 1.2 shows how different sets of claims are related to the major components of the theory of action. Because this manual is focused mostly on the assessment component, this chapter deals mostly with the two rightmost boxes (middle and lower level). The *claims about assessment for given purpose and use* have been discussed in previous sections.

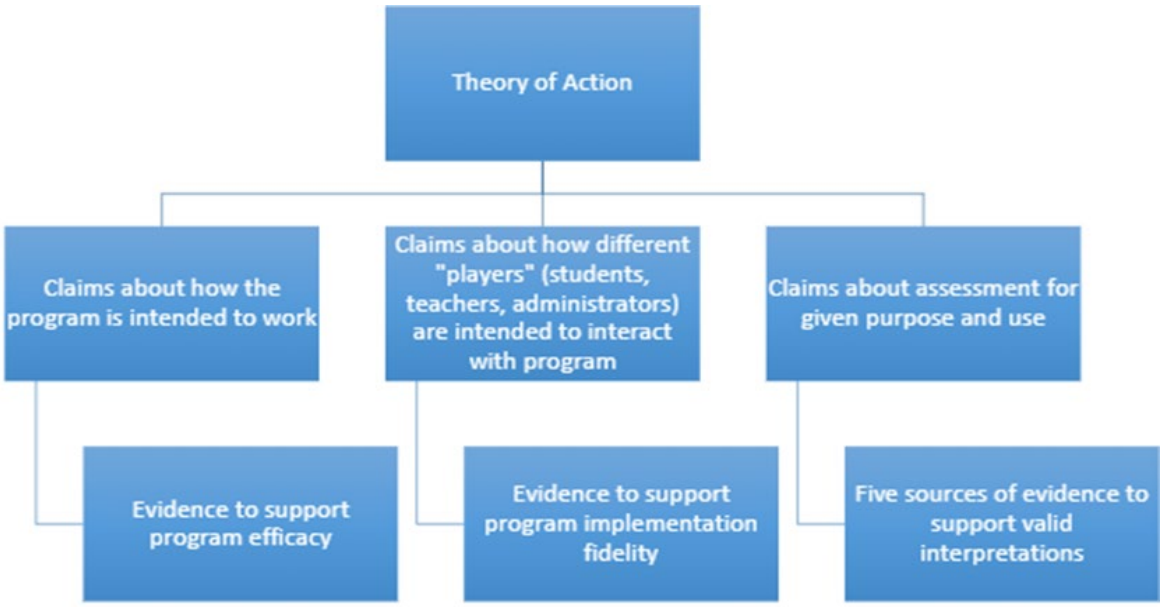


Figure 1.2. Types of Validity Evidence Organized by the Theory of Action (based on Marland & Huff, 2016)

The *five sources of evidence to support valid interpretations* (i.e., test content, response process, internal structure, relationships with other variables, and test consequences) concern the various claims that are made based on the theory of action along with the source of validity evidence that supports each claim. In Chapter 12, the evidence will be woven into a narrative using the five sources of evidence as a framework to support valid interpretations. It should also be noted that, while they are beyond the scope of *i-Ready Inform* and of this technical manual, evidence to support efficacy and implementation fidelity of the comprehensive *i-Ready* system (bottom level of Figure 1.2, two leftmost boxes) can be found on the Curriculum Associates website (CurriculumAssociates.com) and on *i-Ready Central*® (i-ReadyCentral.com).

1.4.4.3 The Technical Advisory Committee

Curriculum Associates has convened a Technical Advisory Committee (TAC) composed of internationally recognized experts in assessment and psychometrics [5]. Key internal stakeholders such as the psychometrics, assessment design, and assessment editorial teams meet with the TAC on a regular basis to seek advice on

[5]As of 2025, the TAC is composed of Dr. Derek Briggs, University of Colorado Boulder, Dr. Andrew Ho, Harvard Graduate School of Education, Dr. Erika Landl, The Center for Assessment, Dr. Richard Leucht, University of North Carolina Greensboro, and Dr. Ye Tong, National Board of Medical Examiners.

pertinent issues. Such engagement with our TAC helps to ensure that our content development, assessment design, and psychometric solutions withstand scrutiny from impartial experts.

1.5 Organization of the Technical Manual

This technical manual provides a detailed description of the critical processes and procedures Curriculum Associates used to develop, administer, and report scores for *i-Ready Inform* and Growth Monitoring assessments. Furthermore, it reports results from a recent administration year and validation studies as sources of evidence for the overall validity argument. As we describe *i-Ready Inform*, we will note the intention of the design and situate evidence in support of this overarching validity argument.

1.5.1 Chapter Structure

The first five chapters of the technical manual focus on procedural evidence of validity, describing the development and administration of *i-Ready Inform*. Chapter 1 provided the foundations of the assessment, including description of the intended (i.e., target) student population, theory of action, and assessment claims, organized around the validity framework. It also described the assessment's purpose, intended uses and interpretations of scores, and an overview of the key system features. Chapter 2 describes content development processes, including the creation of the *i-Ready Inform* item pool, as well as how items were field tested and how the pool is monitored. Chapter 3 describes the *i-Ready Inform* administration process and accessibility features. Chapter 4 describes the computer-adaptive test (CAT) algorithm, including how items are selected and how test flows are managed by *i-Ready Inform*. Chapter 5 provides the psychometric procedures for building and maintaining the *i-Ready Inform* vertical score scale, including item response theory (IRT) calibration, scaling, and equating. Also included in Chapter 5 are score estimation and quality control procedures.

The next portion of the technical manual presents evidence of internal and external validity. First, Chapter 6 provides test reliability metrics and how these metrics are used to report and monitor internal consistency of the assessment. Next, the manual summarizes work that supports the valid interpretation and use of *i-Ready Inform*'s criterion-referenced scores and placement levels. Then, Chapter 7 describes the standard-setting process used to establish the cut scores for the *i-Ready Inform* placement levels. Chapter 11 then details the various types of reports that can be generated to understand and contextualize student scores and various levels of aggregations of those scores.

i-Ready Inform also provides norm-referenced scores and growth scores; Chapter 8 addresses the development of these normative scores, including current percentile rank tables. Chapter 9 addresses Typical Growth and Stretch Growth measures, which rely on the vertical scale. Chapter 10 addresses projected proficiency—combining growth model research and linking studies between *i-Ready Inform* and state summative tests—to report students' performance on *i-Ready Inform* linked to the state test scale and how projected state test proficiency varies based on growth relative to the *i-Ready Inform* growth measures.

In the final chapter, we summarize and organize the validity evidence in support of the intended interpretations and uses of scores we presented earlier in this chapter. Chapter 12 demonstrates how each piece of evidence supports the *i-Ready Inform* assessment claims and how the assessment itself aligns to its validity framework as described in Chapter 1.

Chapter 2: Test Design and Development

2.1 Chapter Summary

Chapter 2 documents how Curriculum Associates approaches content and item development for *i-Ready Inform*. The first section describes how content is specified, starting with a conceptual framework grounded in best practice and research findings. The major content areas of reading and mathematics are defined and related to the development of construct maps, placement levels, and instructionally relevant assessment scores. The next section describes how assessment items are specified, including how content is sampled, cognitive processes are isolated and measured, and what item types are employed for *i-Ready Inform*. The following two sections describe the ways Curriculum Associates ensure students have full access to the assessment content, including a discussion of accessibility supports and accommodations, as well as the Spanish transadapted version of the *i-Ready Inform* for Mathematics assessment. The final three sections of the chapter describe how items are written and reviewed, how items are selected for the assessment using test specifications, and how items are field tested in *i-Ready Inform*.

2.2 Introduction

The National Research Council (2001) recommended that assessments be based on the best available understanding of how students represent knowledge and how they develop competency in the domain of interest. A *construct* is the body of knowledge, understanding, skills, and/or abilities that an assessment seeks to measure. Assessments generally measure a hierarchy of constructs, with the highest level (i.e., coarsest grain size) representing a subject area and the lower levels representing increasingly fine-grained skills. For *i-Ready Inform*, the highest-level constructs—referred to as major content areas—are reading and mathematics. Reading is composed of six domains: Phonics, Phonological Awareness, High-Frequency Words, Vocabulary, Comprehension: Literature, and Comprehension: Informational Text. Mathematics is composed of four domains: Algebra and Algebraic Thinking, Number and Operations, Geometry, and Measurement and Data.

2.2.1 Content Classification Taxonomies

Curriculum Associates adopted an approach grounded in best practice and research to define the test constructs. Initially, the construct definition process entailed the use of construct maps, and it was later extended to incorporate Placement Levels. This section describes how the test construct has been defined and operationalized over time to reflect fidelity to the assessment's conceptual framework.

At present, there are two parallel taxonomies used to categorize items according to the skills that they measure. The elements of these taxonomies are briefly defined next to orient the reader to the key concepts and how they are related.

Grade: All items are aligned to a single grade. Students can receive off-grade items with some exceptions (see Chapter 5 for additional details).

Subject: The major content areas assessed are referred to as subjects. For the *i-Ready Inform* and Growth Monitoring assessments, these are Reading and Mathematics.

Placement level: Providing the foundation for a criterion-referenced interpretation of scores, placement levels were articulated and are mapped to our score scale as part of the standard-setting process.

Domain: The finer-grained constructs within the major content areas of reading and mathematics are referred to as domains.

Indicator: The broad skills represented within each domain and grade can be referred to as indicators and may include content with a wide range of difficulty. They are also aligned to state standards.

Summary claim: Like indicators, summary claims describe the broad skills that are represented at each grade and domain and may also include content with a wide range of difficulty. They are also aligned to state standards. Summary claims and indicators are of a similar grain size.

Anchor claim: The most granular level of skills categorization to which items are aligned in the PLD framework is the anchor claim. Anchor claims are assigned to a summary claim and a single domain placement level.

Item: At present, items are aligned to both indicators and anchor claims.

The *placement-level descriptor (PLD) framework* refers to categorizing items using summary claims, domain-level placements, and anchor claims; the *indicator framework* refers to categorizing items using indicators. A more detailed description of these frameworks follows in the next section.

2.3 The Indicator Framework

Construct maps detail the finer-grained skills within each domain, referred to internally as *indicators*, that should be mastered at each grade level (Wilson, 2005). The construct maps represent how a given construct develops along a difficulty continuum and were created for each domain based on recommendations from an advisory board, subject-matter experts (SMEs), CCRS, well-established practices by professional groups, and experienced researchers. Note that the term *construct* can be used to refer to various levels of granularity with respect to the object of measurement; our use of the term here refers to the skills described by indicators rather than the major content area or domains of the assessment.

Each construct map went through numerous reviews and revisions by SMEs, assessment designers, and pedagogical and instruction experts to ensure instructional relevancy. The resulting construct maps served as a basis for test development and as a guide to item writing. At present, the *i-Ready Inform* score reports (see Chapter 11) remain centered on the indicators that are connected to these construct maps, but item-writing practices have shifted in recent years to leverage the advantages of the PLD framework.

2.4 The PLD Framework

2.4.1 Summary Claims

A summary claim is a description of the knowledge and skills that students are expected to acquire in a particular domain and grade level. It is a broad statement that summarizes the learning objectives for students and provides an overview of the content they should know and be able to do. They describe the knowledge and skills in which students should be proficient in the various domains within reading and mathematics.

Although the grain size is similar, indicators and summary claims are two different taxonomies of organizing skills. Indicators are aligned to one or more summary claims. Several examples from the Grade 5 Number and Operations domain are shown in Figure 2.1 below.

2.4.2 Placement Levels

Placement levels—which are referred to elsewhere in the assessment industry as *performance levels* or *achievement levels*—are used in Grades K–12 assessment programs to set ranges on a test score scale that define and categorize the measured knowledge, skills, and abilities of students in useful ways. The Standards for Educational and Psychological Testing (AERA et al., 2014) define a performance level as the "[l]abel or brief statement classifying a test-taker's competency in a particular domain, usually defined by a range of scores on a test." Our placement levels create the criterion-referenced context for the interpretation of scores. The standard-setting procedures that were used to define and map placement levels to the score scale are described in Chapter 7. Current placement level cut scores can be found at CurriculumAssociates.com/Reviews/Assessment/Diagnostic-Scores-Placements.

2.4.2.1 Placement-Level Descriptors

A *placement-level descriptor (PLD)* describes qualitatively what test-takers know and can do at each of the placement levels. The use of such descriptors has long been associated with standard setting and score reporting. However, in recent decades, performance standards have been used to guide test design and development in advance of setting cut scores or preparing score reports. PLDs are generally developed within a system of descriptors that vary by their granularity, which, in this case, refers to the degree of specificity the PLD provides and relates to the purpose and intended use (Egan et al., 2012). Curriculum Associates therefore developed the PLDs within a system of interrelated descriptor levels that vary in their intended uses and audiences, from very general (*policy-level PLDs*) to very specific (*range PLDs*). The policy-level PLDs are shown in Table 2.1 and apply to all grades across the entire program.

Table 2.1. Policy-Level PLDs

Placement Level	Description
Below Grade Level	Remediation focused on below-grade level material is recommended to help fill in gaps in students' foundational knowledge. Students in this level are not close to meeting the expectations of CCRS for their grade level.
Early Grade Level	Students in this level will benefit from on-grade level instruction to help them meet the expectations of CCRS for their grade level. Students in the early grade level have only partially met these grade-level expectations.
Mid Grade Level	Students in this level will benefit from instruction in late on-grade level topics. These students have met the minimum requirements for the expectations of CCRS in their grade level.
Late Grade Level	Students in this level will benefit from late on-grade level enrichment and will be ready for instruction focused on topics typically covered in the beginning of the subsequent grade level. Students in the late grade level have successfully met or surpassed the grade-level expectations of CCRS.
Above Grade Level	Students in this category will benefit from above-grade level instruction. Students in the above grade level have successfully met or surpassed all the expectations of CCRS for their grade level as well as some expectations from subsequent grade levels.

2.4.3 Anchor Claims and Range PLDs

Next, range PLDs were created. The range PLDs articulate the skills within each grade level, the progression of those skills within and across grade levels, and the expected complexity of those skills across a range of skills/competency levels.

The range PLDs are composed of anchor claims, which are more detailed statements of evidence that explain what is expected of students in relation to the summary claim and at various placement levels. Each item is mapped to a single anchor claim, and anchor claims are each aligned to a single summary claim and placement level combination. Therefore, anchor claims are statements of evidence that explain what is expected of students in order to demonstrate the knowledge and skills of the summary claim at each of the three on-grade placement levels (Early, Mid, and Late) and to be considered proficient in a particular subject area or domain. Figure 2.1 below provides an example of how each of the components discussed in this section relate to each other.

Grade Level and Domain	Indicator Code(s)	Summary Claim	Early Grade Level	Mid Grade Level	Late Grade Level
Grade 5 Number & Operations	INO5-1	Add and subtract decimals through hundredths.	Student represents addition or subtraction of decimals through hundredths using a base ten model or a number line. Student adds or subtracts decimals with the same number of non-zero decimal place values through hundredths using a visual model.	Student adds or subtracts decimals with the same number of non-zero decimal place values through hundredths, presented in a real-world context, without the aid of a visual model. Context does not include dollar amounts.	Student adds or subtracts decimals with different numbers of decimal place values through hundredths, presented in a real-world context, one of which may have zero decimal place values. Context does not include dollar amounts. Student explains a process for adding or subtracting decimals through hundredths.
Grade 5 Number & Operations	INO5-2 INO5-3	Multiply and divide three- and four-digit whole numbers by two-digit whole numbers.	Student multiplies a three- or four-digit whole number by a two-digit whole number.	Student divides a three- or four-digit whole number by a two-digit whole number to find a whole-number quotient.	
Grade 5 Number & Operations	INO5-4	Interpret a fraction as the division of the numerator by the denominator.	Student expresses a fraction, a/b, as a division expression, $a \div b$, or a division expression, $a \div b$, as a fraction, a/b, where a and b are represented symbolically, or represented numerically with $b > a$.	Student expresses a fraction, a/b, as a division expression, $a \div b$, or a division expression, $a \div b$, as a fraction, a/b, where a and b are represented numerically or in a real-world context, and b is less than or equal to a.	

Figure 2.1. Illustration of Summary and Anchor Claims: Number and Operations

2.4.4 A Principled Approach to PLD Development

2.4.4.1 Instructional Relevance and Consideration of Consequences

SMEs were trained to approach the development of the PLDs as being grounded in instructional relevance, including the consequences or outcomes of the assessment. Using a principled approach, writers started with instructional goals. They considered the full set of content and skills (or indicators, in our case) that curriculum/instruction and assessment are intended to reflect, including the conceptual organization and prioritization of the content. The SMEs then modeled the domain, creating claims and specifying evidence to support those claims. Finally, they created a framework that defined how items could be developed for the modeled domain within the constraints of the assessment. This framework became the item specifications used for item development.

The consequences of using the summary and anchor claims were considered at this stage, with an emphasis on instructional recommendations based on the assessment results. To this end, the PLDs were worked into the principled approach to assessment design because the PLDs were developed to be used for:

- Aligning items to more granular levels than is possible with indicators
- A basis for item specifications with a reference tool for writing content mapped to Early, Mid, and Late placement levels
- Setting future performance standards (i.e., determining cut scores)
- Reporting statements about what students likely can do and what they should work on next (i.e., *Can Dos and Next Steps*, see Chapter 11 for more details) with ties to anchor claims.

Since the primary function and purpose of *i-Ready Inform* is to make appropriate instructional recommendations and placement decisions for students performing at different levels, the transition to the new PLD framework has precipitated extensive, ongoing research to collect validity evidence for criterion-referenced score reporting and instructional recommendations that emerge from *i-Ready Inform*. See Chapter 7 for more discussion of the placement levels with respect to standard setting and cut scores and Chapter 9 for a discussion of the placement levels in relation to growth targets. The relationship between the PLD framework and item development is discussed further in the remaining sections in this chapter.

2.4.4.2 Horizontal and Vertical Articulation

In the development of the PLDs, attention was paid to the horizontal alignment of content (i.e., within grade) within a summary claim. The SMEs started with the indicators available within each grade. They worked with instruction team members to explain the progression of skill development (i.e., anchor claims) as students work toward, and sometimes beyond, proficiency status relative to the summary claim.

Measuring student achievement from one year to the next with a vertical scale depends on the vertical articulation of content over the grade levels. Curriculum Associates used the PLDs as extensions of the construct maps to build in the vertical articulation of content. Using the PLDs to guide item development is critical to supporting the claim that *i-Ready Inform* allows for an accurate assessment of student knowledge that can be monitored over time to gauge improvement. SMEs wrote items for the item pool to ensure that content was clearly articulated across grades. They asked the following question: Do changes from one grade to the next represent appropriate developmental and educational progression? The related PLDs were designed to be appropriately differentiated from one grade to the next. Knowledge and skill expectations increased in the next grade. When related (summary) claims were indistinguishable over two consecutive years, clarifying information was included to explain how the expectations are different. They also worked to ensure no unreasonable leap in expectations from one grade to the next.

2.4.4.3 The Relationship between Indicators and PLDs

The existing indicators served as a foundation for the development of the components of the PLD framework. Because the PLDs are grounded in the indicators, and because the indicators are grounded in CCRS, the PLDs reflect CCRS and articulate the skills that guide current item development.

SMEs who write items begin by familiarizing themselves with the PLDs for each grade level. As a result of the robust CCRS coverage and carefully designed horizontal (i.e., within-grade) and vertical (i.e., grade-over-grade) progressions within the PLDs, these skills can be used to align items to CCRS by:

- Targeting individual items to one anchor claim as an intentional way of ensuring that the item bank includes content to measure a specific part of CCRS
- Ensuring proper coverage of all anchor claims through carefully created item development plans that reflect the breadth, depth, and range of complexity of the grade-level content, according to CCRS

2.4.4.4 Transitioning from the Indicator Framework to the PLD Framework

Curriculum Associates is currently in the process of transitioning from using the indicator framework to using the PLD framework for reporting. To accomplish the transition successfully, a crosswalk procedure was applied to match indicators to both CCRS and the summary claims in the PLD framework. All items in the bank were then also aligned to the PLD framework. Each item in the existing bank was aligned to the new PLDs by two independent SMEs using rigorous adjudication processes. SMEs considered the scope and sequence of content in the development of PLDs to ensure coherence.

Empirical item score data were examined to look for expected increases of item difficulty for groups of items to verify that content "unfolds" in the order reflected in the progression through the placement levels. This is one way to empirically validate that PLDs articulate a coherent progression through a grade level, and to ensure vertical articulation across grades. Item alignment analyses have further informed refinements to the PLDs and item or item alignment revisions.

The types of adjustments that were made to the PLDs or items based on analyses that identified items as inconsistent with the vertical articulation were as follows:

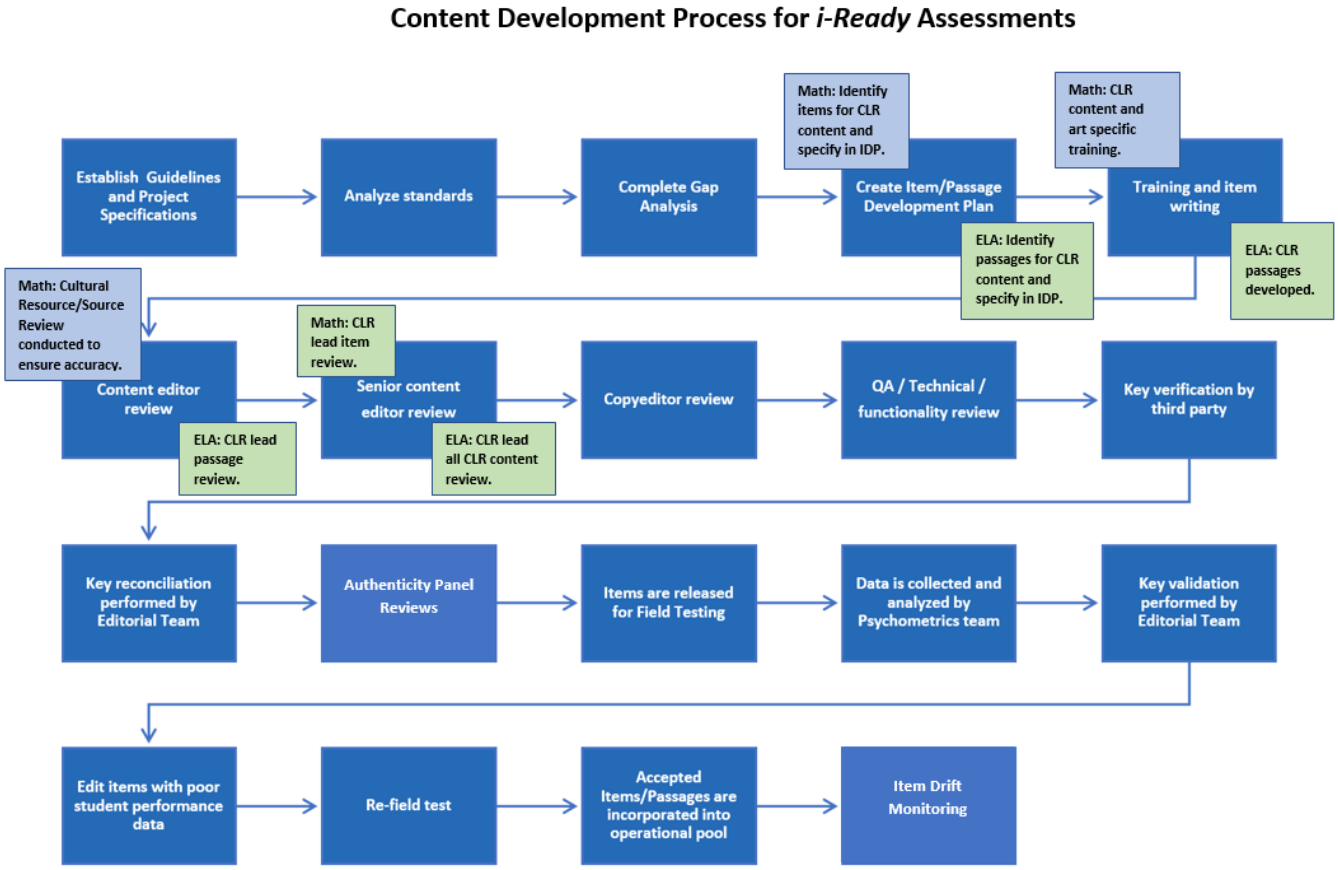
- Disable or replace the inconsistent item
- Edit and re-field test the inconsistent item
- Edit the placement level of the aligned anchor claim (e.g., change the anchor claim in the PLD from a Mid placement to an Early placement)
- Edit the alignment of the item to a new anchor claim

On the path to a full in-system implementation of PLDs, the small number of items that do not align to a PLD are being disabled in the bank and will no longer be eligible for selection for test events.

2.5 Item Development

2.5.1 Process

Item development, shown in Figure 2.2, is a comprehensive process involving many steps. The process encompasses specification and guideline development, analysis of content standards, development planning, and training SMEs to write items with fidelity to the specifications. Editorial review and quality assurance are followed by reviews for fairness, field testing, and empirical evaluation and validation. Each of these steps will be described in narrative form in the following subsections, with key terms identified in italics.



Note: CLR: Cultural and Linguistic Relevance; ELA: English Language Arts (i.e., Reading); QA: Quality Assurance

Figure 2.2. Overview of the i-Ready Inform *Development Process*

2.5.2 Planning

The first step in the item development process is to outline the *guidelines and specifications* for the process, which are the steps and details that follow. Next, item development is driven in part by a *gap analysis* that details the robustness of our item bank relative to anchor claims and *standards* across both major content areas, all associated domains, and all grades. From this analysis, the details of an *item/passage development plan* can be articulated, including timelines, staffing, and production goals.

2.5.3 Item Writing

2.5.3.1 Training

Consistency in item design is established throughout the assessment development process to provide students with a coherent testing experience. SMEs writing items and conducting reviews follow a strict set of item-writing specifications that guide item formatting, question prompts, item types, and both audio and visual presentation. Item writing was guided by the *i-Ready Inform* Item Specifications for each content area.

For each content area, domain, and grade level, writers participate in in-depth training in which they practice developing items that are appropriately aligned to specified anchor claims, each of which resides in a single placement level. They are also trained to write items to specific levels of cognitive complexity as described by both depth of knowledge (DOK) (Webb, 1997, 1999) and aspects of rigor (AOR) (Achieve, 2019), as appropriate to the subject matter, which are discussed in greater detail below.

In the training, writers are provided with a range of available item types and classifications as well as direct instruction on the virtual tools available to them, including manipulatives like counters, charts, base-10 blocks, rulers, protractors, and calculators for mathematics. Exemplar items are provided for each domain and item type as well as explicit direction on content limits and accessibility considerations. Stimulus features are also specified as well as the characteristics, rationale, and limitations of distractors.

2.5.3.2 Measures of Cognitive Complexity

The cognitive complexity, or cognitive demand, of an item is another area of specification used to develop *i-Ready Inform* and Growth Monitoring items. These levels describe the relative complexity in terms of the content (e.g., simple versus complex graphs; literal versus figurative language) and nature of the task (e.g., solving routine versus non-routine problems). Each content-area assessment uses an appropriate definition of cognitive complexity and a specific approach to guide item development.

The level and complexity of the skill and cognitive demand in the PLDs is a major driver for targeting item complexity and cognitive demand. For example, in the mathematics items, the action verb(s) in a PLD describe increasing sophistication of engagement with the content as students progress through the year. To ensure consistency in how (1) PLD creators phrased the PLDs, (2) item developers interpret the verbs in the PLDs, and (3) item writers apply the verbs in writing individual items, the PLD creators collaborated to co-create a Math Skills Glossary, which defines and categorizes verbs found in the PLDs according to both procedural and conceptual complexity.

For example, the following PLDs aligned to the summary claim:

> *Contrast and describe two numerical data distributions using measures of center and/or variability.*
> *Evaluate two numerical data distributions using measures of center and variability.*

are clearly articulated across the placement in Grade 7 using increasingly sophisticated evidence delineated by the verbs, describes, contrasts, and evaluates as shown in the following table:

Table 2.2. Example Anchor Claim Text

Placement	Anchor Claim Text

Early	Student describes the degree of visual overlap of two numerical data distributions with similar variabilities.
Mid	Student contrasts two numerical data distributions using measures of center and/or variability when given a visual representation, such as box plots or dot plots.
Late	Student evaluates inferences made from numerical data for two populations using measures of center and/or variability.

The Reading items and PLDs on the *i-Ready* assessments use a similar process. The cognitive complexity found in the i-Ready PLDs correlates to—and is informed by—other measures of cognitive complexity such as Webb's Depth of Knowledge (DOK) (1997,1999) and Achieve's Aspects of Rigor (AOR; 2019). Note that cognitive complexity is different from item difficulty, which is determined through item calibrations (see Chapter 5).

2.5.3.3 Item Specifications

Item specifications provide detailed expectations for the development of each item and ensure that items are mapped to the intended anchor claim. Knowledge beyond the skill level of the anchor claim should not be necessary to answer the item correctly, and SMEs review the items and the adjacent anchor claims—which may be in different placement levels—to ensure this is the case. In general, item specifications cover alignment to content and cognitive complexity, characteristics of item content, quality of the item stem, quality of the item answer options, and rationale for the answer choice (see Table 2.3 for additional details). Note that there are also specifications for creating art to ensure it renders properly and meets online accessibility guidelines and the use of online tools.

Table 2.3. General Considerations for Quality Item Development for *i-Ready Inform*

Alignment, DOK, and Rigor
Item assesses the anchor claim it is aligned to.
Assigned cognitive complexity measure is accurate.
Item meets the cognitive demands of both the anchor claim and the complexity designation.
Item requires comprehension of the passage or multimedia element to be answered.
Item does not require prior knowledge.
Item has only one correct response. No arguably correct answers are included.
Item has less than a 50/50 chance of the student guessing the correct answer.
Item Content
Items contain grade-appropriate language, vocabulary, and content.
Item type is appropriate and effective for the content of the item (especially technology-enhanced items [TEIs]).
Cueing is avoided between items in a passage set.
Overlapping/repeated content is avoided between items.
Items are clear, concise, and complete. • Items include accurate and sufficient content to complete the item. • Items have correct punctuation and are grammatically correct. • Items are free from spelling and typographical errors.
Items within a passage set cover as much of the text as possible and ask meaningful questions toward understanding the overall content.
Stems
Closed stems are clear, succinct, and follow assessment best practices in construction: • Avoid the use of negatives (e.g., "not," "none") in the stem. • Avoid excessive wording (e.g., "of the following"). • Only necessary words (1–2 at most) should be emphasized in all caps in the stem. • Fiction-based item stems are in present tense.
Answer Options

Alignment, DOK, and Rigor
Avoid cueing within an item (e.g., repeating a word from the stem only in the correct answer).
Distractors are related to the anchor claim being assessed.
All answer choices are plausible and related to or derived from the text without contradicting the text.
Answer options follow assessment best practices in construction: • Avoid outliers. (Answers, or pairs of answers, should all be comparable in language, length, complexity, grammatical form, and have a balance in connotation.) • Do not include pairs of opposites in answer choices. • Avoid throwaway options. • Avoid negative distractors following a negative stem (i.e., double negatives). • Avoid items having absolutes, "none of the above" or "all of the above" as distractors.
Answer-Choice Rationale Criteria
Correct-answer rationales should clearly and concisely explain why an answer choice is correct in relation to the anchor claim, including appropriate references to text content. Each rationale should: • Include relevant language from or about the assessed anchor claim • Be followed by a text-dependent explanation of why the response is correct
Incorrect-answer rationales must describe the mistake, misconception, or misinterpretation the student has likely made as related to the skill in the anchor claim. Each rationale should: • Clearly and concisely explain why a distractor is incorrect in terms of its plausibility within the text • Be followed by text-dependent explanations and/or references that support why the response is incorrect in relation to the assessed anchor claim

2.5.3.4 Item Types

Items for *i-Ready Inform* utilize a range of item types. Multiple-choice items written for Grades K and 1 have three response choices, whereas multiple-choice items written for Grades 2 and above have four response choices. While many items are multiple choice, *i-Ready Inform* also contains drag-and-drop, dropdown menu, number line, short answer, and text highlight items. TEIs continue to be a major focus of our item development process. All *i-Ready Inform* and Growth Monitoring items are automatically scored as either correct or incorrect.

2.5.3.5 Passage Specification

All reading passages are either created or chosen from the public domain according to comprehensive specifications. Specification of passage development starts with the topics to emphasize. Here constructive and altruistic activities (e.g., sports and physical activities, chores, family and household support, acts of kindness, charitable activities), educational activities and aspirations, environmentally friendly activities, and topics that embody cultural and linguistic responsiveness are emphasized. For older students, civic involvement, work, entrepreneurship, and careers are emphasized.

Passages are evaluated for their word counts, text complexity, range of genres, and relative decodability using the Phonics and High-Frequency Words domains. Vocabulary is reviewed with reference to graded word lists, and SMEs take care to identify problematic words and the level of context. Since *i-Ready Inform* is a CAT and students may read passages at various grade levels, specific mentions of grades are avoided in passages (see Chapter 4 for details about the grade-level eligibility of items).

Quantitative measures such as Lexiles® and the Flesch–Kincaid (F–K) reading ease scores are also considered in passage evaluation. These measures reflect such characteristics as the total number of words and sentences, the average syllable count per word syllables, the average number of words per sentence, and the readability level of the text. A qualitative review of passages considers whether a passage is grade-level appropriate in

terms of overall conceptual load. SMEs ask questions such as, *Is the topic too mature/complex for the student? Is there too much information to comprehend? Is there too little information provided to comprehend the topic? Can the student (without teacher assistance) understand the information as presented?* In addition, attention is given to bias and sensitivity issues by seeking to avoid topics deemed inappropriate for assessment purposes; an internal checklist is used to guide this process.

Passage reviews check source materials for accuracy of information, and introductory text is limited to what is useful to students for context. Passages are presented with specified text features that have clear titles and section headings. Poetry and drama passages may have line numbers, and footnotes can be included where appropriate. Public domain and commissioned pieces are included in *i-Ready Inform*, and culturally authentic texts are included. To ensure consistency, exemplar passages are provided, and passage review checklists are used and checked for quality assurance purposes.

2.5.3.6 Multimedia Guidelines

Art and photographs are also specified. All illustration and photo specifications include age ranges, gender, and ethnic and cultural requirements if the art will include people. Text features, including charts, graphs, maps, or other text features, and alternate text (alt text) for images are included to maximize accessibility for visually impaired students.

When a video clip, audio clip, or specific image used for multimedia is paired with a passage, introductory text must be included. Introductory text might make mention of the origin of the piece, give background information for students to understand the piece, or provide other helpful context that would be relevant for students to understand what they are about to see or hear. Occasionally a multimedia element will fall within a passage. When this happens, a transitional introductory line of text informs the reader that something new is being presented. Audio clips are never more than 90 seconds in length. Note that all Grade K passages include audio.

2.5.4 Authenticity Panel Review

The next stage in the item development process is the authenticity panel review, which is one way we promote authentic cultural representations in our products. Panelists volunteer their time and expertise to help make our products more accurate and respectful of the diverse cultural experiences of teachers and students. We define expertise broadly, meaning it can come from personal experience, research, work with students, and more. Note that item reviews are guided by detailed, content-specific checklists.

2.5.5 Psychometric Analyses

For a newly written item to be added to the *i-Ready Inform* item bank, it must first be *field tested*. Field testing serves two primary purposes. First, it allows us to conduct basic statistical analyses to ensure that the item is functioning as expected and is keyed correctly. This process helps us identify item flaws or unexpected behavior. Second, it allows us to estimate the item difficulty parameter according to the IRT model upon which our score scale is built (see Chapter 5 for more details). Item difficulty parameters are used in the

adaptive item selection algorithm as well as in determining the final score and associated measurement error. Items that do not perform as expected are flagged and sent to SMEs for review through a process known as *key validation*. This discovery process can result in item revisions, which require that they be field tested again. Alternatively, items can be accepted as is or discarded entirely. Finally, existing items in the bank are sometimes re-field tested to determine if they are functioning according to the parameters on file. This is known as an *item drift analysis* (see Chapter 5 for additional details).

2.6 The Accessibility and Inclusion Framework

The purpose of the Accessibility and Inclusion Framework is to articulate an overall vision for an inclusive assessment and how accessibility is included in each component of the assessment. We recognize that technology and student needs change, so our approach includes an evaluation and continuous improvement plan as part of the validation process. It is important to emphasize that we do not consider these goals in isolation nor as an add-on to the other elements of the *i-Ready* technical manual. Instead, this framework is embedded through each element in the *i-Ready* Theory of Action described in Chapter 1 with the goal of having a causal impact on the outcomes in the theory of action for all learners. Specifically, we aim for these design elements to impact the intermediate outcomes of (1) increasing self-directed learning, (2) teachers understanding the root cause of learning problems, and (3) teachers setting appropriate growth expectations. These short-term outcomes lead to the long-term outcome of increased academic achievement. In addition, we specify how we will evaluate the impact of our design considerations through studies that will supplement our validity plan (described below) and how we will prioritize future improvements so that accessibility and inclusion are not an afterthought but instead part of our iterative product improvement and development.

2.6.1 A Systems Approach to Inclusive Design

Our approach to accessibility and inclusion recognizes the importance of inclusive design across all phases of testing, from item development to test score use. In addition, we recognize that assessments like *i-Ready* are part of a larger educational system, so accessibility features and tools should be similar even when the test purpose may result in different policies. For example, some testing accommodations may be appropriate for mathematics but not reading. Likewise, the same accommodation may be useful in an instructional context but not appropriate for summative assessments. For these reasons, our strategy is rooted in a systems-based approach to universal design (Laitusis & Karvonen, 2024). The Universal Design of Assessment Systems framework emphasizes three core principles that impact not only students, but their teachers, parents, and administrators. These principles require that assessment systems (1) meet basic human needs for autonomy, competence, and relatedness, (2) maximize flexibility for learner variability, and (3) promote equity and mitigate unintended consequences. In addition, this framework incorporates federal requirements for accessibility, such as the W3C Web Content Accessibility Guidelines (WCAG) (World Wide Web Consortium, 2022) and ideas about designing for learner variability with origins in Universal Design for Learning (UDL) (CAST, 2024) .

2.6.2 Designing for Self-Directed Learners and Teacher Autonomy

The *i-Ready* Theory of Action specifically mentions students becoming "self-directed learners" as one of the intermediate outcomes. Since many students with disabilities and Multilingual Learners are performing below

grade-level expectations, our framework prioritizes design features that enhance student motivation and engagement and measure student proficiency across the full continuum of achievement levels. We recognize that testing and test scores impact educator motivation and engagement too. For these reasons, both the student and teacher elements of *i-Ready* aim to embed the three key psychological nutrients that allow for appropriate goal setting and foster motivation: autonomy, competence, and relatedness (Ryan & Deci, 2018). The following questions help our design team consider how these psychological nutrients can be embedded within *i-Ready* while maintaining the necessary standardization to allow for comparability.

2.6.2.1 Autonomy for Inclusive Design

Following are some guiding questions around autonomy for inclusive design:

- Do students have access to tools, accommodations, and assistive technology and the ability to independently choose when to use them?
- Do teachers have choice in which accommodations will yield the most instructionally relevant information?
- Do teachers and administrators have a choice of when, where, and how often to administer assessments? For example, are test days and times flexible to avoid religious holidays, cultural events, and student health needs?

2.6.2.2 Competence for Inclusive Design

The following are some guiding questions around competence for inclusive design:

- Does test content elicit evidence of skills within a student's zone of proximal developments (e.g., with adaptive test design, growth scores)?
- Do accommodations and universal tools remove construct irrelevant barriers, allowing all students to show what they know and can do?
- Do test score reports yield interpretable information that aids students and teachers in setting achievable instructional goals?
- Do teacher score reports provide support for actionable instructional next steps?

2.6.2.3 Relatedness for Inclusive Design

The following are some guiding questions around relatedness for inclusive design:

- When possible, are accessibility tools and accommodations embedded in the testing platform to increase student autonomy and reduce stigmatization?
- Do test items reflect a wide range of lived experiences and ways of knowing that reflect diversity in cultures, language, and abilities?
- Are alternative pathways provided when important constructs are not appropriate based on a student's disability or English language proficiency (e.g., adaptive pathways for foundational reading skills in deaf students)?

2.6.3 Designing for Learner Variability

Designing for learner variability is an essential component of the *i-Ready* framework that primarily impacts test design, item development, and test administration. Our CAT design and vertical scale acknowledge the diversity of students performing at all academic achievement levels. Adaptive test designs result in unique challenges because in most current implementations they require that every item in the pool be accessible to students with diverse abilities. Our approach is to evaluate each domain for appropriateness. When a domain is appropriate for a student's unique abilities, we work to make each item in the domain accessible. In some cases, domains are determined to be not appropriate (e.g., audio-dependent phonics items for students who are deaf). For these limited cases, we have incorporated a skip foundational domain feature. For the majority of tested domains, our team is committed to ensuring every item is accessible to all students. This is an ongoing process that includes a broadly defined construct, inclusive item development, and accessible delivery platforms. Some broad areas we consider are described below.

2.6.3.1 Broadly Defined Construct

In *i-Ready Inform*, we recognize that within each construct, students with diverse needs may demonstrate these skills in different ways. For example, foundational reading skills like decoding need to be defined in such a way that students who are deaf can show how they mastered foundational word recognition skills. Questions we consider are:

- Do construct definition teams represent those knowledgeable about disability-specific approaches to instruction and mastery of the construct we aim to measure?
- Are alternate pathways to mastery considered when developing expected progressions in mastery?
- Do teachers have autonomy to select accommodations that may alter the construct but provide information to identify the root cause of a student's learning challenges or set achievable learning goals?

2.6.3.2 Multiple Means of Response and Presentation

While some adaptive tests isolate students with sensory disabilities to a limited item pool, the *i-Ready* development team is committed to ensuring that all items within appropriate domains can be administered to every student. To do this, we aim to comply with the most recent version of WCAG (Worldwide Web Consortium, 2022) Level AA compliance and provide three levels of supports: embedded accommodations, designated supports, and universal supports. These are discussed in detail in Chapter 3. In addition, teacher administration guides are provided when teacher assistance is needed. Our item development and delivery teams consider questions such as:

- Does the item development team include representation of people with lived experiences and knowledge of disability-specific approaches to instruction and assessment?
- What evidence of proficiency will this item elicit when it is presented in audio, sign language, tactile, or visual formats?
- What evidence of proficiency will this item elicit when it is responded to in audio (including sign language), tactile, or visual formats?
- How can this item type meet WCAG guidelines and be perceivable, operable, understandable, and sufficiently robust to work with current and future technologies?

- Does the item pool include a broad representation of lived experiences?

2.6.4 Designing for Safe Use

Designing for safe use is the final core principle of the Universal Design of Assessment Systems. Operationalizing this principle involves professional learning to ensure the assessment literacy of score users, the development of a theory of action, ongoing validity studies, and clear consequences for score misuse. The three core principles of the Universal Design of Assessment Systems outlined above are applicable to all test-takers, but special attention is needed for students with disabilities and Multilingual Learners. By integrating these principles of accessibility and inclusion with each key component of the *i-Ready* theory of action, we can better serve the diverse needs of all students, thereby fostering an inclusive learning environment where every student can thrive. In this section, we articulate how the core design principles apply to each component. Since student demographic information is protected by the Family Educational Rights and Privacy Act (FERPA), action needs to be taken to identify schools and districts willing to collaborate on research to evaluate the *i-Ready* theory of action for all student groups. Once partner schools are identified, both quantitative and qualitative research should be conducted to answer fundamental validity questions related to safe use such as:

Do special education and general education teachers report similar utility of *i-Ready* scores for:

- Measuring student learning
- Increasing self-directed learning
- Understanding the root cause of learning problems
- Setting appropriate growth expectations
- Understanding score reports and how scores can be used

Do special education and general education students report similar abilities to:

- Access test questions independently
- Answer test questions independently
- Show what they know and can do (with appropriate accommodations)
- Achieve growth targets (included in score reports)
- Understand score reports and how scores can be used

2.6.5 Ongoing Evaluation and Iterative Improvement

Since validation is an iterative and ongoing process, the accessibility features of *i-Ready* evolve as new assistive technologies become available and additional students' needs are identified. We are committed to using teacher and student feedback gathered through self-report surveys, cognitive labs (see Chapter 12), and systematic analyses of customer feedback to improve *i-Ready* accessibility and inclusivity. In addition, we will actively seek partner schools to collaborate on secondary analyses of user data. These studies would include differential item functioning (DIF) (by accommodation, language, race, gender, and/or disability group) and analyses of usage data from teachers in different settings (general education inclusion, specialized instruction classrooms, or specialized schools). As technology advances and our network of research partner schools grows, we plan to evaluate how iterative changes in our design impact academic growth by student groups.

2.7 Transadapted Spanish *i-Ready Inform* in Mathematics

At the start of the 2019–2020 academic year, Curriculum Associates began offering the transadapted Spanish version of both the *i-Ready Inform* and the Growth Monitoring for Mathematics assessments to support students and educators with useful data on mathematics skills proficiency in Spanish, allowing teachers to better understand the mathematics content knowledge and skills of Spanish-speaking English Learners and students receiving mathematics instruction in Spanish, independent of their English language proficiency.

In the Spanish mathematics item bank, items were transadapted from the English item bank following the *International Test Commission Guidelines for Translating and Adapting Tests* (ITC, 2017) using the translation design of double independent translation plus adjudication. The primary goal of translating the item bank is to deliver Spanish items that are equivalent in meaning and difficulty, maintain the intended reading level of the English item, and that are culturally and linguistically appropriate.

When translating an item for the English assessment, our assessment editorial team took into consideration the cultural diversities in commonplace objects, activities, and sports that would reflect the experiences of the cultures of the Spanish-speaking students in the *i-Ready* population and tried to include concepts that would be more familiar to our audience instead of going for a straightforward translation. Notions and concepts that would be foreign to the Latin American cultures were localized for familiar concepts, or vocabulary, without changing the construct measured by the item.

i-Ready Inform was transadapted into a standard variety of Spanish with the different nationalities of US Spanish speakers in mind. During the translation process, the directive and aim was to translate into a standard Latin American dialect. This was ensured by using translators, reviewers, copy editors, and voice artists with backgrounds comprehending the different regions of Latin America as well as the use of terminology that would be more standard and accessible to all. Reference materials include the internal Curriculum Associates English–Spanish glossary, and other Curriculum Associates products in Spanish, such as Teacher Toolbox.

2.8 Embrace All

Curriculum Associates' mission is to make classrooms better places for all teachers and students. Curriculum Associates believes that the combination of engaging materials and strong pedagogy encourages academic excellence and leads to high achievement for all students, regardless of cultural identity, linguistic background, economic status or circumstance, and disability. Curriculum Associates also believes that connection creates impact. Connection requires relevance. Relevant instruction occurs in classrooms where students feel safe and accepted. Furthermore, Curriculum Associates believes in:

- Celebrating students' identities and helping teachers use students' experiences to shape content and instruction
- Helping all students become independent learners and engaged global citizens

These are the values with which Curriculum Associates developed *i-Ready Inform*.

2.8.1 Bias and Sensitivity

Note that bias in an assessment is the presence of some characteristic that results in differential performance for two individuals of the same proficiency level but from different student groups. Curriculum Associates strives to avoid bias by developing content that is culturally and linguistically responsive. Bias may occur through over- or underrepresentation of any given student group. Our content represents student groups as realistically as possible and is varied in terms of ethnicity, national origin, race, region/location, age, gender, socioeconomic class, religion, and disability. Additionally, Curriculum Associates' programs aim to represent multiple viewpoints, remain true to the culture of origin, and capture everyday experiences of people from different cultures and communities

2.8.2 Development Guidelines

Curriculum Associates' product development guidelines outline key process steps that affect content development of *i-Ready Inform*.

- **Collaborate to develop and document a formal plan** for achieving a culturally relevant and responsive assessment during the discovery and definition phases of development.
- **Aim for a well-rounded development staff**, including writers, editors, designers, artists, vendors, reviewers, and program authors or SMEs, etc., varied communities, including people with disabilities and English Learners.
- **Consider cultural relevance and responsiveness when developing the visual strategy and image plan.** Images are an important form of communication of the meaning and/or point of view of a text. It is important to look at the distribution of representation of the product to ensure representations are appropriately varied. It is also important that background setting (including people) and context be specified, discussed, and potentially tracked in instances when only the foreground is specified.

Chapter 3: Test Administration

3.1 Chapter Summary

Chapter 3 provides detailed evidence of how *i-Ready Inform* is designed to be administered. Ensuring that the assessment is accessible and is administered in a consistent manner helps ensure that scores can be interpreted as intended. We describe steps that test proctors and administrators should take before, during, and after the administration of *i-Ready Inform*. This chapter also describes administrative details such as testing windows, administration times, and expiration rules. *i-Ready Inform*'s accessibility features and accommodations, which seek to expand access to *i-Ready Inform* to as many students as possible (e.g., students with disabilities, Multilingual Learners, etc.) are discussed. Also discussed are test security protocols and monitoring strategies that promote consistent administration practices and provide the appropriate conditions for students to produce responses that best reflect their proficiency, ensuring that scores can be both interpreted and used as intended.

3.2 Introduction

Detailed and thorough test administration protocols are critical to upholding the integrity, reliability, and validity of any assessment system. Administering *i-Ready Inform* is a multiphase process that necessitates careful planning, diligent implementation, and thoughtful follow-up procedures. The *i-Ready Inform* delivery platform is an essential part of this administration process, serving as the tool through which students engage with the assessment. Developed according to best practices in assessment design and administration, the platform offers an effective and user-friendly medium for test delivery and provides comprehensive support and resources for those involved in the testing process.

3.3 Test Administration

3.3.1 Before the Test

i-Ready Inform is typically administered up to three times per year: fall, winter, and spring. The administration process is managed at the school and classroom levels, and the test is delivered through a secure, online system with key roles of administrators and teacher proctors.

3.3.1.1 Role of Test Proctors

The role of proctors is crucial to the effective administration of *i-Ready Inform*. Proctors are responsible for creating a conducive testing environment, ensuring all necessary materials are available, and guiding the students throughout the process. Their role extends beyond the day of the test, encompassing pretest preparations and posttest procedures as well.

3.3.1.2 Overview of the Test Session Setup

Proper setup of a test session on the *i-Ready Connect*™ platform is critical to a smooth administration process. This includes ensuring the availability of necessary equipment, confirming that the test has been assigned to the right students, preparing the testing environment to minimize distractions, and confirming student readiness to take the test.

3.3.1.3 Necessary Equipment and Materials

For the administration of *i-Ready Inform* via the *i-Ready Connect* platform, compatible devices[6] with headphones and reliable internet connections are needed. Additionally, any physical materials needed for non-embedded accessibility supports and accommodations (e.g., scratch paper and pencil, calculator, dictionary, etc.) need to be arranged prior to the beginning of the administration.

3.3.2 During the Test

3.3.2.1 Overview of Testing Experience

i-Ready Inform follows a set of protocols to maintain consistency across test sessions. This includes distinct components such as the initial login, an overview of test rules, practice items, an overview of tools (mathematics only), and responding to the actual assessment items.

i-Ready Inform includes intermittent breaks (e.g., Brain Breaks, game breaks) to maximize student performance and promote investment and engagement with the assessment. It also has built-in features to accommodate potential interruptions and allow students to complete the test over multiple shorter sittings. Students can pause and resume when needed. Due to the adaptive nature of the test, most students are expected to respond correctly to approximately 50 percent of the items administered (see Section 5.4.1 for more information).

3.3.2.2 Test Expiration Rules

All students in a district, school, and grade are recommended to complete the test within the same two-to-four-week window to support valid data comparison. An assessment remains active in student accounts for a specific duration before expiring. Administrators and proctors should monitor the status of *i-Ready Inform* assessments to ensure that students complete each assessment as soon as possible once started.

An *i-Ready Inform* test event expires 28 days from when a student initiates the assessment, and a new test event is automatically reassigned. A Growth Monitoring assessment expires seven days after a student begins. If a student is actively working on *i-Ready Inform* at the time it is set to expire, the expiration will occur if they exit the test prior to completion.

When a student has seven days or fewer to complete their in-progress *i-Ready Inform*, a red hourglass symbol with the number of days remaining before expiration will appear next to the student's name on the *i-Ready Inform* Status report and teacher dashboard.

3.3.2.3 Monitoring Student Progress and Activity

The role of proctors during the test includes active monitoring of student progress and engagement. Proctors observe students during an administration to ensure they are on task and encourage them to exert their best effort. The *i-Ready* platform has features that support real-time progress tracking, enabling proctors to oversee each student's test status and expiration date. Teachers can reset a test in progress if a student appears to be rushing through the test.

---

[6]Information regarding system requirements and supported devices can be found here: https://cdn.i-ready.com/instruction/content/system-check/iReady_System_Requirements.pdf.

3.3.3 After the Test

3.3.3.1 Posttest Procedures

Following an administration, Curriculum Associates recommends that teachers check the Inform Results report for any students with yellow or red Rush flags. Rush flags use item-level timing data to determine if it is likely that a student progressed through the test more quickly than would be expected if the student was demonstrating a sufficient level of effort. If a student takes less than a specified amount of time to complete a test event, a flag will appear next to the Overall Placement & Scale Score if it is probable (yellow flag) or evident (red flag) that a student rushed through the test. This alert means the student's test results should be evaluated for alignment to other indicators of the student's performance. Depending on the district's policy, the student may be asked to retake the test; however, the original *i-Ready Inform* results remain in the system and are not overwritten.

Rush flags are based on extensive item-level timing research and are determined using the following criteria:

- **Yellow Flag:** Student rushed on at least 10% of items AND got less than 45% of items correct, OR the total test duration was less than eight minutes.
- **Red Flag:** Student rushed on at least 25% of items AND got less than 43% of items correct.

Test duration is calculated as the sum of all item durations to the nearest second; intermissions are calculated as zero duration.

3.3.3.2 Reviewing Results, Setting Goals, and Making Instructional Decisions Based on the Results

After the test, the results are reported to educators to help them enhance student learning, as shown in the Theory of Action in Chapter 1. Teachers are encouraged to review the outcomes of *i-Ready Inform* in collaboration with students as they set academic goals. Reports also identify areas of strength and recommend instructional priorities. If Personalized Instruction is enabled for the student, they are automatically assigned a set of lessons based on their *i-Ready Inform* results, inform instructional decisions, and ensure teaching strategies align with the specific needs and competencies of each student.

3.3.4 Resources and Support Materials for Administrators and Proctors

To facilitate smooth and effective test administrations, a comprehensive library of resources and support materials is available on *i-Ready Central*, an online repository of resources developed by Curriculum Associates available to districts that have purchased *i-Ready Inform*. This site answers frequently asked questions (FAQs) about the entire *i-Ready* product suite, and includes brief training videos, helpful reference sheets, and many other resources. The document Proctoring Guidance for the Diagnostic provides helpful recommendations about what to do before, during, and after test administration. The Get Good Data page provides checklists and other tools to ensure successful administrations that yield reliable data about student performance. These documents can be accessed using the embedded links on the Get Good Data page or by entering the title or keywords into the search bar on *i-Ready Central*.

3.4 Testing Windows and Administration Time

3.4.1 Testing Windows

The *i-Ready Inform* assessment is administered in three windows spanning the academic year: fall, winter, and spring. The dates for each window are presented in Table 3.1. Note that the windows changed slightly beginning in the 2020–2021 school year to better reflect district preferences and usage patterns. It is recommended that school districts set testing dates that lie within these three windows according to their unique scheduling needs, but testing is available year round. Note that the percentile rank associated with an *i-Ready Inform* scale score is computed within each window, and thus may vary across these testing windows (see Chapter 8 for more details).

Table 3.1. Recommended *i-Ready Inform* Administration Windows

School Years	Fall	Winter	Spring
2020–2021 School Year and Beyond	August 1 to November 15	November 16 to March 1	March 2 to June 15
Prior to the 2020–2021 School Year	August 1 to November 30	December 1 to March 15	March 16 to June 15

3.4.2 Administration Time

While *i-Ready Inform* is not a timed assessment, school districts are advised to plan testing sessions based on the recommended total testing times for each grade level, which are displayed in Table 3.2. Proctors are encouraged to allot additional time for transitions, logging in, viewing tutorial videos, and accommodating students who may be absent or require more time to complete *i-Ready Inform* before it expires. The testing times and sessions are suggestions for planning purposes, and we recognize that students may need more or less time depending on a variety of factors. Note that Curriculum Associates recommends a four- to six-week delay in the start of testing in Grade K so that students can get acquainted with the classroom experience and interacting with technology. Resources are provided to support educators in preparing their students accordingly.

Table 3.2. *i-Ready Inform* Scheduling Recommendations by Grade

Grade	When to Schedule the First *i-Ready Inform*	Recommended Testing Session Times*
K	Four to six weeks into the school year	Three 20-minute sessions
1	Start as soon as possible.	Two 20- to 30-minute sessions
2–5	Start as soon as possible.	Two 40- to 50-minute sessions
6–8 and 9–12	Start as soon as possible.	Plan number and length of testing session based on your schedule. Most students complete within 60–90 minutes (e.g., two 45-minute sessions); some will need additional time.

Note: Recommended testing session times do not necessarily reflect actual testing times, rather they are provided to help support scheduling and administration.

3.5 Accessibility and Accommodations

To meet the needs of the students and districts we serve, we engage in ongoing work to evaluate and improve our educational tools and resources. Consistent with our Accessibility and Inclusion Framework described in Chapter 2, we have developed a systematic approach to accessibility that includes:

- WCAG (Worldwide Web Consortium, 2022) AA compliance and the UDL framework guiding our accessibility efforts
- An internal team of access and equity, curriculum, assessment, policy, and research experts who are dedicated to finding new ways for our educational tools and resources to be used by a diverse range of learners
- A rigorous review process that involves outside accessibility experts to ensure our thinking and approach reflect established and evolving best practices
- Guidance and feedback from the school districts and educators we serve

In addition to the above, we have enlisted the help of student testers to evaluate the effectiveness of our accessibility enhancements. Our commitment to and work toward increased accessibility is ongoing. All our accessibility efforts and enhancements are offered to educators as part of Curriculum Associates' Software-as-a Service (SaaS) model.

i-Ready Inform incorporates several levels of accessibility features that are available to students. The first level of features is referred to as *universal supports,* which are available to all students regardless of disability status or any other type of need. That is, they are designed to make *i-Ready Inform* a broadly inclusive assessment. The next level of accessibility features is referred to as *designated supports,* which are not automatically available, but they can be made available based on the determination of an educator familiar with a student's characteristics and needs. The last level of accessibility features are *accommodations*, which are granted in accordance with a student's individualized education plan (IEP), 504, or EL plan.

Some of the accessibility features can be delivered directly within the *i-Ready* platform. Such features are referred to as *embedded*. Other available accessibility features cannot easily be provided directly within the *i-Ready* platform, so they are provided to students externally. These features are referred to as *non-embedded*.

Curriculum Associates offers a publicly available document[7] that provides additional details regarding the accessibility features of the program. This document is updated regularly to reflect the most current accessibility features. See Table 3.3 for an overview of the various types of supports. Please note, however, that not all supports are available for all grades, subjects, and domains.

Table 3.3. Types of Accessibility Features for the *i-Ready Inform* and Growth Monitoring Assessments

	Universal Supports	Designated Supports	Accommodations
Embedded	• Audio support	Audio support is available for *i-Ready Inform* for Mathematics	Audio support is available for *i-Ready Inform* for Mathematics

[7]Available at CurriculumAssociates.com/Reviews/iReadyAccessibility.

	• Audio description • Keyboard access • Calculator • Color contrast • Closed captioning • Digital Math Tools • Presentation of material for age-appropriate pedagogy and legibility	Grades 6+ items. As of July 2024, audio support will be available for *i-Ready Inform* for Reading items where universal audio support is not available. In both Reading and Mathematics, audio support is available as either a designated support or an accommodation.10 Educators need to enable this feature. To learn more about audio support in *i-Ready Inform*, please refer to the Feature Overview: Audio Support.	Grades 6+ items. As of July 2024, audio support will be available for *i-Ready Inform* for Reading items where universal audio support is not available. In both Reading and Mathematics, audio support is available as either a designated support or an accommodation.10 Educators need to enable this feature. To learn more about audio support in *i-Ready Inform*, please refer to the Feature Overview: Audio Support. Educators have the ability to manually exempt students who are blind, low vision, deaf, or hard of hearing from foundational reading domains, such as Phonological Awareness, Phonics, and High-Frequency Words, due to the domains' dependence on audio or visual content.*
Non-Embedded	• Audio amplification • English dictionary (when appropriate) • Noise buffer (e.g., earmuffs, audio aids) • Scratch paper (i.e., blank paper) • Thesaurus (when appropriate)	• Bilingual word-for-word dictionary • Magnification device • Native language translation of directions • Student reads test aloud.	• Abacus • Alternate response options • Calculator • Extended time breaks and flexible scheduling • Graphic organizer/reference • Sheet/checklist • Human reader • Human signer • Multiplication table • Scribe • Screen readers • Tactile graphics

Note: Not all supports are available for all grades, subjects, or domains. Please see Accessibility and Accommodations Update as well as other materials on the Curriculum Associates accessibility hub.

3.6 Test Security and Privacy

By implementing comprehensive monitoring procedures, prioritizing platform security, safeguarding data, and establishing effective policies and oversight, the test administration process maintains integrity, ensures data privacy, and upholds the trust and confidentiality of stakeholders involved in the assessment process.

3.6.1 Supervision

Ensuring the integrity and security of test administration is crucial to maintaining the validity and reliability of assessment results. To support this, Curriculum Associates provides training and detailed recommendations

for scheduling, student engagement, and monitoring throughout the test administration process. School leaders, teachers, or other designated proctors are assigned to supervise the testing environment, closely monitoring test-takers to prevent any irregularities or violations and monitor status reporting to ensure timely completion. Proctors serve to enforce adherence to test administration protocols and offer test-takers encouragement or breaks when necessary, ensuring a fair and controlled testing environment.

3.6.2 Secure Browser

Curriculum Associates offers an optional secure browser solution that prevents students from printing, copying, accessing other websites, or using other applications while taking the *i-Ready Inform* and/or Growth Monitoring assessments. The feature is designed for educators wanting increased security during the administration. Implementing secure testing will:

- Lock the testing environment when students begin the assessment
- Give educators confidence that students are on task by minimizing distractions
- Ensure the accuracy and reliability of assessment data

The secure browser can be enabled for an entire district, or for specific schools. Once the secure browser is enabled, all students accessing *i-Ready* from the enabled district or school must use the secure browser when taking *i-Ready Inform* or Growth Monitoring assessments.

3.6.3 Data Transmission and Storage

We conduct data transmission for *i-Ready* using a secure internet protocol (e.g., Secure Sockets Layer [SSL] and Hypertext Transfer Protocol Secure [https]). As a web-based, SaaS application*, i-Ready* is accessible only through https. In cases where we need to transfer data with a district student information system, we employ a Secure File Transfer Protocol (SFTP) server through which data is encrypted. *i-Ready* employs the following types of encryption:

- *Database encryption*: All *i-Ready* database servers are encrypted using industry standard AES-256 encryption.
- *Network encryption*: *i-Ready* is accessible only via https, and all public network traffic is encrypted with the latest encryption standards.
- *Data-at-rest/data-in-motion encryption*: Encryption of data at rest is implemented for all data stored in the *i-Ready* system. All data transfers are conducted using a secure internet protocol (e.g., SSL and https).
- *Backup encryption:* As indicated above, all *i-Ready* database servers are encrypted using industry-standard AES-256 encryption.

i-Ready provides various layers of security (network, server, and application security), employing the concept of "defense in depths" throughout the architecture and operational procedures. Some examples of this include:

- Only SSL (i.e., secure file transfer protocol implementation) ports are available to the internet.
- Security best practices from our cloud hosting provider, Amazon Web Services, are implemented.

- A virtual private network (VPN) is required for all administrative connections to the architecture.
- Multifactor authentication is required for administrative access.
- Data are encrypted at rest.
- After initial http connection, all user communication is forced to https.
- A small operations team is the only authorized personnel with administrative access to production systems.
- Anti-virus and intrusion detection systems are utilized.
- Security logs are kept and analyzed.
- Strict server and application configuration management is enforced.
- Third-party security audits are conducted annually.

Curriculum Associates hosts all our servers at Amazon® Web Services; all data is stored within the United States. AWS implements network-level security measures in accordance with industry standards that are layered on top of our network controls such as security groups, firewalls, and virtual private networking. Continuous maintenance plans are in place to keep all servers optimized. Online servers are connected to the Internet by one of the most redundant networks in the industry. AWS is certified across multiple industries (e.g., health, government, and finance) as one of the highest-level security and availability service providers in the country. Our data hosting facility includes 24x7x365 on-site security and does not allow any physical access to the server area by non-authorized personnel or outside visitors at any time.

Amazon® Web Services and all related marks, including logos, graphic designs, and service names, are trademarks or trade dress of AWS in the US and other countries.

3.6.4 Data Security

Security is an ongoing focus for Curriculum Associates, and as such we regularly plan security initiatives such as advanced staff training, additional security oversight on various systems, improved policies, and access control. *i-Ready* adheres to an extremely high degree of data security in its operations and data management. We follow specific procedures to maintain secure and reliable data for our customers and protect against unauthorized access and use. Our Data Handling and Privacy Statement—posted on our website at CurriculumAssociates.com/Support/Privacy—informs our customers about our current data security policies and practices, which are intended to safeguard the sensitive information in *i-Ready*.

i-Ready tracks each user's last access time to the system, as well as when users perform certain activities, such as taking a test. The program tracks all failed user logins and changes to access rights in the system. Administrators, teachers, and students are timed out after 30 minutes of inactivity.

On the server side, we log any server errors that are detected, along with detailed user information. This allows us to troubleshoot a potential problem, should one be detected. Our IT operations team keeps a constant vigil—through automated and human methods—to ensure that our customers' data is kept secure and confidential.

We adhere to a comprehensive Security Policy and Implementation guide that outlines our overall approach to data security. Based on a variety of reference documents, the guide primarily follows the ISO-27002 standard for Information Technology–Security Techniques–Code of Practice for Information Security Controls.

3.6.5 Data Privacy

Curriculum Associates adheres to strict standards for student data privacy and accessibility, including the FERPA and the Children's Online Privacy Protection Act (COPPA). We provide FERPA training for our staff and maintain a detailed security plan. An overview of that plan is available at CurriculumAssociates.com/Support/Privacy-and-Policies/i-Ready-Data-Handling-Privacy.

Our data integration process is FERPA compliant and never shares student data with unauthorized outside parties. Additionally, Curriculum Associates has joined the Future of Privacy Forum and the Software and Information Industry Association Student Privacy Pledge to safeguard student privacy. The pledge outlines commitments regarding collection, maintenance, and use of student personal information.

While authorized teachers and school administrators can access student information and related *i-Ready* usage data, this information is not made available to other students or to the public. We do not solicit personally identifiable information directly from students—all student information is provided by school districts or created through the use of the *i-Ready* platform.

3.6.6 Use of Student Data

Curriculum Associates provides reporting capabilities to the educators who use *i-Ready*, and these reports are generated based on collected *i-Ready* usage information. This usage data may also be used on an aggregated basis to allow our partner success management, customer service, and technical support teams to provide services that more specifically meet the needs of our educator customers. We use student data collected from the use of *i-Ready* for the purpose of making the programs available to our customers and improving content and effectiveness.

Curriculum Associates collects and uses aggregated, de-identified student data for core product functionality to make *i-Ready* a more effective, adaptive product. "De-identified student data" refers to data generated from usage of *i-Ready* from which all personally identifiable information has been removed or obscured so that it does not identify individual students and there is no reasonable basis to believe that the information can be used to identify individual students. We also use de-identified student data for research and development purposes, including analyzing the efficacy of *i-Ready*.

While some of this work is done internally, Curriculum Associates does share de-identified data with trusted third-party research partners as part of these research initiatives. We do not re-identify de-identified student data, and we protect against the re-identification of de-identified student data. We also prohibit our research partners from attempting to re-identify de-identified student data. Curriculum Associates does not sell student data or otherwise share student identifiable data with third parties. We do not include advertisements within *i-Ready* nor does it use student data for targeted advertising in any manner.

Chapter 4: Computerized Adaptive Testing

4.1 Chapter Summary

The *i-Ready Inform* and Growth Monitoring assessments are computerized adaptive tests (CATs) whose item selection algorithm aims to maximize measurement precision while ensuring that students are assessed on the appropriate content. This chapter includes an overview of the core components of our CAT: the vertical scale, our approach to content balancing—referred to as our *test flows*—across domains within each grade and subject, item selection and stopping rules, methods for estimating student proficiency, and the item bank. We conclude the chapter with a summary of simulation research that Curriculum Associates conducts to refine processes and improve the effectiveness of *i-Ready Inform*.

4.2 Introduction

Computerized adaptive testing is an assessment administration method defined by its ability to tailor the items administered to each student based on their responses to previous items. The *i-Ready Inform* and Growth Monitoring assessments leverage CAT in conjunction with grade- and, in some cases, proficiency-specific test flows, which specify the number of items administered from each domain. As such, our assessments provide a strong degree of measurement precision across a wide range of the proficiency spectrum, which helps ensure that resources provided to teachers are appropriately aligned with students' needs.

4.2.1 Vertical Scale

The *i-Ready Inform* and Growth Monitoring assessments report student proficiency on a vertical scale with separate scales for Reading and Mathematics. That is, regardless of a student's chronological grade level and the items administered, scores are reported on the same 100–800 score scale. Critical components of the vertical scale, such as field testing, item calibration methodology, scale transformation constants, item drift monitoring, and scale maintenance, are discussed at length in Chapter 5.

4.2.2 Computerized Adaptive Testing

CAT programs must achieve a balance between item selection methods that support measurement precision while meeting specific content requirements and managing item exposure. *i-Ready Inform* exemplifies a strategic balance between content representation and measurement efficiency. This balance is achieved by structuring the test to present questions from all major content domain areas without requiring the administration of items in skills that are likely incongruent with the student's proficiency level, thus avoiding unnecessary time expenditure. By taking advantage of CAT, *i-Ready Inform* utilizes items that are more closely aligned with students' proficiency levels, which may also result in a shorter test than would be possible with a fixed-form (i.e., non-adaptive) assessment.

The *i-Ready Inform* and Growth Monitoring assessments are initialized using a starting proficiency estimate. The methods for selecting the starting items are discussed later in this chapter. After the administration of the first item or testlet, the item selection algorithm compiles a pool of eligible items for that student and selects an item appropriate for a student with that proficiency level. As the student progresses through the

assessment, the system adapts in real time, selecting items based on the student's performance up to that point. The system continues this process, progressively narrowing in on the student's proficiency level, until the stopping rule is satisfied. Upon completion of *i-Ready Inform*, a variety of reports are generated, and an online lesson queue is generated for the Personalized Instruction program based on performance in each domain, thereby providing a course of action to facilitate academic growth.

Ideally, the items selected for students during an adaptive assessment should be neither too easy nor too difficult, but rather appropriately challenging for that student. These types of items maximize measurement precision for that student. In general, students who answer items correctly are subsequently administered more-difficult items, while students who answer items incorrectly are subsequently administered less-difficult items until the algorithm identifies each student's proficiency level.

The five major components of our CAT are:

- Content balancing as accomplished through test flows
- Adaptive item selection logic
- Proficiency estimation
- Operational item bank
- Test termination criteria

Each of these are discussed in detail in the following sections.

4.3 Test Flows

A *test flow* details the number of and order in which students are administered items from each domain. Both the *i-Ready Inform* and Growth Monitoring assessments have pre-defined test flows for each subject and grade. Note that the domains assessed can vary by grade level [8] as well as interim proficiency estimates during the assessment. The test flows remain the same for every *i-Ready Inform* assessment a student completes during a given grade. Results from *i-Ready Inform* assessments administered previously within that grade do not impact the test flow on *i-Ready Inform* assessments taken for the remainder of the grade.

4.3.1 Reading Test Flows

The design of *i-Ready Inform* for Reading promotes accurate assessment of a student's reading proficiency. All students, regardless of grade level, are presented with items from the Vocabulary, Comprehension: Literature, and Comprehension: Informational Text domains. Students in lower chronological grade levels or performing at lower proficiency levels may also be assigned items from the Phonics, Phonological Awareness, and High-Frequency Words domains. Reading test flows route students into these foundational skill domains if it is

[8] If a student has a developmental level (i.e., a grade level that is different than the chronological grade level that is set manually by the teacher or administrator) set, the test flow is based on the developmental grade level, but the reporting is based on the chronological grade level.

determined by the interim proficiency estimate that these additional domains are appropriate to measure or if the school or district participates in the Foundational Reading Extension (see Section 1.3.2).

The reading test flows are shown in Figure 4.1, Figure 4.2, Figure 4.3, and Figure 4.4. For example, Figure 4.2 illustrates the Grade 2 test flow for Reading as follows: students are first administered Phonics items, then Phonological Awareness items if the student has an estimated proficiency of less than 421 (and thus would benefit from being assessed in that domain), followed by (for all students) items from the Vocabulary, Comprehension: Literature, Comprehension: Informational Text, and High-Frequency Words domains.

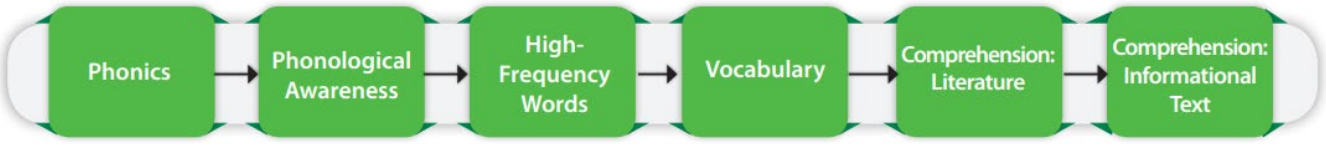


Figure 4.1. Reading Test Flow—Grades K and 1

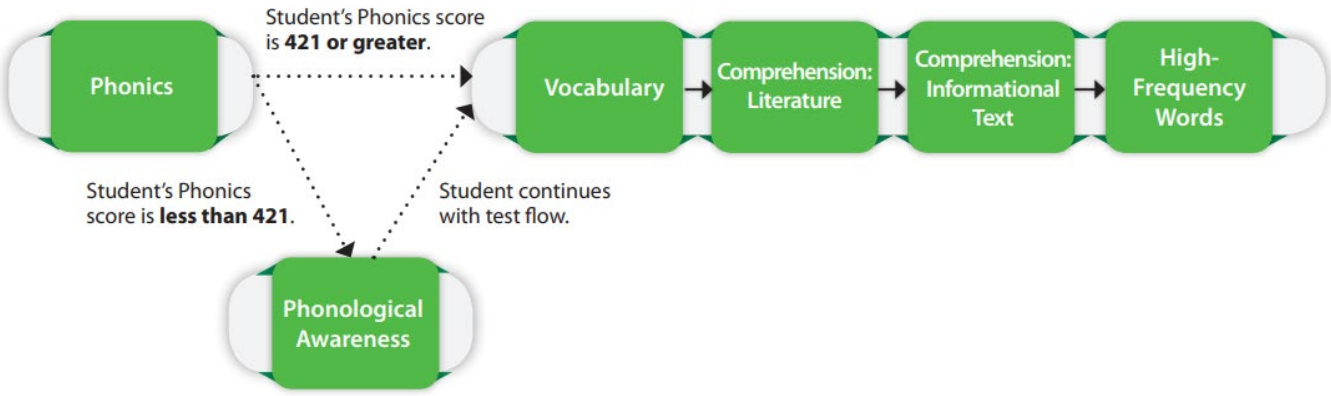


Figure 4.2. Reading Test Flow—Grade 2

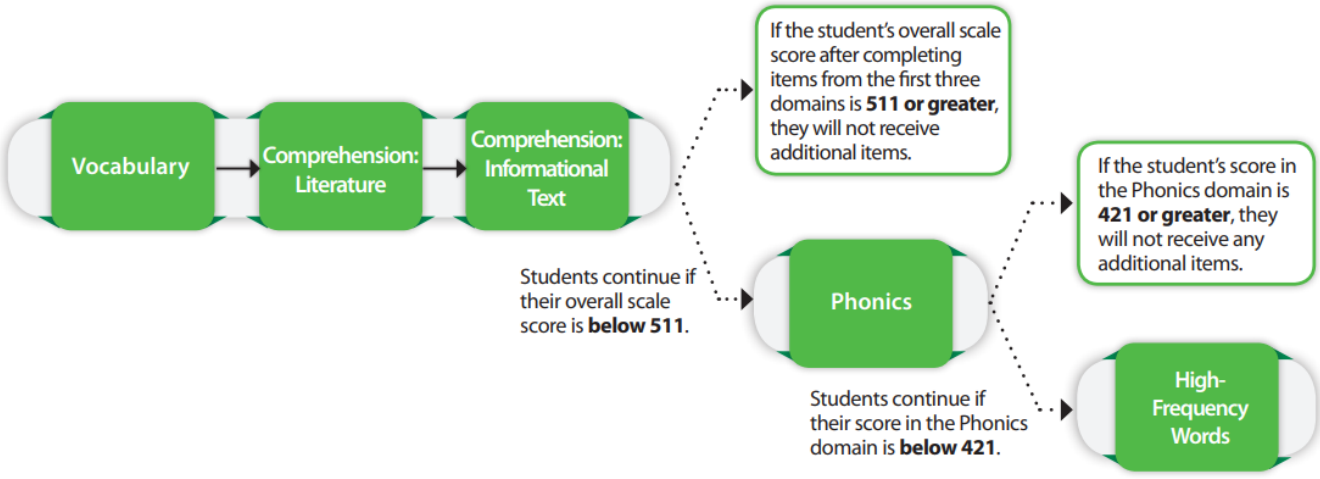


Figure 4.3 Reading Test Flow—Grades 3–8

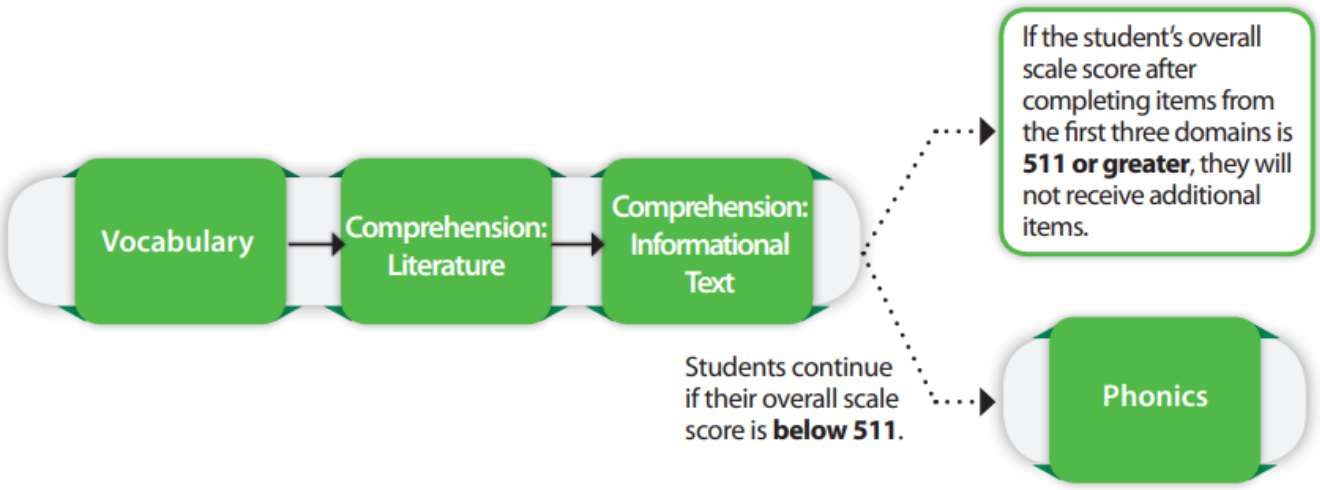


Figure 4.4. Reading Test Flow—Grades 9–12

Table 4.1. Reading Items by Grade and Test Flow

Grades K–1

Phonics	12
Phonological Awareness	12
High-Frequency Words	12
Vocabulary	12
Comprehension: Literature	12
Comprehension: Informational Text	12
Total Operational Items	72

Grade 2

	Phonics Score < 421	**Phonics Score ≥ 421**
Phonics	12	12
Phonological Awareness	12	—
High-Frequency Words	12	12
Vocabulary	12	12
Comprehension: Literature	12	12
Comprehension: Informational Text	12	12
Total Operational Items	72	60

Grades 3–8

	Overall Score < 511 and Phonics Score < 421
Phonics	12
Phonological Awareness	—
High-Frequency Words	12

Vocabulary	18
Comprehension: Literature	18
Comprehension: Informational Text	18
Total Operational Items	78

Grades 9–12

	Phonics Score < 421	Phonics Score ≥ 421
Phonics	12	—
Phonological Awareness	—	—
High-Frequency Words	—	—
Vocabulary	18	18
Comprehension: Literature	18	18
Comprehension: Informational Text	18	18
Total Operational Items	66	54

4.3.2 Mathematics Test Flows

The design of *i-Ready Inform* for Mathematics also promotes accurate assessment of a student's proficiency in each domain. For Mathematics, the domains assessed include Algebra and Algebraic Thinking, Number and Operations, Geometry, and Measurement and Data. All students in Grades K–8 receive a total of 66 items in the same sequence of domains, each with adaptive item selection based on continuously updated estimates of student proficiency. However, high school students whose proficiency is less than 480 following the Algebra and Algebraic Thinking domain will receive a total of 66 items from domains that are ordered the same as for Grades K–8 students. High school students whose proficiency is 480 or higher following the Algebra and Algebraic Thinking domain will encounter a unique sequence of domains and a total of 72 items.

Figure 4.5 and Figure 4.6 illustrate the test flows for Mathematics for Grades K–8 and 9–12, respectively.

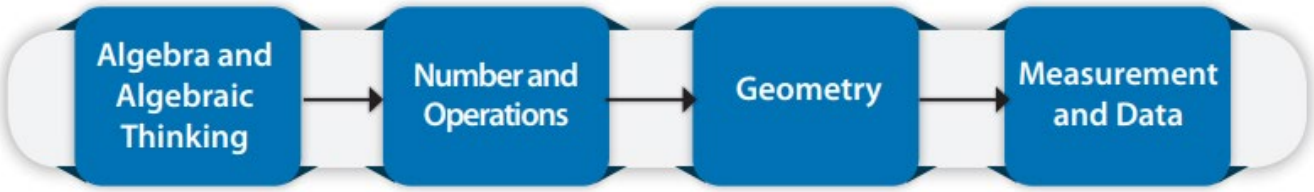


Figure 4.5. Mathematics Test Flow—Grades K–8

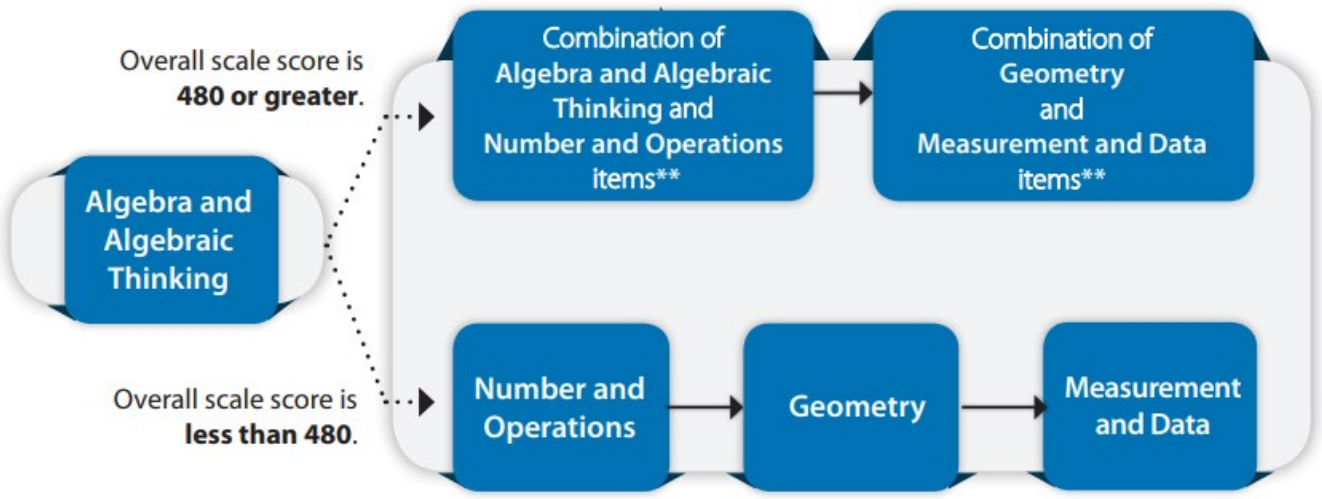


Figure 4.6. Mathematics Test Flow—Grades 9–12

Table 4.2 Mathematics Items by Grade and Test Flow

Grades K–8

Algebra and Algebraic Thinking	18
Number and Operations	20
Geometry	14
Measurement and Data	14
Total Operational Items	66

Grades 9–12

	Overall Score < 480	**Overall Score ≥ 480**
Algebra and Algebraic Thinking	18	18–36
Number and Operations	20	0–18
Geometry	14	36
Measurement and Data	14	36
Total Operational Items	66	72

4.3.3 Growth Monitoring Test Flows

Growth Monitoring assessments produce an end-of-year (EOY) projected score, which is based on all *i-Ready Inform* and Growth Monitoring assessments taken within a given school year. The test flows are described here. The details of the projection model are discussed in Chapter 5, and an example score report is shown in Chapter 11.

4.3.3.1 Reading

Students may receive 19–21 items from the Comprehension: Informational Text, Comprehension: Literature, Phonics, and Vocabulary domains.

- For Grades K–2, students receive the following:
  - Four Comprehension: Informational Text items
  - Four Comprehension: Literature items
  - 12 items from either the Phonics or Vocabulary domains, with no fewer than four items administered in each domain
- For Grades 3–8, students receive the following:
  - Nine to 11 items from either Comprehension: Informational Text or Comprehension: Literature, with no more than six items administered from either domain
    - If another testlet must be selected for administration after the ninth item from the Comprehension domains, the comprehension section will terminate instead of selecting and administering another testlet.
      Otherwise, this portion will end on item 11.
  - 10 items from either the Phonics or Vocabulary domains, where the number of Phonics items can range from zero to five

4.3.3.2 Mathematics

Students are administered 20 Mathematics items. Unlike *i-Ready Inform*, items on the Growth Monitoring for Mathematics assessment are not administered in specific domain sets, but rather in one multi-domain set of 20 items. In this multi-domain set, there will be six Number and Operations items, six Algebra and Algebraic Thinking items, four Measurement and Data items, and four Geometry items. These items may appear in any order regardless of domain.

4.3.4 Grade-Level Content Restrictions

4.3.4.1 Typical Grade-Level Caps

The vertically scaled item bank allows for the administration of both on-grade level items, which match the student's chronological grade, and certain off-grade level items, as necessary, to best measure each student's proficiency. While off-grade level items support measurement precision across the proficiency spectrum, it is also important that students receive items that measure the appropriate content. As such, we impose a maximum grade level for each chronological grade and domain, which is referred to as a *typical grade-level cap*. Note that there are exceptions to these caps, which are discussed in the next section.

For Reading, the typical grade-level cap is set at three grade levels above the student's chronological grade, except for Grade 6, where the typical grade-level cap is set to two grade levels above (i.e., Grade 8) to ensure that sixth grade students are not administered high school content. For example, a fourth grade student administered a Reading assessment could be served items from the bank representing Grades K–7. For Mathematics, the typical grade-level cap is set at one grade level above the student's chronological grade to ensure the appropriateness of content. Thus, a fourth grade student administered a Mathematics assessment could be served items from Grades K–5.

4.3.4.2 Grade-Level Cap Exceptions

Other exceptions to the typical grade-level cap rules are intentionally built into the *i-Ready Inform* item selection algorithm. The purpose of grade-level cap exceptions is to increase the range of allowable items for students performing at a level exceeding the difficulty range of the item pool within the grade-level caps. These exceptions provide appropriately challenging items for high-achieving students, which support reliable measurement outside the range of achievement typical of students at a certain grade level. Note that these items are also subject to a separate set of caps by grade level and domain.

The items selected as grade-level cap exceptions are reviewed by the assessment editorial team to ensure content appropriateness for the respective grades. Additionally, these items must have difficulty levels that are appropriately high. Students are only administered items from grade levels above the traditional grade-level caps if their interim estimated proficiency level is high in relation to the difficulty of the item pool. Items in the grade-level cap exception pool are administered on an item-by-item basis for Mathematics and most Reading domains.

4.4 Adaptive Item Selection

4.4.1 Starting Items

Selection of the first item depends on whether the student has previously taken an *i-Ready Inform* or a Growth Monitoring assessment. For students without a prior score, an item is randomly selected from a predetermined set of items with difficulty closest to a specified starting location, which is one placement level below the student's grade. The list of items for each grade has been reviewed by the content team to ensure appropriateness of the items as starting items for the student. The first item is always a standalone item, never a testlet. These items are also eligible to be administered during the remainder of the assessment if they were not used as the starting item. For students who have previously taken an *i-Ready Inform* or a Growth Monitoring assessment, their starting item will follow the typical item selection rules outlined in the next section.

4.4.2 Item and Testlet Selection

Under the Rasch (1960) model, measurement information (i.e., test score reliability) is maximized when item difficulty is equal to student proficiency, which can be shown using Equations 4.2 and 4.3. As the item difficulty and student proficiency diverge, information for that student decreases. Consequently, items that are too easy or too difficult for a student will provide little information about the proficiency level of that student because the probability of a correct or incorrect response will be close to one and zero, respectively. Rather, selecting an item that minimizes the difference between item difficulty and student proficiency, a method known as *b-matching*, optimizes the precision of proficiency estimation at that point in the test.

Prior to the b-matching procedure, a subset of eligible items for a student is identified. That subset depends on the following:

- The student's grade level
- The domain from which the next item will be selected, as defined by the test flow

- Grade-level caps and grade-level cap exceptions
- Whether the test is an *i-Ready Inform* or Growth Monitoring assessment
- Which items the student had been administered previously

Note that items and testlets are not repeated within an administration, and repetition across testing occasions are minimized.

Next, the b-matching procedure is carried out by identifying one or more items with difficulty levels close to the student's proficiency estimate at that point in the test is obtained. If multiple items fit this criterion, one of those items is randomly selected for administration. In Mathematics and in most Reading domains, all items are selected item by item using this procedure. For the Reading domains Comprehension: Literature and Comprehension: Informational Text, whole testlets containing multiple items are selected using the same b-matching procedure but based on the testlet difficulty, which is defined as the average difficulty of the associated items. If a testlet near the end of the test is selected in which the number of associated items is greater than the number of items remaining in that domain, the algorithm stops administering items from that testlet once the prescribed number of items from that domain has been administered.

4.4.3 Stopping Rule

The test terminates when a student has completed the number of required operational items according to their test flow. Table 4.3 below shows the possible operational item counts—depending on the test flow—by grade and subject. Note that the total test length could be slightly longer if field-test items are included (see Chapter 5 for details).

Table 4.3. Number of Operational Items by Grade and Subject

Grade	Reading	Mathematics
K–1	72	66
2	60 or 72	66
3–8	54, 66, or 78	66
9–12	54 or 66	66 or 72

4.5 Proficiency Estimation

4.5.1 Starting Proficiency

The starting student proficiency depends on whether the student has a prior *i-Ready Inform* or Growth Monitoring score stored in the assessment system. If the student has a prior score, the starting proficiency estimate is their final overall proficiency estimate from their most recent *i-Ready Inform* or Growth Monitoring assessment. Note that, while Growth Monitoring assessments are used to calculate an EOY projected score, it is possible to calculate a score on the same scale as *i-Ready Inform* because it draws from the same item bank.

If a student has not previously been administered an *i-Ready Inform* or Growth Monitoring assessment for the subject in which they are testing, their starting proficiency estimate is set to the cut score for students from

their grade who place one grade level below at the overall assessment level. For example, a Grade 4 student in this scenario would have their starting proficiency set to a scale score of 434, which is the lower bound of a Grade 3 placement level for a Grade 4 student in mathematics. Starting values are also evaluated to ensure there are an adequate number of items with difficulty levels near the starting value on the Rasch logit scale.

4.5.2 Proficiency Estimation

4.5.2.1 Overall Estimates

Student proficiency estimates, which are based on the Rasch (1960) model, are updated after each item response and are used in subsequent item selection. To control the magnitude of adjustments to the proficiency estimate, particularly near the beginning of the assessment, the maximum allowable adjustment after the administration of any single item is limited to ± .5. Generally, the magnitude of the adjustments to a student proficiency estimate decreases as one progresses through the test because the estimation algorithm progressively narrows in on a student's proficiency.

4.5.2.2 Domain Estimates

The procedure for domain-level proficiency estimation is similar to overall score estimation but with the estimate being calculated using only the items within the domain according to the test flow for that grade and subject. The starting domain level student proficiency estimate for all domains is equal to the overall student estimate when the first item in the domain is administered. Domain-level estimates can be used for routing purposes for certain test flows during the assessment and are derived after each item and at the end of the domain, just like estimation for the overall score.

4.5.2.3 Extreme Score Adjustment

In the case of extreme scores—either all correct or all incorrect item score patterns at the overall or domain level—a slight adjustment is made to the scoring algorithm to produce a finite estimate.

4.6 Item Bank

As of late 2022, the *i-Ready Inform* item bank comprises nearly 10,000 mathematics and reading items. The mathematics operational item bank contains more than 4,000 items, and the reading operational item bank contains nearly 6,000 items. The *i-Ready Inform* for Mathematics in Spanish uses items that have been transadapted from the English-language versions. The operational item bank continues to grow as new items are field tested and evaluated for inclusion in the bank. Most *i-Ready Inform* items are also eligible for use in the Growth Monitoring assessments. Items are not included when their properties don't align with the shorter format of the assessment.

Table 4.4 displays the number of items both assigned to each grade and eligible for administration—due to the grade-level caps and grade-level cap exceptions—in each grade by grade and subject. Table 4.5 and Table 4.6 show item counts by domain in Mathematics and Reading, respectively. Note that as of the date of publication, there are more than 10,000 items across both subjects.

Table 4.4. Number of Items by Grade and Subject

Mathematics

Grade	By Grade	Eligible
K	348	958
1	373	1,378
2	434	1,849
3	606	2,500
4	474	2,825
5	398	3,213
6	501	3,611
7	353	3,882
8	289	3,997
9	126	4,099
10	125	4,170
11	143	4.170
12		4.170

Reading

Grade	By Grade	Eligible
K	763	2,544
1	730	2,912
2	559	3,268
3	492	3,758
4	368	4,093
5	356	4,501
6	446	4,868
7	453	5,127
8	461	5,273
9	293	5,223
10	244	5,223
11	252	5.223
12	259	5.223

Table 4.5. Number of Items by Domain: Mathematics

Algebra and Algebraic Thinking	Geometry	Measurement and Data	Number and Operations	Total
1,192	662	844	1,472	4,170

Table 4.6. Number of Items by Domain: Reading

High-Frequency Words	Phonological Awareness	Phonics	Vocabulary	Comprehension: Literature	Comprehension: Information	Total
209	244	822	867	1,805	1,729	5,676

Note that the existing item bank does not contain mathematics items aligned to grade 12 because of the current item alignment approach, which reflects an integrated mathematics curriculum for high school. Under this approach, many items are appropriate for either grade 11 or grade 12, so they are aligned to grade 11. Starting in the 2026–2027 school year, Curriculum Associates will implement a PLD framework for item writing and alignment (discussed in Chapter 2), and this framework will support item alignments to grade 12 PLDs.

Test information functions (TIFs) are unique to each student and are based on the items administered, therefore they are not easily presentable here. However, information functions that are averaged across all eligible items for a grade level or grade band can be used to understand the nature of the item bank. Figure 4.7 and Figure 4.8 show the average item information curves at each grade level or band for Mathematics and Reading, respectively. These curves show high levels of information across a broad range of proficiencies. Additional information and results, including details on student-level measurement precision, are available in Chapter 6.

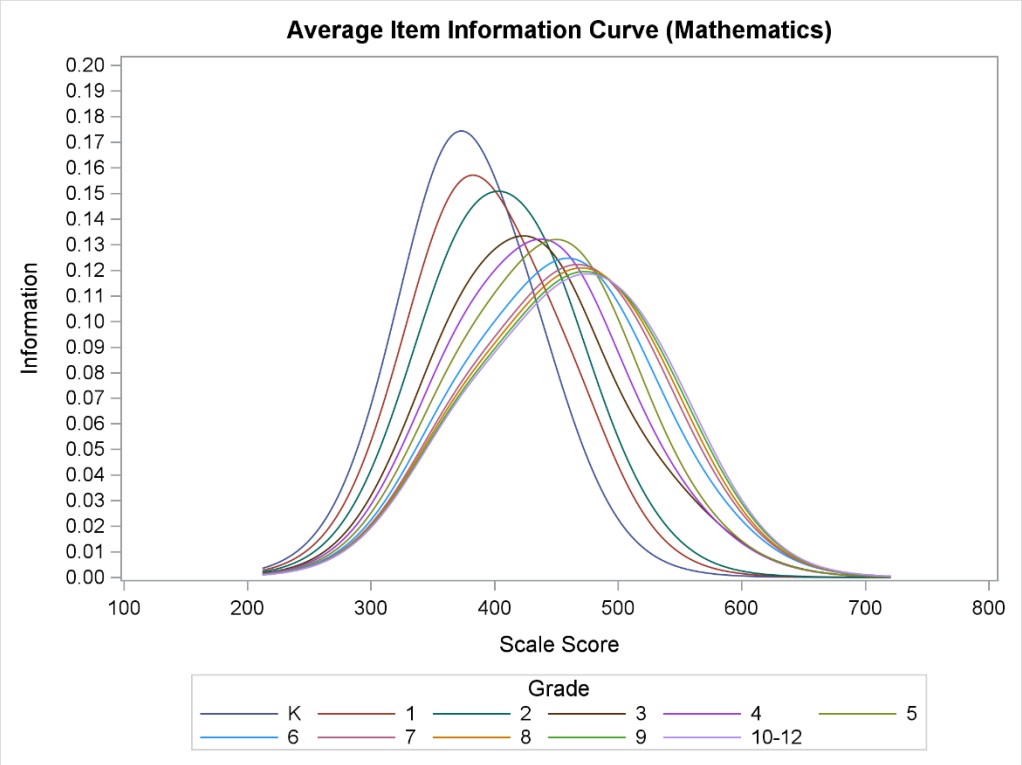


Figure 4.7. Average Item Information Curve by Grade for Mathematics

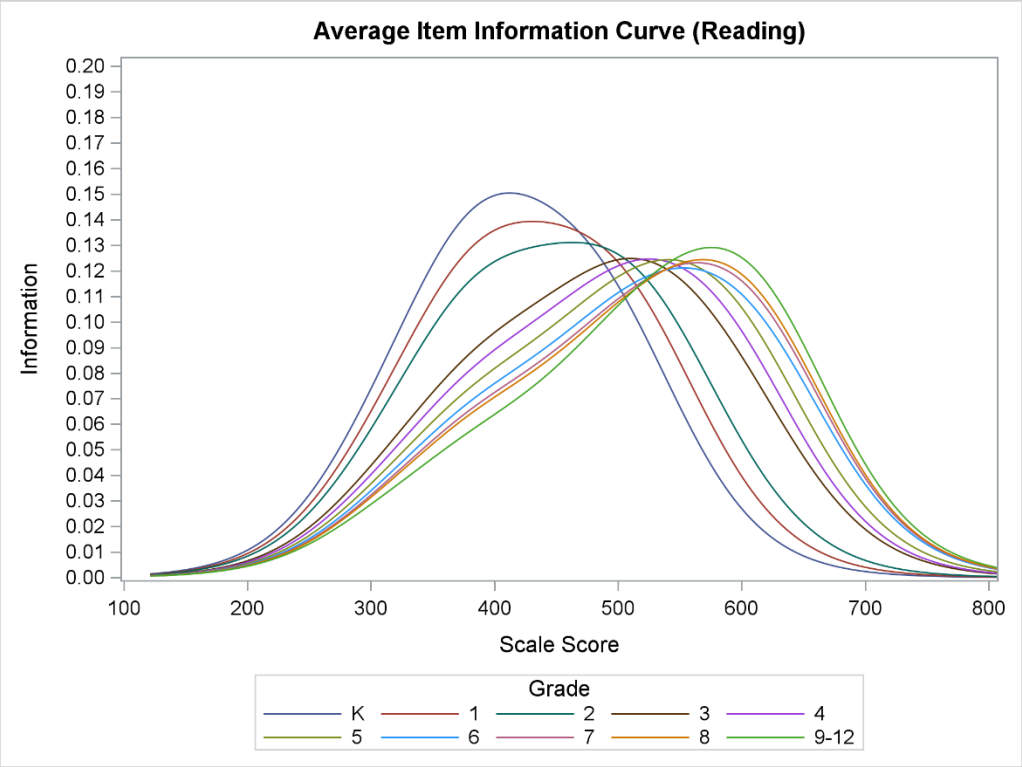


Figure 4.8. Average Item Information Curve by Grade for Reading

The distribution of items in the bank for *i-Ready Inform* is well suited for an adaptive assessment. Figure 4.9 and Figure 4.10 present the distributions of all Reading and Mathematics items as histograms in relation to an estimate of the Grades K–12 student population density. These figures demonstrate that the item and student proficiency distributions are well aligned. Additionally, these figures show that our item development procedures have produced a substantial number of difficult items, allowing for reliable measurement of highly proficient students. The objective of this intentional development was to ensure that enough items were available to measure student proficiency across three testing events. The same is not required in the same way in the lower extreme of the scale where students often experience large rapid growth.

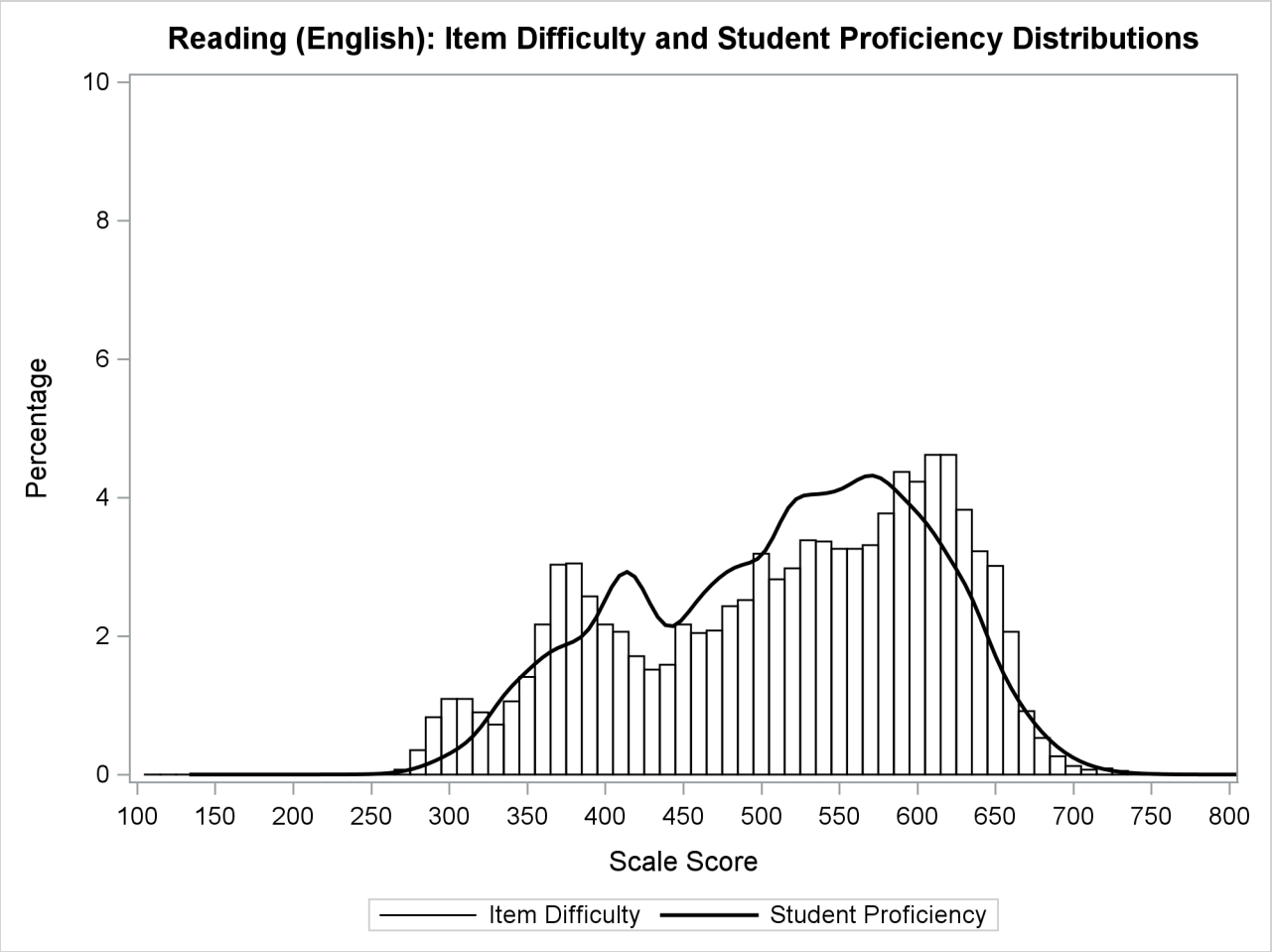


Figure 4.9. Item Difficulty and Student Proficiency Distributions for Reading

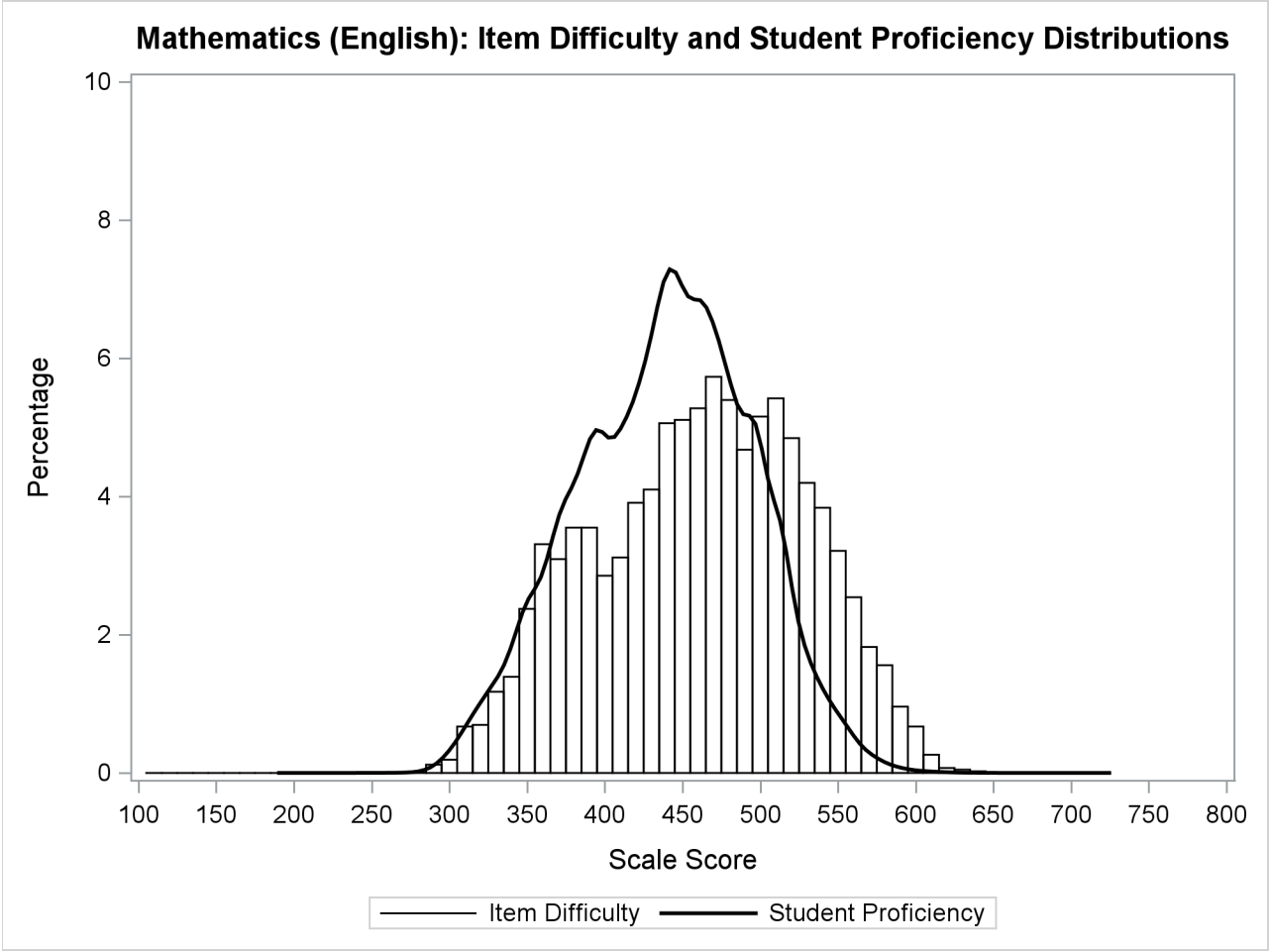


Figure 4.10. Item Difficulty and Student Proficiency Distributions for Mathematics

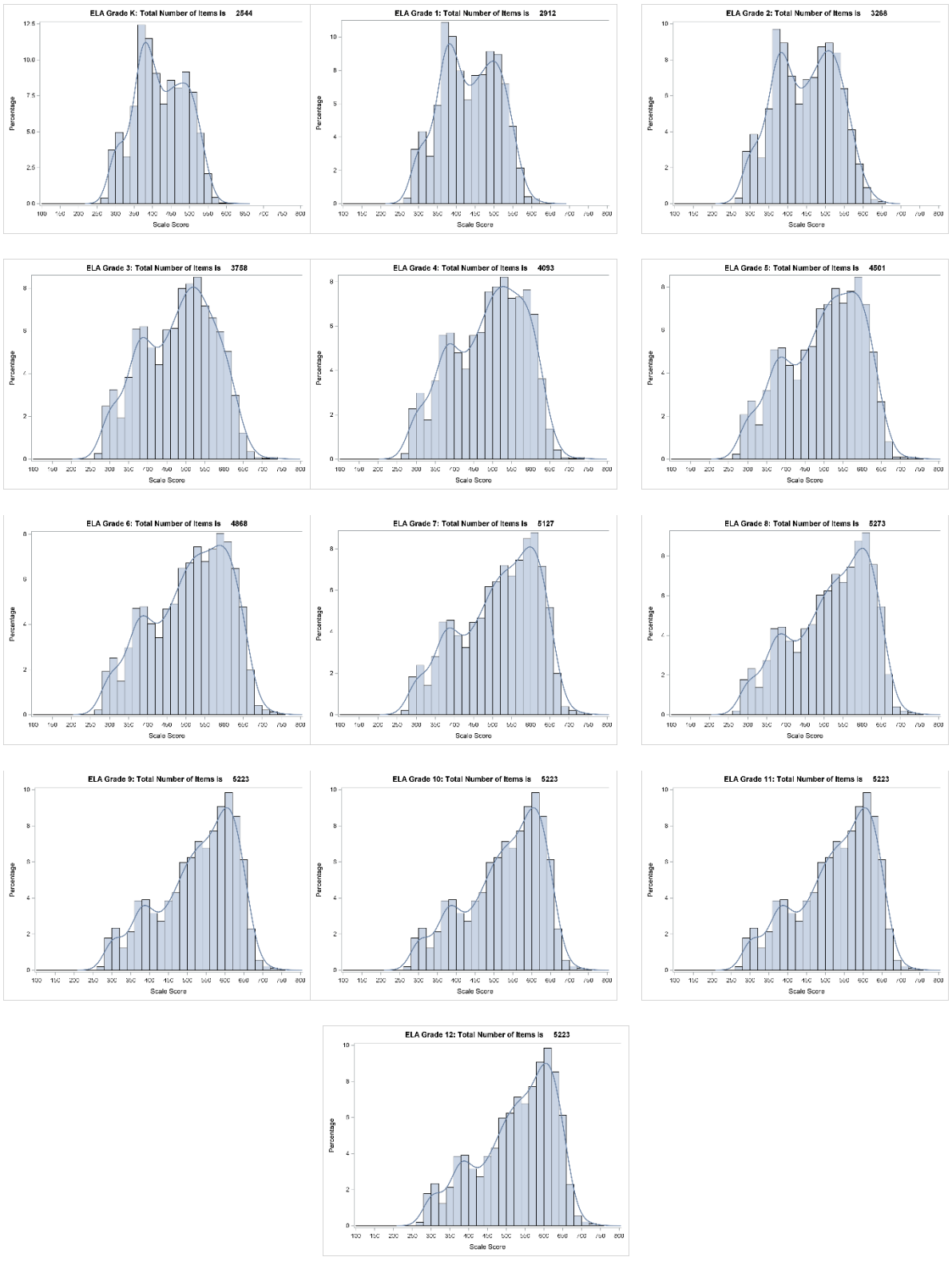


Figure 4.11. Eligible Items: Reading—Grades K–12

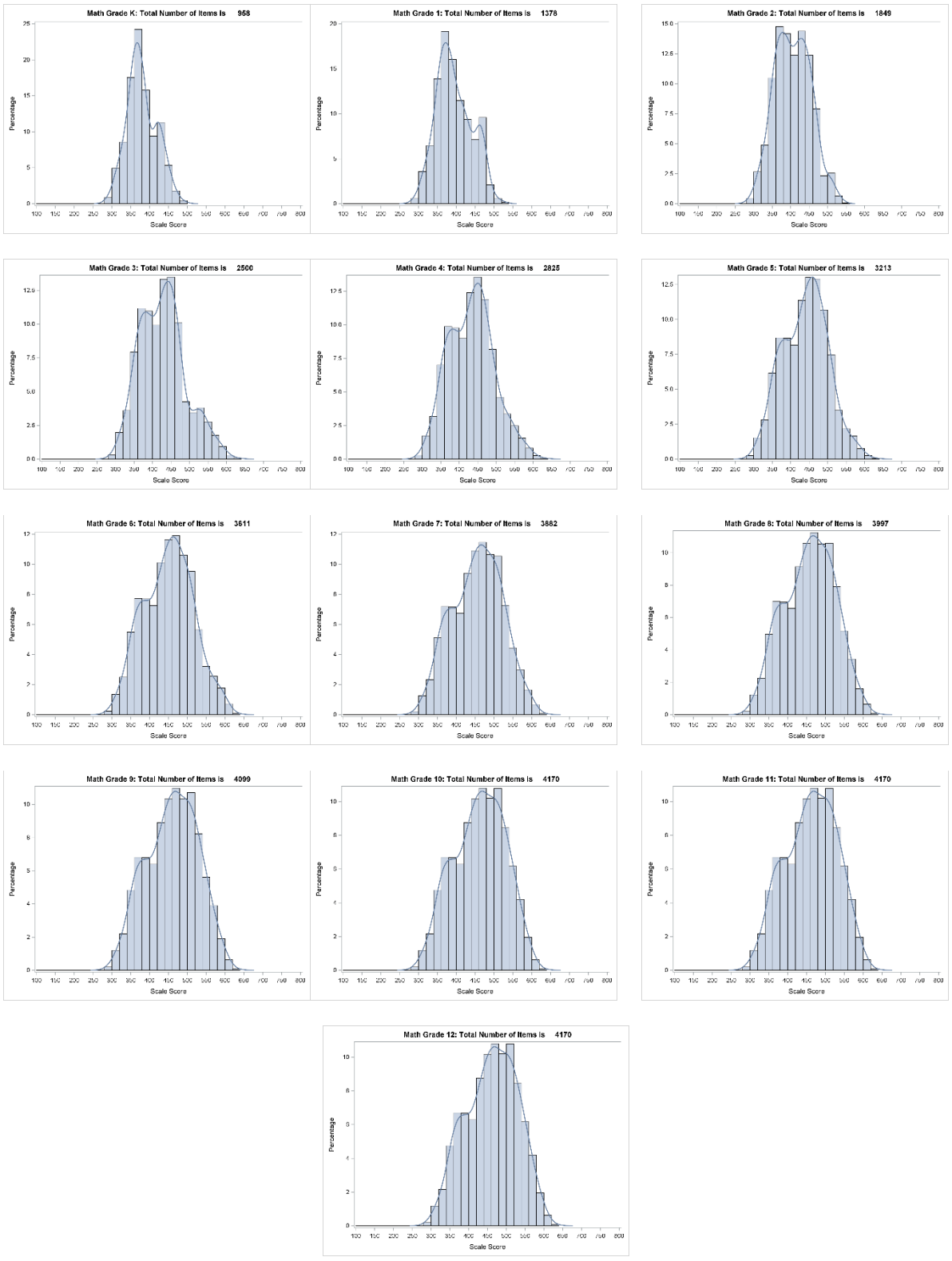


Figure 4.12. Eligible Items: Mathematics—Grades K–12

4.7 CAT Simulation Studies

Curriculum Associates utilizes simulations to refine its processes and improve the effectiveness of its assessments. A custom-developed simulator is used to evaluate the efficiency of decisions regarding item selection, proficiency estimation, and item development priorities.

4.7.1 Simulator Structure and Capabilities

The simulator is a modular tool, mirroring the structure of the operational CAT algorithm to allow for accurate simulations. It features interchangeable modules such as the item bank, student proficiency distribution, student proficiency estimation method, test flow, and item selection methods. This approach allows for changes and adjustments as needed, while maintaining the integrity and functionality of the CAT algorithm.

The core advantage of this simulator is its capacity to evaluate changes to the item bank, item selection methods, and test flows, among other variables. It measures and provides detailed analyses of mean conditional standard error of measurement (CSEM), root mean squared error (RMSE), and bias. (See Chapter 6 for the definition of CSEM and the associated results.)

4.7.2 Simulator Outputs and Usage

The simulation studies' outcomes significantly inform the maintenance of the *i-Ready Inform* item bank. Simulations are conducted with both simulated representative normal and simulated uniform proficiency distributions to reflect the current *i-Ready* user base and to understand how students at the tails of proficiency distribution would be impacted by changes to the item bank and CAT functionality. Simulation results allow Curriculum Associates to understand how changes to various parameters may influence student assessment outcomes. As such, they serve as a validation tool for adjustments made to the item bank, test flows, and item selection methods, and are also used to inform item development. By leveraging data-driven insights, Curriculum Associates continually refines its assessments to provide more precise and beneficial results for all students.

Chapter 5: Calibration and Scaling

5.1 Chapter Summary

Chapter 5 presents the psychometric theory and supporting research on which *i-Ready Inform* is built. Evidence in Chapter 5 includes evaluating the functioning of items to identify potential issues, calibration procedures for field-test items, the development and maintenance of the vertical score scale, scale score transformation equations, and the Growth Monitoring projection model. The methods presented in this chapter demonstrate our use of best-practice approaches to build and monitor the *i-Ready Inform* score scale to establish a reliable measure for measuring and reporting student knowledge for all content domains, which provides the foundation upon which scores can be used in the various ways outlined in Chapter 1.

5.2 Introduction

Curriculum Associates uses industry-standard methods rooted in both classical test theory (CTT) and IRT to develop and maintain the *i-Ready Inform* item bank. This chapter describes the processes and procedures used to field test new items, estimate item parameters, monitor item and scale drift, transform scores to the reporting scales, and produce EOY *i-Ready Inform* score projections using the Growth Monitoring assessments.

5.2.1 Field-Testing Design

Embedded field-test (EFT) items are inserted into the test flow for eligible students in separate sections of *i-Ready Inform* during the test administration, but are not included in the Growth Monitoring assessments. Students are only administered EFT items when new items are being developed or currently operational items require evaluation. It is possible for a student's test to include no field-test items. Note that EFT items are not used in calculating a student's score.

For reading, there are two possible sections of EFT items. In the first, EFT items can be from the Phonics, Phonological Awareness, High-Frequency Words, and/or Vocabulary domains. In the second, they can be from the Comprehension: Informational Text and/or Comprehension: Literature domains. Items associated with a passage, or paired passages, are field tested together as a set. For mathematics, EFT items are grouped into two EFT sections: Algebra and Algebraic Thinking and Number and Operations, or Geometry and Measurement and Data items.

A within-test check is conducted to determine if a student is eligible for administration of EFT items. Students performing well below grade level are not included in EFT data collection. For eligible students, the system compares the student's chronological grade to the assigned grades of the EFT items to determine which items are eligible for administration. An eligible student could receive EFT items that are on grade level, meaning the item's assigned grade level matches the student's chronological grade level (e.g., a Grade 4 student receiving a Grade 4 EFT item), and/or also receive EFT items that are off grade level, wherein an item could be one grade level below (e.g., a Grade 4 student receiving a Grade 3 EFT item) or one grade level above (e.g., a Grade 4 student receiving a Grade 5 EFT item) the student's chronological grade.

Once a student is determined to be eligible to receive EFT items and the *i-Ready* system has determined the EFT items eligible to be administered, an item is randomly selected from the pool of eligible EFT items

associated with a given section. Adaptive logic is not used, meaning that there is no attempt to match EFT item difficulty to the student's proficiency level. The *i-Ready* system will continue administering random EFT items in this way until the student responds to the maximum number of items in the EFT section or the system runs out of EFT items to administer. When items are passage based, the ones associated with a passage are selected as a bundle and field tested together. Note that although EFT items may be administered to eligible students, the number of EFT items included in any given student's test is based on both test development considerations and the impact to students.

5.3 Key Validation

Key validation (i.e., key check or statistical key check in other assessment programs) is the process used to evaluate EFT items and uses a variety of item statistics to help identify potential issues with the items. This process applies several statistics to evaluate the items.

5.3.1 Classical Test Theory

CTT provides a statistical model for the number of items students answer correctly on a test. When model assumptions are satisfied, CTT facilitates the evaluation of test score precision (e.g., reliability and the standard error of measurement). In CTT, student performance is generally represented by the total number of items answered correctly, and an item's difficulty is estimated by its *p*-value, which, for dichotomous items (those scored incorrect/correct), is calculated as the proportion of students who answered the item correctly.

CTT analyses are based on the assumption that an observed score on a test may be modeled as a sum of two components: the test-taker's "true" (or expected) score and residual error specific to the observed score. In other words, CTT assumes that students possess stable characteristics that manifest in observed test scores which can be represented mathematically as:

$$X_i = T_i + e_i \qquad (5.1)$$

where X_i is the observed score for respondent i, T_i is the respondent's true score on the test, and e_i is random error that influences the expression of respondent i's true score.

5.3.2 CTT Indices Used for Evaluating Items

The key validation analysis uses a variety of indices rooted in CTT for items that have been field tested. Extreme or unexpected values can be used to help identify items that deserve closer scrutiny. However, the items that are identified based on these criteria are not automatically rejected or retained.

5.3.2.1 Item Difficulty

The *p*-value is a sample-dependent measure of item difficulty. For field-test items being evaluated for possible addition to the *i-Ready* bank, the *p*-value is calculated by dividing the number of students who responded correctly to an item by the total number of students who attempted the item. The examination of *p*-values, which range from .0 to 1 with higher values representing a greater proportion of students answering the item correctly, helps identify items that may be too easy or difficult for a given group of students. Moreover, this examination helps ensure the item bank includes items spanning the full range of item difficulties, which is necessary to precisely measure the proficiency of a wide range of students.

5.3.2.2 Item Distractor Response Proportion

The proportion of students that select each answer choice provides insight into how students interact with an item. This measure is calculated by dividing the number of students that selected the response option by the total number of students who answered the item. The response proportion is computed for each distractor, which are response alternatives that are not correct. Note that when this statistic is computed using the correct response, which is referred to as the response key, the *p*-value described in the previous section is obtained. Ideally, a higher proportion of students select the key (correct answer) compared to the distractors. Special attention is given when response proportions are higher for distractors compared to the key.

5.3.2.3 Item Discrimination

The item-total correlation (i.e., the point-biserial correlation) is the correlation between the item score and the total test score. This measure, which ranges from -1 to 1, indicates how well an item discriminates between low- and high-achieving students. Positive values indicate students with higher scores on the overall test are more likely to select the key (expected), while negative values indicate students with higher scores on the overall test are less likely to select the key (not expected). Values near zero indicate little to no relationship between a student's score and selecting the key. Higher-achieving students should be more likely to select the key than lower-achieving students, resulting in a positive correlation. A range of positive point-biserial correlations is desired, though higher values are generally preferable since items that exhibit high positive discrimination contribute most to measurement precision.

5.3.2.4 Item Distractor-Total Test Correlation

The item distractor-total test correlation, which is also a point-biserial correlation, is the correlation between each distractor and the total test score and ranges from -1 to 1. When computed on the response key, this is referred to as item discrimination; however, this value is also computed for each distractor. Positive values indicate students with higher scores on the overall test are more likely to select the distractor (not expected), while negative values indicate students with higher scores on the overall test are less likely to select the distractor being evaluated (expected). Values near .0 indicate little to no relationship between a student's score and selecting the distractor being evaluated. We would expect there to be a negative point-biserial correlation for distractors, indicating the higher the students' score, the less likely they are to have chosen the incorrect response. Special attention is provided to items with positive distractor-total correlations.

5.3.3 Item Evaluation

Field-test data are used to compute and empirically evaluate the classical item statistics described in the prior section. Field-test data, rather than operational data, are preferred because the data collection process results in a wide representation of student proficiency due to the field-test design. The item evaluation procedure evaluates if an item is functioning in a manner that is consistent with expectations and flags items with potential issues due to, for example, very low or high *p*-values, near zero or negative discrimination values, distractors (i.e., incorrect answer options) selected by a high proportion of students, INFIT or OUTFIT values that are out of a prespecified range, or a high proportion of high-achieving students selecting the incorrect answer. Ideally, items will have a positive item-total correlation for the key and negative correlations between the distractors and the total score. However, a weak positive correlation for a distractor does not necessarily indicate that the item is problematic, especially if the distractor correlation is substantially smaller in

magnitude compared to the item-total correlation. In some cases, it may simply mean that the distractor is attractive to moderate-ability students and unattractive to low-ability students.

Once items are flagged by the psychometrics team, they are thoroughly reviewed by subject-matter experts (SMEs). SMEs determine if the item is functioning as expected and should be accepted for the item pool, if the item should be revised and re-field tested, or if the item should be removed. For example, adaptive assessments require items of extremely low difficulty and extremely high difficulty to adequately measure students of low and high proficiency. In this case, the SME review might result in an item with extremely low or extremely high difficulty being retained for the item pool because the item is functioning as expected. To aid in this evaluation, SMEs are also provided IRT difficulty parameter estimates of the item.

5.4 Item Calibration and Scaling

Curriculum Associates follows industry-standard methods for calibrating items and maintaining score scales for *i-Ready Inform*. This section describes the general IRT procedures used to create the *i-Ready Inform* base scale and to transform new items onto that scale.

5.4.1 Rasch Measurement Model

Items for the *i-Ready Inform* assessments are calibrated using the Rasch (1960) measurement model. A unidimensional Rasch model assumes that the assessment measures one underlying construct—either mathematics or reading in the case of *i-Ready Inform*. In the Rasch model, the probability of a correct response for student i on item j, $P_{i,j,}$ is modeled as:

$$P_{i,j} = \frac{e^{(B_i - D_j)}}{1 + e^{(B_i - D_j)}} \quad (5.2)$$

where B_i is the proficiency estimate of the student, and D_j is the difficulty parameter of item j. One key advantage of this model is that it places item difficulty and student proficiency estimates on the same scale.

The difficulty parameter b_j in the Rasch measurement model is the location on this scale at which a student with proficiency equal to difficulty has a probability of .5 of answering the item correctly. The difficulty parameter also corresponds to the inflection point of the item-response function, which is where that function has the steepest slope and where the item provides the most information about student proficiency. This principle guides the selection of items in a computer-adaptive assessment like *i-Ready Inform*. For example, an item with difficulty (b_j) of 1 maximizes measurement information in the Rasch framework for a student with a proficiency (θ_i) of 1. If such a student were administered a large set of items, all with item difficulty equal to 1, the student would be expected to answer half of the items correctly, and the resulting proficiency estimate $(\hat{\theta}_i)$ would be highly reliable.

5.4.2 Implementation Overview

The IRT calibration process aims to generate estimates of item parameters that are independent of the specific test-taker population. These estimates are used to maintain a consistent scale over time, allowing for the

interpretation of student proficiencies and comparison of individual student scores across different populations, timeframes, and sets of items administered.

However, applying IRT poses a challenge due to the inherent identifiability characteristics of the Rasch measurement model, which refers to the fact that this model is identified up to a linear transformation. In other words, there are infinitely many equivalent solutions (i.e., sets of item parameter and student ability estimates) that fit the data equally well due to the relative nature of person and item parameter estimates. Consequently, item parameters estimated with data from a given examinee sample are not on the same scale as item parameters estimated with data from a different examinee sample, but they are linearly related. Linking procedures utilize that relationship to maintain a common scale across different test forms and administrations.

The *i-Ready* items are calibrated using the WINSTEPS® software (Linacre, 2014) using maximum likelihood estimation (MLE) to estimate Rasch parameters for all items and examinees within a particular calibration sample. The scale is maintained over time by using the calibration design to place EFT items on our scale.

WINSTEPS® is a registered trademark of John Michael Linacre.

5.4.3 Calibrating and Equating the Embedded Field-Test Items

An IRT-based scale underpins the *i-Ready* item bank and requires that new items be transformed to this scale when added to the item bank. A major advantage of this type of scale is that CAT programs like *i-Ready* can construct and deliver a test form using items previously calibrated to the bank scale. This design permits administering unique combinations of calibrated items to create custom test forms that optimize measurement precision for individual students.

To continually improve the item bank, students may encounter field-test items during the administration of *i-Ready Inform*. This section focuses on how the EFT item parameters are transformed to the *i-Ready Inform* test scale.

5.4.3.1 Sample

Student and item data are extracted from the *i-Ready* database during an EFT administration window. An EFT window extends for the length of time necessary to ensure that all EFT items have a minimum number of responses from one grade level below, the given grade level, and one grade level above. Only student responses from test scores adhering to predetermined business rules are analyzed. The data include student responses to operational and EFT items in all grade levels within the designated timeframe. Following extraction, the item responses undergo quality control checks and conversion into an incomplete data matrix before being sent to WINSTEPS for calibration.

5.4.3.2 Calibration

Calibration involves fitting the Rasch measurement model to the item-level data using WINSTEPS. The calibration design for *i-Ready Inform* involves two calibrations. In the first calibration, responses to both operational and EFT items are submitted to WINSTEPS, but all operational items are considered anchor items and their parameters are fixed to their banked values; EFT item parameters are freely estimated. This calibration design places EFT items on the existing scale; no other transformation equating is necessary.

5.4.3.3 Anchor Stability

Following the first calibration, during which all operational items are anchored to their banked values, the stability of these anchor items is evaluated using a precision criterion and a stability criterion.

Precision: Items are flagged for stability if the standard error of the Rasch difficulty parameter for the item is large relative to other items within that domain.

Stability: A Robust-*z* statistic is used to flag anchor items with extreme displacement values, which are calculated as the difference between the banked Rasch difficulty estimate and a freely estimated difficulty estimate for that item, with every other item fixed to the anchored item difficulty parameter (banked value stored for test administration). Note that WINSTEPS computes this value in the background and does not require a separate calibration run. This statistic uses the displacement interquartile range (IRQ), the median displacement, and each item's displacement value to flag items with substantial instability. The Robust-*z* statistic is calculated as:

$$z = \frac{displace - Median_{displace}}{IQR_{displace} x 0.74}. \quad (5.3)$$

Curriculum Associates computes the domain-specific median and IQR based on all items that are not flagged by the precision criterion. Using this information, items are flagged by the Robust-*z* statistic if the value falls outside a predefined range.

Items flagged by these criteria are removed from the anchor item set (i.e., their anchored status and values are removed from the anchor set and they are freely estimated) and a second calibration is conducted. It is typical for approximately 10 percent of the operational items to be removed from the common item set during this step. After the second calibration, the difficulty parameters for the EFT items are retained.

5.4.4 Item Bank Maintenance

The stability of the vertical scale over time is essential to ensure the validity of inferences about student achievement and growth based on *i-Ready Inform* results. To help ensure that scores remain comparable over time, Curriculum Associates employs several methods to maintain the item bank.

5.4.4.1 Monitoring Method

Items in the *i-Ready Inform* item bank receive a Rasch difficulty estimate when they are initially added to the item bank. These estimates are based on a random (i.e., the adaptive item selection is not activated during this process) sample of students that responded to the item, resulting in a wide range of student proficiency for the sample responding to the EFT items. While these estimates reflect the best information available at the time the item is operationalized, factors such as repeated exposure over time, changes in instructional focus, and student familiarity with the contexts of certain items (e.g., items that discuss pandemic and health issues) can affect the accuracy of these estimates over time. In other words, the difficulty value of the item can drift from its initial estimate. When items drift, their initial estimates may be less accurate representations of their difficulty, and a new estimate should be obtained.

Curriculum Associates has developed a process for monitoring items in the item bank for potential item parameter drift. Using a logistic regression procedure, the method flags items for potential drift and, with appropriate consultation with the assessment editorial team, items are either re-field tested, revised and re-field tested, or removed from the item bank.

5.4.4.2 Scale Stability Study

In January 2020, Curriculum Associates undertook a research study to evaluate the stability of the scaling and equating results for *i-Ready Inform* as new items are added to the item banks across time. For this study, the operational item bank was split into two groups: (1) items that were part of the bank prior to July 2019, and (2) items that were new as of July 2019. For *i-Ready Inform*, newly field-tested items are transformed to the *i-Ready Inform* scale via an anchored common-item calibration wherein the operational items are anchored to their banked values. In this research, the two groups of items were freely calibrated (i.e., with no items anchored to their banked values) based on fall 2019 operational student responses. Next, item parameter estimates were then equated to the bank scale.

The correlations between banked and freely estimated Rasch difficulty parameters were high (.99) for both groups of items in both subject areas. The equating scale transformation constants were very similar for the two groups of items. Overall, this research suggested that the scaling and equating methods used in *i-Ready Inform* are stable and replicable across multiple methods and with both older and newer items.

5.4.4.3 Fall 2016 Update to the Item Bank

In August 2016, the item bank was updated with recalibrated item parameters obtained from a calibration study conducted in spring 2015. The ranges, means, and standard deviations of the updated scale score distribution were either exactly the same as, or very similar to, those from the scale score distribution prior to the August 2016 scale update, which allows for approximate year-over-year comparisons of *i-Ready* scale scores.

5.4.5 Quality Control

Curriculum Associates works with a third party to conduct an independent replication of the calibration and equating steps, and results must match before moving on to the final process step. Both Curriculum Associates and the third party perform various quality control (QC) checks, and then a comparison between the two results is conducted to ensure accuracy of the results obtained. An overall summary of the calibration and equating results is created and reviewed for reasonableness before final signoff on the results is given. The following is a short description of each of the checks performed.

Accuracy checks. For both the initial and final calibration run, checks are done to ensure the anchored values in the WINSTEPS output file match the anchored values input to WINSTEPS.

Descriptive statistics. For the final calibration run, standard descriptive information (e.g., mean, standard deviation, and ranges) is obtained on various statistics (i.e., item difficulty, INFIT, OUTFIT, displacement) for the three groups of items—operational anchor items, operational non-anchor items, and EFT items.

Increasing difficulties. For the final calibration run, mean difficulty is examined by item grade level to determine whether items show increasing difficulties as grade level increases.

Anchor purification summary. After items are flagged for stability concerns based on the initial calibration, the number of anchor items identified is reported overall, by grade, and by domain. Note additional breakouts that show counts of items flagged for each criterion as well as grade by domain comparisons are also available, but they are omitted from this summary for brevity.

Summary of Rasch difficulties for first and second calibration. After both calibrations, an inspection of the initial and final item difficulties is conducted by comparing the item difficulty estimates from the initial and final calibrations. Scatterplots and correlations are generated to conduct this comparison.

INFIT/OUTFIT. EFT items with extreme INFIT/OUTFIT values, which represent how well the data for each item fit the Rasch model, are further examined.

5.5 Transformation to the Reporting Score Scale

The *i-Ready* assessments use linear transformations of proficiency estimates on the underlying logit scale for the two vertical scales: $\hat{\theta}_R$ for Reading and $\hat{\theta}_M$ for Mathematics. The primary advantages of linear transformations are simplicity and supporting score interpretation. A single transformation is applied across the entire scale for each subject. In addition, the distributions of the reported scale scores remain proportionally congruent with the distributions of the estimated θ scores, and the conditional standard errors are proportionally scaled relative to the slope constant and maintain the same convex, U-shaped pattern on both the θ metric and reported score scale.

All estimates at both the overall test level and the domain level are transformed from their operational θ (or logit) scale to a scale score using the following linear transformation functions for Reading and Mathematics, respectively:

$$Unrounded\ Reading\ Scale\ Score = 499.38 + \left(37.81 \times\ \hat{\theta}_R\right), \text{and} \tag{5.4}$$

$$Unrounded\ Mathematics\ Scale\ Score = 466.41 + \left(25.42 \times\ \hat{\theta}_M\right). \tag{5.5}$$

After student scores are scaled using the above transformations, they are rounded to the nearest integer, and a floor and ceiling of 100 and 800, respectively, are applied by replacing any scale scores below 100 with the lowest observable scale score (LOSS) of 100 and replacing any scale scores above 800 with the highest observable scale score (HOSS) of 800.

5.6 Growth Monitoring Projected Scores

Students who are administered Growth Monitoring assessments receive a projected EOY score, rather than a score based solely on the items administered. This projection is based on all assessments taken by that student—either *i-Ready Inform* or Growth Monitoring—during that school year.

5.6.1 Projection Model

The Growth Monitoring projection model was developed after the first full-year implementation of the assessment. Several models were evaluated in an extensive research study, in collaboration with independent researchers from Harvard University. The model that had the best psychometric characteristics (e.g., low residual, low residual bias, consistent projection precision across the school year, etc.) and was operationally feasible was selected. The final projection model has the following key structural features:

- The projection is based on a weighted combination of two values:
  - The average across all test scores a student receives during the academic year, including *i-Ready Inform* and Growth Monitoring (grand mean, or GM)
  - Predicted EOY scale score based on a simple linear regression (linear prediction, or LP)
- Weighting of the GM and the LP is determined by fitting multiple linear regression models to the preceding year's assessment data on the relationship between GM and LP and the actual EOY *i-Ready Inform* test scores students obtained in the previous year.
- A set of multiple regression intercepts and weighting factors is derived for each of the nine grades (K–8), two subjects, three ability groups based on fall percentile rank (bottom 25%, middle 50%, and top 25%), and eight months (October to May). Thus, a total of 432 ($9 \times 2 \times 3 \times 8$) sets of model parameters are developed.

5.6.2 Example

A Grade 1 student takes the first *i-Ready Inform* for Reading test in November and one Growth Monitoring test in each month of December through February. The four test scores are 450, 463, 476, and 469. The GM across these test scores is 465; the LP projection is 497. The Reading projection model for Grade 1 students in this proficiency group when the last test is taken in February is:

$$Projected\ Score = 84.29 + GM \times 0.78 + LP \times 0.09. \quad (5.6)$$

Thus, the final projected EOY score for this student is 492.

Model parameters are obtained for each month, so the projection error stays low even at the beginning of the school year when the number of data points is small. Note that at least three assessment data points taken in three different months are required to receive a projection. Also, since model parameters are obtained for three proficiency groups, the differential growth rate for students at the high and low ends of the proficiency spectrum is taken into consideration.

5.6.3 Supporting Research

To illustrate the accuracy of the Growth Monitoring projection model, all students from the 2014–2015 school year were randomly assigned into one of two samples: the training sample or the validation sample. The training sample was used to derive the weighting parameters for each of the 432 models. These parameters were then applied to the validation sample. Figure 5.1 and Figure 5.2 below show the normalized root-mean-square error (NRMSE) from the validation sample. NRMSE is zero when the prediction matches perfectly to the

actual test score; an NRMSE of less than .10 is considered adequate fit. The figures for both Reading and Mathematics show that while the prediction error is relatively higher in October when only three months of test data are available and the projection is more than six months out, it quickly drops to a lower level (i.e., most are below .10) in November and stays low and stable across the rest of the year.

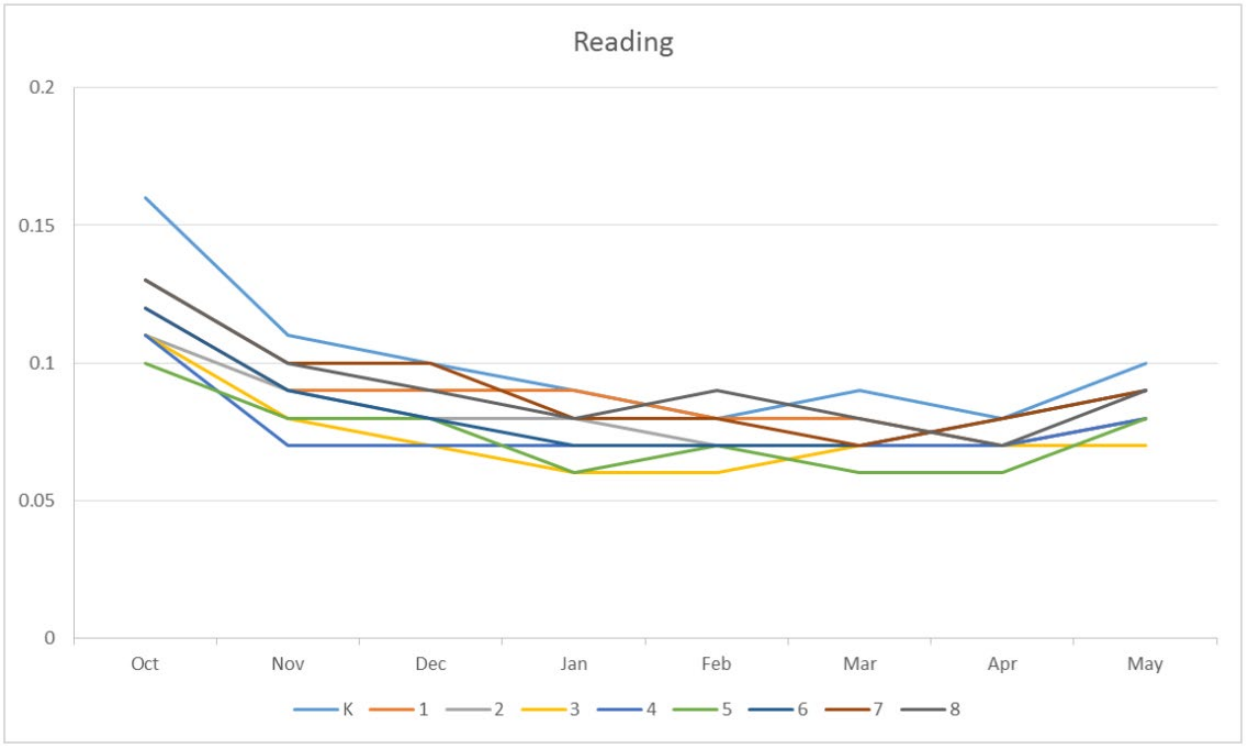


Figure 5.1. Normalized Root-Mean-Squared Error (NRMSE) for Reading

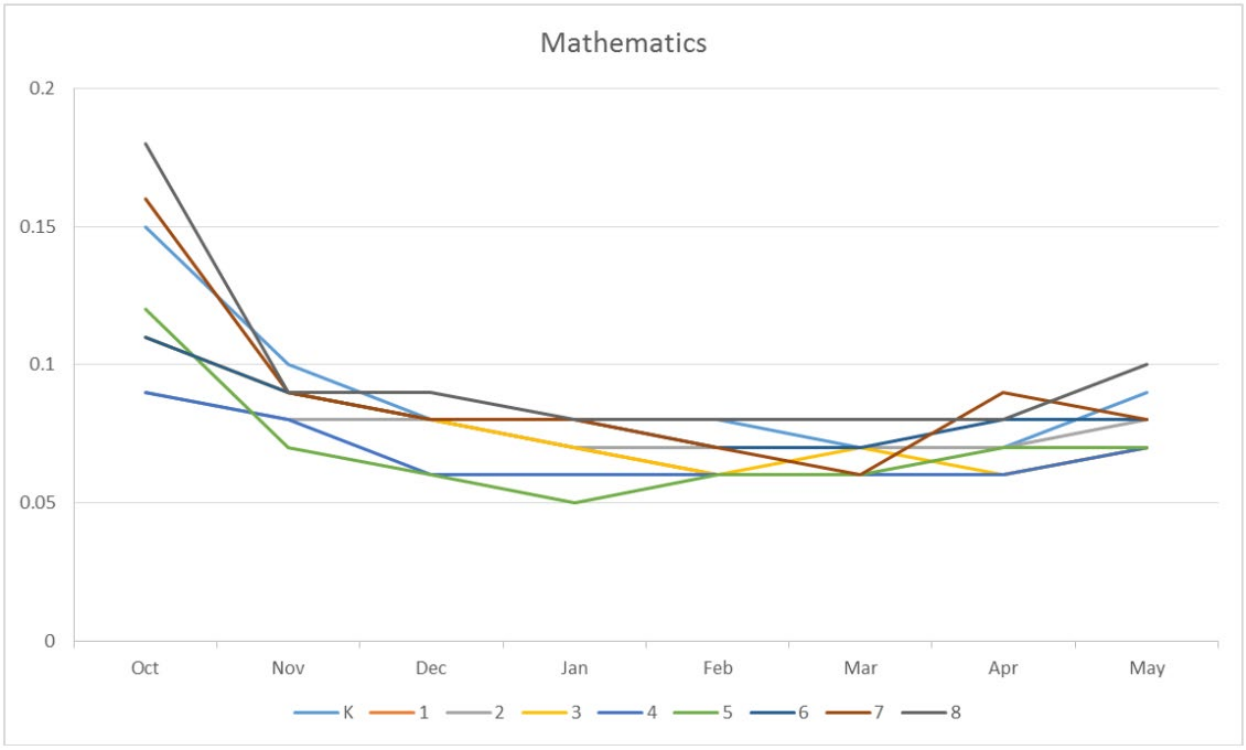


Figure 5.2. Normalized Root-Mean-Squared Error (NRMSE) for Mathematics

The Growth Monitoring projection model was built on a large amount of empirical data. The NRMSE values from the study presented above suggest the current model provides adequate accuracy in predicting the students' EOY performance level.

Chapter 6: Reliability

6.1 Chapter Summary

Chapter 6 reports marginal reliability estimates for the overall assessment, domains, and various demographic groups as critical evidence of the overall internal consistency of *i-Ready Inform*. Test–retest reliability estimates are also reported. The overall marginal reliability estimates for the total population and for various student groups are provided, and domain marginal reliability estimates are reported for both reading and mathematics. The reliability estimates reflect the relative cohesiveness and integrity of the test's internal structure, which provides evidence that the test measures its intended constructs consistently with precision. The chapter concludes by presenting classification accuracy and consistency results.

6.2 Introduction

In the context of educational measurement, reliability generally refers to the measurement precision of scores from a particular instrument or assessment. Reliability can be thought of as an estimate of internal consistency or the correlation between scores from parallel forms, but it can also be thought of in a broader sense that includes the study of measurement consistency across many variable facets of assessment contexts (AERA, APA, NCME, 2014).

Reliability is necessary—although not sufficient in itself—for the valid interpretation and use of assessment scores. Test scores are reliable when an instrument produces similar scores across repeated administrations, assuming certain conditions (e.g., measurement occasions occur within a short period of time, there is no learning or intervention, the items are varied, there are no context effects, etc.). Because it is frequently impossible or undesirable to administer multiple test forms to students at once or within a short period, reliability coefficients are commonly estimated with data from a single testing occasion.

Marginal reliability and test–retest reliability are both coefficients that estimate the theoretical correlation between parallel forms of a test. In the case of the former, the CTT definition of reliability is extended for use in an IRT-based CAT, in which students are administered different combinations of items and only test once within an administration window. To estimate the latter, we identified a sample of students who tested multiple times and computed the correlation between their scores.

The CSEM is a property of test scores that quantifies the degree of measurement imprecision. CSEMs can be averaged across a population to produce a test-level statistic known as the standard error of measurement (SEM). CSEMs and SEMs are useful for interpreting test scores because they quantify the expected variability of test scores across multiple test administrations. When reliability is high, the standard error is low, therefore the variability of test scores across administrations would be small.

Marginal reliability and SEM analyses are conducted at the test level, the domain level, and for select student groups. Computing these metrics at the domain level helps demonstrate that our domain scores produce useful information; computing these metrics for select student groups provides evidence supporting the fairness of our assessments.

Classification accuracy and consistency measure the reliability of the classifications into placement levels. Classification accuracy reflects to the degree of certainty that students are placed in the correct placement levels, whereas classification consistency reflects the degree of certainty that students would be placed in the *same*—but not necessarily the correct—placement levels across two administrations.

Unless otherwise specified, analyses for Mathematics (English) and Reading (English) were conducted using data from the 2021–2022 school year; analyses for Mathematics (Spanish) were conducted using data from the 2023–2024 school year. Filtering rules were applied to retain as much data as possible while removing potentially invalid scores. For *i-Ready Inform* data, a maximum of one test per student per administration window was allowed for each subject and grade combination. The test–retest analyses involved a separate data extraction procedure to identify students who tested twice within a short period.

This chapter is organized around evidence of reliability for *i-Ready Inform*: test-level reliability results are presented first, which include marginal and test–retest reliability as well as SEM, then domain-level results, followed by results for select student groups, and finally estimates of classification accuracy and consistency.

6.3 Test-Level Reliability

Test-level measurement precision is analyzed using both marginal and test–retest reliability coefficients as well as the SEM, which is closely related to the marginal reliability. This section concludes with several plots of students' CSEMs against their scale scores, which demonstrate low levels of measurement error across a broad range of proficiencies due to the effectiveness of our CAT algorithm and the quality and size of our item bank (see Chapter 4 for more information).

6.3.1 Marginal Reliability

Whereas reliability coefficients like coefficient alpha reflect the precision of raw test scores (i.e., number correct), marginal reliability reflects the precision of estimated latent ability, which is more appropriate in the CAT context since there are no fixed test forms, and reported scores are based on proficiency estimates. The calculation of marginal reliability is rooted in the CTT framework for reliability, in which reliability is conceptualized as the squared correlation between two parallel test forms (Lord & Novick, 1968). Using this definition and applying the CTT formulation of an observed test score X being composed of the sum of the true score T and error term E, the formula for reliability can be rearranged and written using quantities that are automatically produced for IRT-based tests (Sireci, Thissen, & Wainer, 1991) such as *i-Ready Inform*. Thus, the formula for the reliability coefficient ($\hat{\rho}$) can be written as:

$$\hat{\rho} = 1 - \frac{\sigma_E^2}{\sigma_\theta^2}, \text{where} \tag{6.1}$$

$$\sigma_E^2 = \widehat{SEM}_\theta^2 = \frac{\sum_{i=1}^{N} CSEM_{\hat{\theta}_i}^2}{N}, \text{and} \tag{6.2}$$

$$\widehat{CSEM}_{\hat{\theta}_i} = \frac{1}{\sqrt{I(\hat{\theta}_i)}}, \text{and} \tag{6.3}$$

$$I(\hat{\theta}_i) = \sum_{j=1}^{J} P_{ij}\left(1 - P_{ij}\right), \tag{6.4}$$

and where i indexes students from 1 to N, j indexes items from 1 to J, σ_E^2 is variance of the error terms, σ_θ^2 is the variance of the observed scores on the logit scale, $\hat{\theta}_i$ is student i's estimated proficiency on the logit scale, $CSEM_{\hat{\theta}_i}$ is student i's $CSEM$ for $\hat{\theta}_i$, $SEM_{(\theta)}$ is the SEM in θ or logit units (not for a particular value of θ), $I(\hat{\theta}_i)$ is the test information function for student i conditional on their proficiency estimate when using a Rasch IRT model and maximum likelihood scoring, and P_{ij} is the probability of a correct response of student i to item j, which is defined in Chapter 5. Note that the summand in Equation 6.4, $P_{ij}(1 - P_{ij})$, is the information for a single item for a given student.

The reliability coefficient can have a value ranging from 0 to 1, with higher values indicating more reliable test scores and lower SEMs. A common rule of thumb is that reliability coefficients greater than or equal to .80 are considered acceptable for tests of moderate lengths.

Marginal reliability coefficients are presented in Table 6.1–Table 6.3. Reliability coefficients for *i-Ready Inform* for Mathematics in English range from .95 to .99, and for *i-Ready Inform* for Reading, they range from .95 to .98. For *i-Ready Inform* for Mathematics in Spanish, reliabilities range from .93 to .97. All reliability coefficients indicate strong measurement precision across grades and languages.

6.3.2 Standard Error of Measurement

An observed test score is not a student's *true* score. Indeed, the true score is a theoretical concept and cannot therefore be measured directly in practice. However, through the careful construction of tests and rigorous psychometric practices, observed test scores can estimate true scores with a high degree of precision. The uncertainty of these estimates can be quantified and is referred to as the standard error of the proficiency estimate, or the CSEM. The SEM is simply the average of the CSEMs across all examinees, as shown in Equation 6.2, but the conceptual discussion in this section will focus on the CSEM. There are a variety of ways to maximize measurement precision, such as increasing test length, using a CAT algorithm, implementing good item writing practices, and developing a robust item bank—all of which contribute to *i-Ready Inform*'s strong measurement precision.

As with reliability, one way to conceptualize the CSEM is in the context of repeated testing events. For example, if an individual was given the same test repeatedly within a short period, assuming the conditions

outlined above, that person would not be expected to earn the same test score every time. Rather, the student would have a distribution of scores. The standard error, or the CSEM, is the standard deviation of that distribution, and it is computed using Equation 6.3. The spread of this distribution (i.e., the CSEM) and the reliability are inversely related. Furthermore, this distribution can be shown to be normally distributed, which allows the units to be interpreted accordingly. For example, we can be 68 percent confident that the true score is contained within one standard error in either direction of the proficiency estimate, and 95 percent confident that the true score is contained within approximately two standard errors in either direction of the proficiency estimate. CSEMs are calculated for each student based on their responses to the specific items they were administered. CSEMs vary in magnitude across the range of proficiency estimates and are usually lower in the middle of the score distribution and higher at the tails—a pattern that is both common and expected.

It is imperative to note that the actual computation of the CSEM in Equation 6.3 occurs on the underlying proficiency scale, often referred to as the "θ scale" or the "logit scale." This is not the metric used for reporting; rather a linear transformation, along with rounding and truncation rules, is used to transform the proficiency estimates and their CSEMs from the underlying scale to corresponding values on the reported scale, which ranges from 100 to 800. The formula to transform the CSEMs is as follows:

$$CSEM_{SS} = B \times CSEM_{\theta}, \tag{6.5}$$

where B is the scale transformation constant (37.81 for Reading and 25.42 for Mathematics, see Chapter 5 for more details). It is important to note that transformed CSEMs are specific to the characteristics of the reporting scale, in particular its range. For example, if the underlying CSEM and all other things are held constant, wider scales result in larger transformed CSEMs—and SEMs—than narrower scales, but this difference is purely an artifact of the transformation.

These transformed CSEMs can be averaged using Equation 6.2, but with the θ subscripts replaced with SS to denote scale score, to produce the SEM on the reporting score scale. Alternatively, the SEM can be transformed directly using the following equation:

$$SEM_{SS} = B \times SEM_{\theta}. \tag{6.6}$$

This relationship between the underlying and reporting scales means that SEMs in reporting scale units cannot be directly compared between assessment programs. Rather, other scale-free metrics such as reliability coefficients, are more amenable to comparison.

SEMs are presented in Table 6.1–Table 6.3. SEMs for *i-Ready Inform* for Mathematics in English range from 6.3 to 6.5, and for *i-Ready Inform* for Reading, they range from 9.3 to 11.1. For *i-Ready Inform* for Mathematics in Spanish, SEMs range from 6.4 to 6.5. Such small SEMs on a broad score scale reflect strong measurement precision across grades and languages. Note that SEMs for Reading are several points higher than the SEMs for Mathematics. Their reporting scales have the same width, yet the SEMs differ mostly due to the difference in the scaling constants *(B)* for Reading and Mathematics (37.81 versus 25.42, respectively); their corresponding SEM_{θ} values are similar.

Table 6.1. Marginal Reliability and Standard Error of Measurement for *i-Ready Inform* for Mathematics in English

Grade	Sample Size	Marginal Reliability	SEM
K	2,281,818	0.95	6.5
1	2,600,489	0.96	6.4
2	2,774,663	0.96	6.4
3	2,870,693	0.96	6.4
4	2,826,144	0.97	6.4
5	2,854,279	0.97	6.4
6	2,352,756	0.97	6.4
7	2,081,161	0.97	6.4
8	1,961,463	0.98	6.4
9	283,093	0.98	6.3
10	177,740	0.98	6.3
11	118,950	0.98	6.3
12	68,945	0.99	6.3

Table 6.2. Marginal Reliability and Standard Error of Measurement for *i-Ready Inform* for Reading in English

Grade	Sample Size	Marginal Reliability	SEM
K	1,878,648	0.95	9.3
1	2,175,773	0.97	9.5
2	2,328,438	0.97	10.5
3	2,486,736	0.97	10.2
4	2,452,357	0.97	10.4
5	2,469,573	0.97	10.6
6	2,023,140	0.97	10.7
7	1,839,402	0.97	10.7
8	1,808,160	0.97	10.8
9	313,476	0.98	10.9
10	214,492	0.98	11.0
11	146,513	0.98	11.1
12	98,845	0.98	11.1

Table 6.3. Marginal Reliability and Standard Error of Measurement for *i-Ready Inform* for Mathematics in Spanish

Grade	Sample Size	Marginal Reliability	SEM
K	79,291	0.94	6.5
1	79,865	0.95	6.4

2	75,411	0.96	6.4
3	48,678	0.96	6.4
4	37,293	0.96	6.4
5	32,963	0.96	6.4
6	22,866	0.96	6.4
7	19,585	0.96	6.5
8	18,753	0.96	6.5
9	5,137	0.97	6.4
10	3,165	0.97	6.4
11	1,300	0.97	6.5
12	518	0.98	6.4

Figure 6.1–Figure 6.3 show the mean CSEM at each scale score of a random sample of students across all grades. The vertical bars identify the location of the 1st and 99th percentiles of the data used to generate the plot, demonstrating that CSEM is both low and relatively consistent across the vast majority of students.

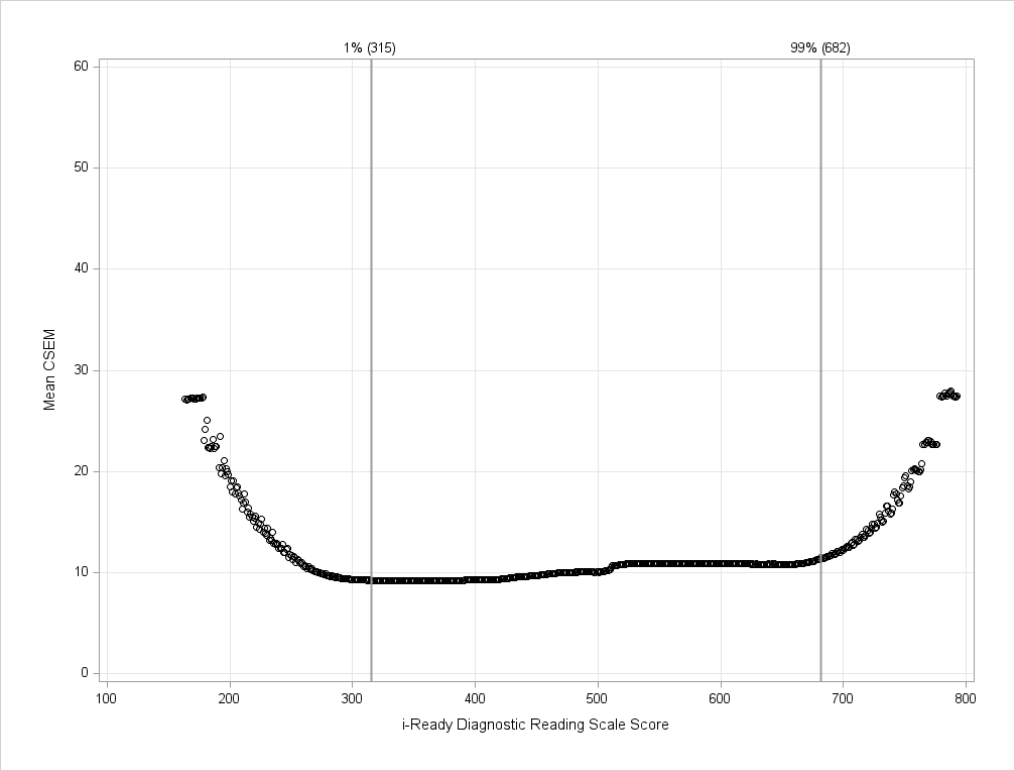


Figure 6.1. Conditional SEM of i-Ready Inform *Assessments for Reading*

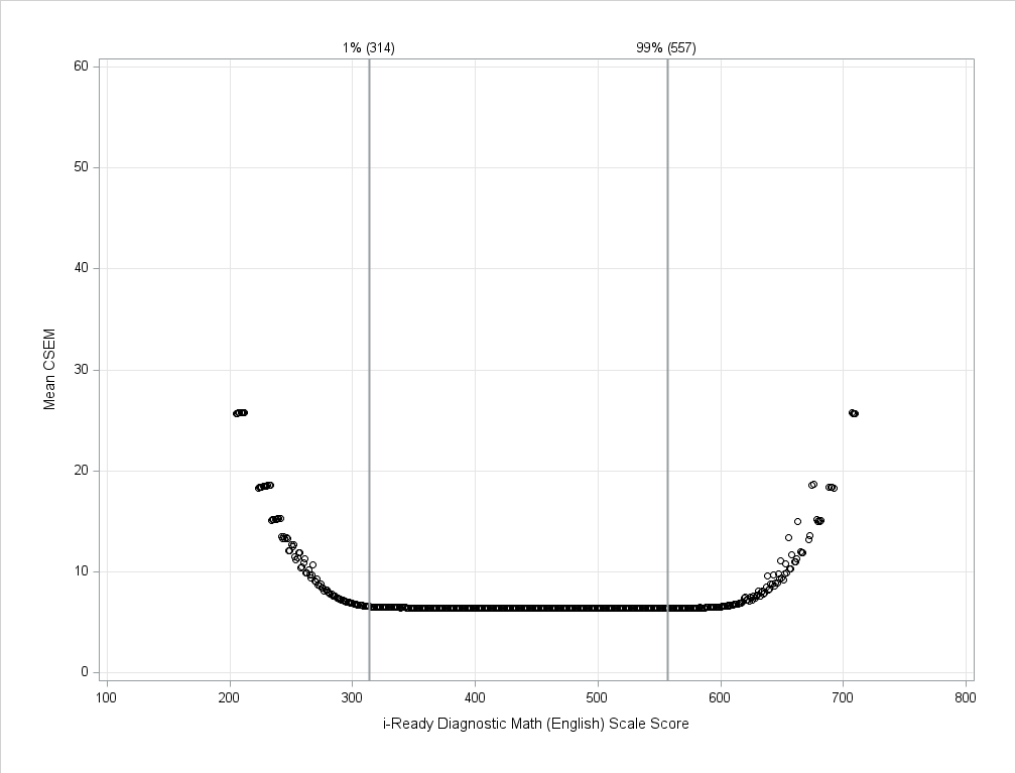


Figure 6.2. Conditional SEM of i-Ready Inform *Assessments for Mathematics*

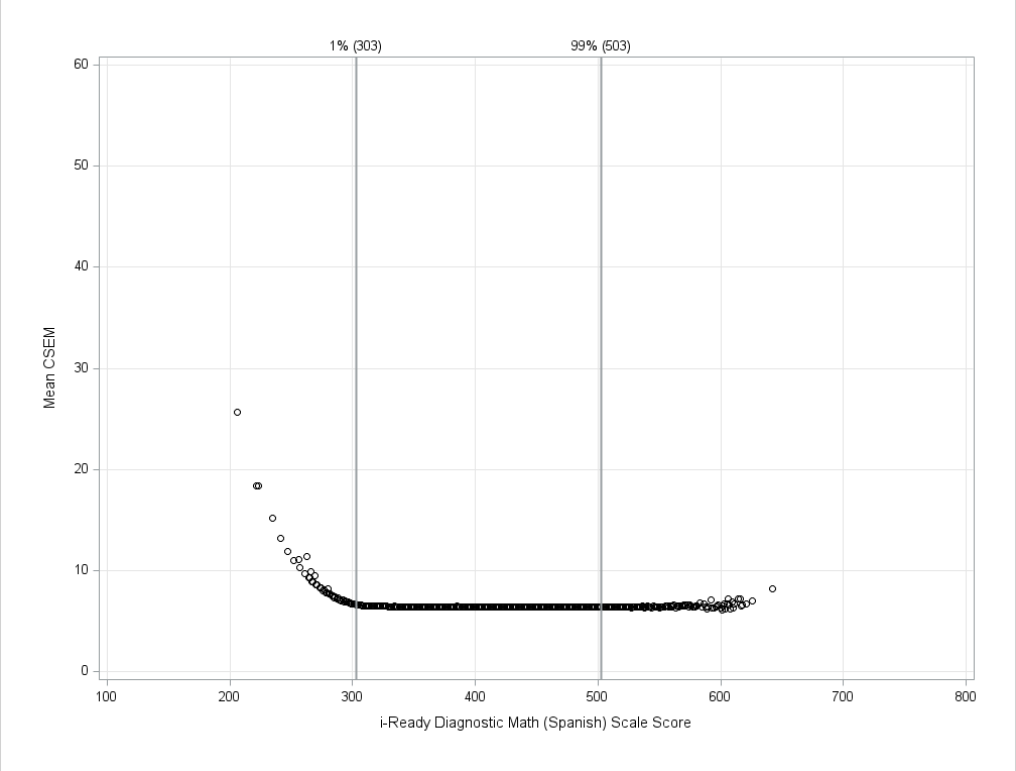


Figure 6.3. Conditional SEM of i-Ready Inform *Assessments for Mathematics (Spanish)*

6.3.2.1 Test–Retest Reliability

Ideally, estimating test–retest reliability would involve a single group of students, in which each student takes two different test forms and there is counterbalancing to remove effects due to the order of administration. The observed correlation between scores on the two test forms would estimate test–retest reliability. Our administration procedures are not designed to collect such data; thus, the analyses that follow in this section do not necessarily reflect typical usage patterns for *i-Ready Inform*.

To estimate test–retest reliability, we identified a sample of students with tests that were administered relatively close together in time. In this data set, the average number of weeks between administrations ranges from approximately 14 to 17, which allows for substantial learning to have taken place, and this could potentially reduce the correlation between testing events. The purpose of the test is specifically to inform instruction and to provide additional instruction and support to students to enhance their learning. Therefore, there is reason to believe that the reliability coefficients reported here are underestimates of test–retest reliability.

Additionally, because the students identified for this analysis are a subset of the full sample, there are some cases in which the reliability coefficients are likely to be further underestimated due to a reduced range of scores in the subsample. Thus, reliabilities corrected for restriction of range are presented, which were calculated as $\hat{\rho}_{TR_c} = 1 - \sigma_{TR}^2(1 - \hat{\rho}_{TR})/\ \sigma_F^2$, where σ_{TR}^2 is the variance of the observed scores in the test–retest sample, which is pooled across the testing events and specific to subject and language, $\hat{\rho}_{TR}$ is the uncorrected test–retest reliability, and σ_F^2 is the variance of observed scores used in the marginal reliability calculation for either Mathematics (English) and Reading (English) in Equation 6.1.

For English-language tests, test–retest reliabilities exceeded .80, with the exception of Grade K, and they often exceeded .90, as shown in Table 6.4 and Table 6.5. Mathematics in Spanish test–retest reliabilities are shown in Table 6.6 and were slightly lower, but the sample size was much lower than the English-language tests.

Table 6.4. Test–Retest Reliability for *i-Ready Inform* for Mathematics in English

Grade	Sample Size	Reliability	Corrected Reliability
K	137,818	0.73	0.79
1	660,603	0.82	0.86
2	750,132	0.86	0.89
3	784,848	0.88	0.90
4	766,923	0.90	0.91
5	764,465	0.90	0.91
6	587,197	0.90	0.91
7	486,787	0.90	0.90
8	432,921	0.89	0.90
9	39,272	0.89	0.89
10	22,942	0.89	0.89
11	14,107	0.89	N/A
12	7,237	0.90	N/A

Note: N/A = not applicable because the variance of the full dataset was not greater than that of the dataset used for estimating test–retest reliability

Table 6.5. Test–Retest Reliability for the *i-Ready Inform* for Reading in English

Grade	Sample Size	Reliability	Corrected Reliability
K	108,387	0.75	0.78
1	547,126	0.87	0.88
2	612,917	0.91	0.91
3	666,317	0.90	0.91
4	635,548	0.90	0.91
5	621,130	0.90	0.91
6	466,000	0.90	0.91
7	390,678	0.90	0.91
8	368,904	0.89	0.90
9	39,492	0.90	0.90
10	25,559	0.90	0.90
11	15,376	0.90	N/A
12	9,098	0.91	N/A

Note: N/A = not applicable because the variance of the full dataset was not greater than that of the dataset used for estimating test–retest reliability

Table 6.6. Test–Retest Reliability for the *i-Ready Inform* for Mathematics in Spanish

Grade	Sample Size	Reliability	Corrected Reliability
K	5002	0.63	0.73
1	20937	0.73	0.79
2	19553	0.81	0.84
3	11087	0.84	0.86
4	8125	0.87	0.87
5	7023	0.87	0.87
6	4496	0.83	0.86
7	3614	0.80	0.85
8	3514	0.78	0.85
9	476	0.74	0.84
10	422	0.78	0.83
11	116	0.74	0.85
12	**	**	**

Note: ** = not reported due to a sample size of less than 100

6.4 Reliability for Select Groups

Although demographic information is limited and not provided for every student, it is prudent to evaluate measurement precision among select student groups where possible. Ideally, marginal reliability and SEM for

select student groups should be similar to the corresponding indices computed using all students, presented earlier; however, if the proficiency distribution is dissimilar to that of the population of test-takers, these statistics can be affected, particularly marginal reliability, which is affected by the overall variance of the scores. Marginal reliability coefficients and SEMs are computed for select student groups at the test level and are presented in Table 6.7–Table 6.11 for *i-Ready Inform* for Mathematics in English, Table 6.12–Table 6.16 for *i-Ready Inform* for Reading in English, and Table 6.17–Table 6.20 *i-Ready Inform* for Mathematics in Spanish. Overall, the reliability coefficients and SEMs for economically disadvantaged students, English Learners, students in special education, different racial/ethnic groups, and different genders are similar to those for all test-takers.

Table 6.7. Marginal Reliability and Standard Error of Measurement for Select Groups for *i-Ready Inform* for Mathematics in English

Economically Disadvantaged

Grade	*N*	Rel.	SEM
K	174,740	0.94	6.5
1	213,573	0.95	6.4
2	232,868	0.95	6.4
3	244,040	0.96	6.4
4	243,304	0.96	6.4
5	247,374	0.96	6.4
6	213,565	0.97	6.4
7	183,742	0.97	6.4
8	176,709	0.97	6.4
9	38,091	0.98	6.3
10	28,162	0.98	6.3
11	20,508	0.98	6.3
12	14,283	0.98	6.3

English Learners

Grade	*N*	Rel.	SEM
K	173,750	0.94	6.5
1	202,882	0.95	6.4
2	224,697	0.96	6.4
3	228,325	0.96	6.4
4	214,233	0.96	6.4
5	195,453	0.97	6.4
6	131,711	0.97	6.4
7	109,310	0.97	6.4
8	98,766	0.97	6.4
9	15,594	0.97	6.4

10	11,699	0.98	6.4
11	7,874	0.98	6.4
12	4,530	0.98	6.4

Special Education

Grade	*N*	Rel.	SEM
K	100,598	0.95	6.5
1	131,972	0.97	6.5
2	157,899	0.97	6.4
3	183,042	0.97	6.4
4	185,289	0.98	6.4
5	196,837	0.98	6.4
6	146,954	0.98	6.4
7	128,553	0.98	6.4
8	119,404	0.98	6.4
9	22,824	0.98	6.4
10	16,701	0.98	6.4
11	12,287	0.98	6.4
12	8,375	0.99	6.5

Note: Rel. = Reliability

Table 6.8. Marginal Reliability and Standard Error of Measurement for Select Groups for *i-Ready Inform* for Mathematics in English

American Indian or Alaskan

Grade	*N*	Rel.	SEM
K	9,377	0.94	6.5
1	10,135	0.95	6.4
2	10,982	0.96	6.4
3	11,261	0.96	6.4
4	11,333	0.96	6.4
5	11,277	0.97	6.4
6	9,653	0.97	6.4
7	8,938	0.97	6.4
8	8,923	0.97	6.4
9	1,532	0.97	6.3
10	1,088	0.98	6.3
11	742	0.98	6.3
12	610	0.98	6.3

Asian

Grade	*N*	Rel.	SEM
K	71,272	0.95	6.5
1	94,166	0.96	6.5
2	101,842	0.96	6.4
3	106,272	0.97	6.4
4	103,955	0.97	6.4
5	100,695	0.97	6.4
6	82,889	0.97	6.4
7	67,878	0.98	6.4
8	66,089	0.98	6.5
9	9,652	0.98	6.2
10	8,072	0.98	6.3
11	6,153	0.98	6.3
12	3,582	0.99	6.3

Black or African American

Grade	*N*	Rel.	SEM
K	320,878	0.94	6.5
1	340,789	0.95	6.4
2	360,404	0.95	6.4
3	370,383	0.96	6.4
4	355,032	0.96	6.4
5	366,349	0.96	6.4
6	295,578	0.97	6.4
7	276,321	0.97	6.4
8	262,295	0.97	6.4
9	47,208	0.98	6.3
10	25,902	0.98	6.3
11	19,140	0.98	6.4
12	10,796	0.99	6.4

Note: Rel. = Reliability

Table 6.9. Marginal Reliability and Standard Error of Measurement for Select Groups for *i-Ready Inform* for Mathematics in English

Hispanic

Grade	*N*	Rel.	SEM
K	499,364	0.94	6.5
1	573,279	0.95	6.4

2	618,633	0.96	6.4
3	651,606	0.96	6.4
4	645,987	0.96	6.4
5	666,042	0.97	6.4
6	545,674	0.97	6.4
7	491,944	0.97	6.4
8	466,649	0.97	6.4
9	71,931	0.98	6.3
10	52,645	0.98	6.3
11	37,893	0.98	6.3
12	24,615	0.98	6.3

Native Hawaiian or Pacific Islander

Grade	*N*	Rel.	SEM
K	9,994	0.94	6.5
1	11,462	0.95	6.4
2	13,347	0.96	6.4
3	13,314	0.96	6.4
4	13,513	0.96	6.4
5	13,802	0.97	6.4
6	12,588	0.97	6.4
7	8,451	0.97	6.4
8	9,041	0.98	6.4
9	2,081	0.97	6.3
10	1,546	0.97	6.2
11	1,449	0.97	6.3
12	721	0.98	6.3

Other

Grade	*N*	Rel.	SEM
K	715	0.95	6.5
1	956	0.95	6.4
2	1,311	0.95	6.4
3	1,655	0.96	6.4
4	1,592	0.96	6.4
5	1,542	0.96	6.4
6	1,357	0.97	6.4
7	1,058	0.97	6.4

8	1,025	0.97	6.4
9	117	0.98	6.2
10	**	**	**
11	**	**	**
12	**	**	**

Note: Rel. = Reliability; ** = not reported due to a sample size of less than 100

Table 6.10. Marginal Reliability and Standard Error of Measurement for Select Groups for *i-Ready Inform* for Mathematics in English

Two or More Races

Grade	*N*	Rel.	SEM
K	80,735	0.95	6.5
1	88,114	0.96	6.4
2	92,976	0.96	6.4
3	90,433	0.96	6.4
4	87,454	0.96	6.4
5	83,691	0.97	6.4
6	65,284	0.97	6.4
7	54,353	0.97	6.4
8	49,480	0.98	6.4
9	7,484	0.98	6.3
10	4,099	0.98	6.3
11	2,235	0.98	6.3
12	1,084	0.99	6.4

White

Grade	*N*	Rel.	SEM
K	642,754	0.95	6.5
1	738,082	0.95	6.4
2	783,399	0.95	6.4
3	792,876	0.95	6.4
4	781,469	0.96	6.4
5	787,927	0.96	6.4
6	653,818	0.97	6.4
7	567,044	0.97	6.4
8	530,213	0.97	6.4
9	76,708	0.98	6.3
10	46,980	0.98	6.2
11	28,379	0.98	6.3

12	15,573	0.99	6.4

Note: Rel. = Reliability

Table 6.11. Marginal Reliability and Standard Error of Measurement for Select Groups for *i-Ready Inform* for Mathematics in English

Male

Grade	*N*	Rel.	SEM
K	1,024,018	0.95	6.5
1	1,155,146	0.96	6.4
2	1,239,741	0.96	6.4
3	1,289,876	0.97	6.4
4	1,265,983	0.97	6.4
5	1,282,971	0.97	6.4
6	1,058,641	0.97	6.4
7	953,358	0.98	6.4
8	902,517	0.98	6.4
9	138,883	0.98	6.3
10	86,822	0.98	6.3
11	58,400	0.98	6.3
12	33,950	0.99	6.4

Female

Grade	*N*	Rel.	SEM
K	1,001,084	0.94	6.5
1	1,132,504	0.95	6.4
2	1,206,957	0.95	6.4
3	1,238,457	0.96	6.4
4	1,216,580	0.96	6.4
5	1,232,326	0.96	6.4
6	1,013,768	0.97	6.4
7	910,939	0.97	6.4
8	859,073	0.97	6.4
9	123,468	0.98	6.3
10	78,029	0.98	6.3
11	52,609	0.98	6.3
12	30,012	0.98	6.3

Note: Rel. = Reliability

Table 6.12. Marginal Reliability and Standard Error of Measurement for Select Groups for *i-Ready Inform* for Reading in English

Economically Disadvantaged

Grade	*N*	Rel.	SEM
K	151,970	0.94	9.3
1	186,083	0.96	9.4
2	201,885	0.97	10.3
3	217,060	0.97	10.0
4	213,969	0.97	10.3
5	216,908	0.97	10.5
6	191,901	0.97	10.6
7	165,036	0.97	10.7
8	161,454	0.97	10.7
9	36,500	0.97	10.8
10	30,501	0.97	10.9
11	23,131	0.97	11.0
12	16,775	0.97	11.0

English Learners

Grade	*N*	Rel.	SEM
K	158,782	0.95	9.3
1	190,562	0.96	9.3
2	203,022	0.97	10.1
3	211,274	0.97	9.8
4	193,202	0.97	10.1
5	176,596	0.97	10.2
6	124,989	0.98	10.3
7	108,337	0.98	10.4
8	99,828	0.98	10.5
9	20,442	0.98	10.6
10	15,560	0.98	10.7
11	11,267	0.98	10.7
12	7,399	0.98	10.8

Special Education

Grade	*N*	Rel.	SEM
K	90,210	0.95	9.3

1	120,706	0.97	9.4
2	142,727	0.98	10.2
3	170,038	0.98	9.8
4	167,098	0.98	10.0
5	177,192	0.98	10.2
6	128,346	0.98	10.3
7	116,396	0.98	10.4
8	110,230	0.98	10.5
9	23,388	0.98	10.6
10	17,355	0.98	10.7
11	13,312	0.98	10.7
12	10,153	0.99	10.7

Note: Rel. = Reliability

Table 6.13. Marginal Reliability and Standard Error of Measurement for Select Groups for *i-Ready Inform* for Reading in English

American Indian or Alaskan

Grade	*N*	Rel.	SEM
K	7,211	0.95	9.3
1	7,913	0.96	9.4
2	8,983	0.97	10.3
3	9,203	0.97	10
4	9,306	0.97	10.3
5	9,119	0.97	10.5
6	7,705	0.97	10.6
7	7,377	0.97	10.7
8	7,485	0.97	10.7
9	1,513	0.97	10.8
10	1,185	0.98	10.9
11	837	0.98	10.9
12	695	0.98	11

Asian

Grade	*N*	Rel.	SEM
K	63,657	0.97	9.4
1	86,392	0.97	9.8
2	93,511	0.97	10.9
3	99,699	0.97	10.4
4	97,043	0.97	10.7

5	94,198	0.97	10.7
6	83,193	0.96	10.8
7	69,869	0.96	10.9
8	72,428	0.96	11
9	15,736	0.97	11.1
10	14,289	0.97	11.3
11	11,619	0.97	11.5
12	9,173	0.97	11.7

Black or African American

Grade	*N*	Rel.	SEM
K	283,395	0.94	9.3
1	305,708	0.96	9.4
2	326,120	0.97	10.3
3	343,942	0.97	9.9
4	324,728	0.97	10.3
5	331,905	0.97	10.5
6	262,939	0.97	10.6
7	241,508	0.97	10.6
8	236,168	0.97	10.7
9	46,772	0.98	10.8
10	28,422	0.98	10.9
11	19,517	0.98	10.9
12	11,829	0.98	10.9

Note: Rel. = Reliability

Table 6.14. Marginal Reliability and Standard Error of Measurement for Select Groups for *i-Ready Inform* for Reading in English

Hispanic

Grade	*N*	Rel.	SEM
K	437,192	0.95	9.3
1	510,666	0.97	9.4
2	556,065	0.97	10.3
3	603,129	0.97	10
4	589,922	0.97	10.3
5	602,707	0.97	10.5
6	495,595	0.97	10.6
7	452,575	0.97	10.7
8	448,864	0.97	10.7

9	81,798	0.98	10.8
10	63,823	0.97	10.9
11	46,820	0.98	10.9
12	34,064	0.98	11

Native Hawaiian or Pacific Islander

Grade	*N*	Rel.	SEM
K	9,063	0.94	9.3
1	10,581	0.97	9.4
2	12,165	0.97	10.4
3	12,285	0.97	10
4	12,410	0.97	10.3
5	12,500	0.97	10.5
6	11,161	0.97	10.6
7	7,714	0.97	10.7
8	8,161	0.97	10.7
9	2,378	0.97	10.8
10	1,897	0.97	10.9
11	1,605	0.97	11
12	1,082	0.97	11

Other

Grade	*N*	Rel.	SEM
K	640	0.95	9.3
1	893	0.96	9.5
2	1,206	0.97	10.6
3	1,657	0.97	10.2
4	1,620	0.96	10.5
5	1,583	0.96	10.7
6	1,413	0.96	10.8
7	1,206	0.96	10.8
8	1,245	0.96	10.9
9	125	0.97	10.8
10	**	**	**
11	**	**	**
12	**	**	**

Note: Rel. = Reliability

Table 6.15. Marginal Reliability and Standard Error of Measurement for Select Groups for *i-Ready Inform* for Reading in English

Two or More Races

Grade	*N*	Rel.	SEM
K	70,714	0.95	9.3
1	77,833	0.97	9.5
2	82,594	0.97	10.6
3	81,410	0.97	10.2
4	76,866	0.97	10.5
5	73,949	0.97	10.6
6	57,842	0.97	10.7
7	48,991	0.97	10.8
8	46,853	0.97	10.9
9	9,099	0.97	10.9
10	5,702	0.97	11.1
11	3,221	0.97	11.2
12	2,025	0.98	11.2

White

Grade	*N*	Rel.	SEM
K	522,918	0.95	9.3
1	613,700	0.96	9.6
2	656,197	0.96	10.7
3	678,494	0.97	10.3
4	660,451	0.96	10.6
5	665,086	0.96	10.7
6	556,596	0.96	10.8
7	502,260	0.96	10.8
8	495,236	0.96	10.9
9	90,027	0.97	11
10	60,048	0.97	11.1
11	37,948	0.98	11.2
12	24,921	0.98	11.2

Note: Rel. = Reliability

Table 6.16. Marginal Reliability and Standard Error of Measurement for Select Groups for *i-Ready Inform* for Reading in English

Male

Grade	*N*	Rel.	SEM
K	859,906	0.95	9.3

1	986,286	0.97	9.5
2	1,070,628	0.97	10.5
3	1,154,716	0.97	10.1
4	1,126,846	0.97	10.4
5	1,138,014	0.97	10.6
6	930,631	0.97	10.6
7	847,720	0.97	10.7
8	837,451	0.97	10.8
9	151,641	0.98	10.9
10	103,641	0.98	11
11	71,994	0.98	11.1
12	48,140	0.98	11.1

Female

Grade	*N*	Rel.	SEM
K	842,696	0.95	9.3
1	969,784	0.97	9.5
2	1,045,928	0.97	10.6
3	1,113,682	0.97	10.2
4	1,088,401	0.97	10.5
5	1,099,893	0.97	10.6
6	901,784	0.97	10.7
7	823,636	0.97	10.8
8	812,477	0.97	10.8
9	140,300	0.97	10.9
10	97,141	0.97	11
11	65,853	0.97	11.1
12	44,122	0.98	11.2

Note: Rel. = Reliability

Table 6.17. Marginal Reliability and Standard Error of Measurement for Select Groups for *i-Ready Inform* for Mathematics in Spanish

Economically Disadvantaged

Grade	*N*	Rel.	SEM
K	20,006	0.94	6.5
1	21,123	0.95	6.4
2	19,730	0.95	6.4
3	7,665	0.96	6.4
4	5,977	0.96	6.4

5	5,425	0.96	6.4
6	3,790	0.96	6.4
7	3,107	0.96	6.4
8	2,997	0.96	6.5
9	533	0.97	6.4
10	453	0.98	6.4
11	278	0.98	6.5
12	155	0.98	6.4

English Learners

Grade	*N*	Rel.	SEM
K	39,749	0.93	6.4
1	37,904	0.95	6.4
2	36,230	0.95	6.4
3	20,313	0.96	6.4
4	16,392	0.96	6.4
5	14,558	0.96	6.4
6	11,099	0.96	6.4
7	10,149	0.96	6.5
8	9,616	0.96	6.5
9	4,237	0.97	6.4
10	2,671	0.97	6.4
11	1,075	0.97	6.5
12	419	0.97	6.4

Special Education

Grade	*N*	Rel.	SEM
K	4,667	0.94	6.5
1	4,937	0.96	6.5
2	4,651	0.96	6.4
3	2,217	0.97	6.4
4	1,517	0.97	6.4
5	1,274	0.97	6.4
6	508	0.97	6.5
7	464	0.97	6.4
8	377	0.97	6.5
9	**	**	**
10	**	**	**

11	**	**	**
12	**	**	**

Note: ** = not reported due to a sample size of less than 100 students; Rel. = Reliability

Table 6.18. Marginal Reliability and Standard Error of Measurement for Select Groups for *i-Ready Inform* for Mathematics in Spanish

American Indian or Alaskan

Grade	*N*	Rel.	SEM
K	161	0.93	6.5
1	139	0.96	6.4
2	140	0.95	6.4
3	**	**	**
4	**	**	**
5	**	**	**
6	**	**	**
7	**	**	**
8	**	**	**
9	**	**	**
10	**	**	**
11	**	**	**
12	**	**	**

Asian

Grade	*N*	Rel.	SEM
K	265	0.94	6.5
1	325	0.96	6.4
2	338	0.95	6.4
3	140	0.95	6.4
4	100	0.95	6.4
5	**	**	**
6	**	**	**
7	**	**	**
8	**	**	**
9	**	**	**
10	**	**	**
11	**	**	**
12	**	**	**

Black or African American

Grade	*N*	Rel.	SEM
K	1,061	0.94	6.5
1	1,281	0.95	6.5
2	890	0.96	6.4
3	406	0.96	6.4
4	367	0.97	6.4
5	380	0.97	6.4
6	148	0.98	6.5
7	**	**	**
8	109	0.97	6.4
9	**	**	**
10	**	**	**
11	**	**	**
12	**	**	**

Note: Rel. = Reliability; ** = not reported due to a sample size of less than 100 students

Table 6.19. Marginal Reliability and Standard Error of Measurement for Select Groups for *i-Ready Inform* for Mathematics in Spanish

Hispanic

Grade	*N*	Rel.	SEM
K	61,562	0.93	6.4
1	62,561	0.95	6.4
2	59,226	0.95	6.4
3	36,543	0.96	6.4
4	27,506	0.96	6.4
5	24,533	0.96	6.4
6	16,135	0.96	6.4
7	13,940	0.96	6.5
8	13,309	0.96	6.5
9	4,679	0.97	6.4
10	2,903	0.97	6.4
11	1,207	0.97	6.5
12	492	0.98	6.4

Two or More Races

Grade	*N*	Rel.	SEM
K	685	0.95	6.5
1	856	0.96	6.5
2	651	0.96	6.4

3	357	0.95	6.4
4	181	0.96	6.4
5	128	0.97	6.4
6	**	**	**
7	**	**	**
8	**	**	**
9	**	**	**
10	**	**	**
11	**	**	**
12	**	**	**

White

Grade	*N*	Rel.	SEM
K	3,383	0.95	6.5
1	3,643	0.96	6.5
2	3,425	0.96	6.4
3	1,745	0.96	6.4
4	1,232	0.97	6.4
5	930	0.97	6.4
6	285	0.97	6.4
7	155	0.97	6.5
8	133	0.97	6.5
9	**	**	**
10	**	**	**
11	**	**	**
12	**	**	**

Note: Rel. = Reliability; ** = not reported due to a sample size of less than 100

Table 6.20. Marginal Reliability and Standard Error of Measurement for Select Groups for *i-Ready Inform* for Mathematics in Spanish

Male

Grade	*N*	Rel.	SEM
K	36,817	0.94	6.5
1	37,647	0.96	6.4
2	35,080	0.96	6.4
3	22,675	0.96	6.4
4	17,505	0.97	6.4
5	15,505	0.97	6.4
6	10,938	0.97	6.4

7	9,624	0.97	6.5
8	9,251	0.96	6.5
9	2,839	0.97	6.5
10	1,787	0.97	6.4
11	680	0.97	6.5
12	265	0.98	6.4

Female

Grade	*N*	Rel.	SEM
K	37,941	0.93	6.4
1	37,986	0.95	6.4
2	36,211	0.95	6.4
3	22,909	0.96	6.4
4	17,483	0.96	6.4
5	15,624	0.96	6.4
6	10,229	0.96	6.4
7	8,792	0.96	6.5
8	8,396	0.96	6.5
9	2,168	0.97	6.4
10	1,302	0.97	6.4
11	566	0.97	6.4
12	240	0.97	6.4

Note: Rel. = Reliability

6.5 Reliability of Classification

Classification accuracy and consistency are metrics that describe the reliability of classification into two or more placement levels, using one or more thresholds. The main factors that impact classification accuracy and consistency are the number and location of the thresholds, the student proficiency distribution, and the CSEM of the students' proficiency estimates (which is in turn affected by the number of items).

Classification accuracy is the degree to which examinees' assigned placement levels match their *true* placement levels. Classification consistency is the degree to which examinees fall into the same placement level on two separate administrations, regardless of whether or not it is the true placement level. Both quantities are expressed as proportions. Because classification accuracy reflects only one source of error (i.e., a single administration) and classification consistency reflects two sources of error (i.e., two administrations), classification consistency will never exceed classification accuracy.

Rudner's (2000, 2005) method is used to estimate classification accuracy. The method utilizes several key results from IRT, namely, that $\hat{\theta}_i$ follows a normal distribution centered at θ_i (i.e., $\hat{\theta}_i$ is an unbiased estimator of θ_i) with a standard deviation equal to $CSEM_{\hat{\theta}_i}$. These results allow us to construct a normal distribution for

each student based on their $\hat{\theta}_i$ and $CSEM_{\hat{\theta}_i}$. The placement-level thresholds can then be overlayed onto that normal distribution, and the area underneath the normal curve can be computed for each placement level to calculate a probability by placement-level vector for each student. We assume the *true* placement level to be the one in which the student's actual score is located. The proportion under the normal curve associated with this placement level is the classification accuracy for examinee i. These examinee-level classification accuracy statistics can then be averaged across all students, resulting in test-level classification accuracy.

Wyse and Hao (2012) extended Rudner's (2000, 2005) method to compute classification consistency from the probability by placement-level vector described above. Since the elements of these vectors represent the probability that a student is classified in each placement level, each element can be squared, representing the probability that they would be classified in the same placement level twice. Since classification consistency does not consider the correctness of classification but only the consistency of classification, these squared terms can be summed within each student to calculate a student-level classification consistency metric. These can then be averaged across all students for the test-level classification consistency statistic.

Classification accuracy and consistency results are displayed in Table 6.21–Table 6.23 for *i-Ready Inform* for Reading and for Mathematics in English and *i-Ready Inform* for Mathematics in Spanish, respectively. Note that the placement levels analyzed here are based on those used to assign growth measures (see Chapter 9). The columns labeled "Two Levels Below," "One Level Below," "Early On Grade," and "Mid On Grade" refer to the cuts associated with attaining those placement levels. The values in each of those columns represent the classification accuracy or consistency of students relative to *only* that cut score. For example, when considering the "One Level Below" columns, the classification results are produced by partitioning students into two categories: two or more levels below versus one level below and above.

Note that the "Three or More Levels Below" placement level does not exist for Grades K–2, and that the "Two Levels Below" placement level does not exist for Grade K. Because of this, it is impossible to categorize students into two groups using the "Two Levels Below" and "One Level Below" cuts for Grades K–2 and Grade K, respectively. Therefore, there are no entries in the table for these grade-placement level combinations. The "All" columns refer to using all four cut scores, which divide students into up to five groups.

Classification accuracy and consistency values associated with a single cut score are .90 or above, with many approaching 1, and classification consistency is equal to or slightly lower than classification accuracy, as expected. Classification accuracy results for all cuts are generally .80 or above, with the exception of Grades 6–8 in Reading, which are just below .80. Classification consistency results are .70 or greater.

Table 6.21. Classification Accuracy and Consistency for *i-Ready Inform* for Mathematics in English

Classification Accuracy

Grade	*N*	Two Levels Below	One Level Below	Early On Grade	Mid On Grade	All
K	2,281,818	-	-	0.93	0.93	0.87
1	2,600,489	-	0.97	0.94	0.96	0.88
2	2,774,663	-	0.95	0.94	0.95	0.84
3	2,870,693	0.98	0.96	0.94	0.96	0.83

4	2,826,144	0.97	0.95	0.94	0.96	0.82
5	2,854,279	0.97	0.95	0.94	0.96	0.82
6	2,352,756	0.96	0.95	0.94	0.96	0.81
7	2,081,161	0.96	0.95	0.95	0.98	0.84
8	1,961,463	0.96	0.94	0.95	0.98	0.83
9	283,093	0.95	0.95	0.96	0.99	0.85
10	177,740	0.95	0.96	0.98	1.00	0.89
11	118,950	0.96	0.97	0.99	1.00	0.91
12	68,945	0.97	0.98	0.99	1.00	0.94

Classification Consistency

Grade	*N*	Two Levels Below	One Level Below	Early On Grade	Mid On Grade	All
K	2,281,818	—	-	0.91	0.91	0.83
1	2,600,489	—	0.96	0.92	0.94	0.83
2	2,774,663	—	0.93	0.92	0.93	0.78
3	2,870,693	0.97	0.94	0.91	0.94	0.76
4	2,826,144	0.96	0.93	0.91	0.94	0.74
5	2,854,279	0.95	0.93	0.92	0.94	0.75
6	2,352,756	0.94	0.93	0.92	0.95	0.74
7	2,081,161	0.94	0.93	0.93	0.97	0.77
8	1,961,463	0.94	0.92	0.94	0.97	0.77
9	283,093	0.93	0.93	0.94	0.98	0.80
10	177,740	0.93	0.94	0.98	0.99	0.84
11	118,950	0.94	0.96	0.98	0.99	0.88
12	68,945	0.96	0.97	0.98	1.00	0.92

Table 6.22. Classification Accuracy and Consistency for *i-Ready Inform* for Reading in English

Classification Accuracy

Grade	*N*	Two Levels Below	One Level Below	Early On Grade	Mid On Grade	All
K	1,878,648	—	—	0.93	0.95	0.88
1	2,175,773	—	0.98	0.96	0.96	0.90
2	2,328,438	—	0.96	0.95	0.95	0.85
3	2,486,736	0.98	0.96	0.95	0.95	0.84
4	2,452,357	0.97	0.97	0.94	0.94	0.83
5	2,469,573	0.98	0.95	0.94	0.95	0.81
6	2,023,140	0.96	0.94	0.94	0.94	0.79

7	1,839,402	0.95	0.94	0.94	0.94	0.78
8	1,808,160	0.95	0.95	0.93	0.94	0.78
9	313,476	0.95	0.94	0.95	0.96	0.81
10	214,492	0.95	0.94	0.95	0.96	0.80
11	146,513	0.94	0.94	0.95	0.97	0.81
12	98,845	0.94	0.95	0.95	0.97	0.83

Classification Consistency

Grade	*N*	Two Levels Below	One Level Below	Early On Grade	Mid On Grade	All
K	1,878,648	—	—	0.90	0.93	0.83
1	2,175,773	—	0.97	0.94	0.95	0.86
2	2,328,438	—	0.94	0.93	0.93	0.80
3	2,486,736	0.97	0.95	0.93	0.93	0.77
4	2,452,357	0.96	0.95	0.92	0.92	0.76
5	2,469,573	0.97	0.93	0.91	0.93	0.74
6	2,023,140	0.94	0.92	0.91	0.92	0.71
7	1,839,402	0.93	0.92	0.91	0.92	0.70
8	1,808,160	0.93	0.92	0.90	0.92	0.71
9	313,476	0.93	0.92	0.92	0.95	0.75
10	214,492	0.93	0.91	0.93	0.94	0.73
11	146,513	0.92	0.92	0.93	0.95	0.75
12	98,845	0.92	0.92	0.93	0.96	0.78

Table 6.23. Classification Accuracy and Consistency for *i-Ready Inform* for Mathematics in Spanish

Classification Accuracy

Grade	*N*	Two Levels Below	One Level Below	Early On Grade	Mid On Grade	All
K	79,291	—	—	0.94	0.95	0.90
1	79,865	—	0.95	0.96	0.98	0.88
2	75,411	—	0.94	0.96	0.97	0.88
3	48,678	0.96	0.95	0.96	0.98	0.85
4	37,293	0.96	0.94	0.96	0.98	0.84
5	32,963	0.94	0.94	0.97	0.99	0.84
6	22,866	0.94	0.95	0.98	0.99	0.87
7	19,585	0.95	0.97	0.99	1.00	0.91
8	18,753	0.96	0.97	0.99	1.00	0.93
9	5,137	0.97	0.98	0.99	1.00	0.95

10	3,165	0.97	0.99	1.00	1.00	0.96
11	1,300	0.99	1.00	1.00	1.00	0.98
12	518	1.00	1.00	1.00	0.99	0.99

Classification Consistency

Grade	*N*	Two Levels Below	One Level Below	Early On Grade	Mid On Grade	All
K	79,291	—	—	0.92	0.93	0.86
1	79,865	—	0.93	0.94	0.97	0.84
2	75,411	—	0.91	0.95	0.96	0.83
3	48,678	0.94	0.93	0.94	0.97	0.78
4	37,293	0.94	0.91	0.94	0.98	0.77
5	32,963	0.91	0.92	0.96	0.98	0.78
6	22,866	0.92	0.93	0.98	0.99	0.83
7	19,585	0.93	0.96	0.99	1.00	0.87
8	18,753	0.95	0.96	0.99	1.00	0.90
9	5,137	0.96	0.98	0.99	1.00	0.93
10	3,165	0.96	0.98	1.00	1.00	0.95
11	1,300	0.98	1.00	1.00	1.00	0.97
12	518	0.99	1.00	1.00	0.99	0.98

6.6 Discussion

The results in this chapter demonstrate that the depth of the item bank, the effectiveness of the item selection algorithm, and the design of the test flows work together to produce reliable estimates of the overall score across a wide range of the proficiency distribution. Classification accuracy and consistency coefficients demonstrate a similar result. Although demographic data is limited, the overall score reliability at the group level demonstrates that scores for each of the groups are generally estimated with a similar degree of precision compared to the entire population. Domain scores, although less reliable because they are based on fewer items, generally have reliability coefficients above .7, but they are often above .8.

Chapter 7: Setting Standards

7.1 Chapter Summary

This chapter outlines the standard-setting methods and statistical techniques used to set the cut scores for *i-Ready Inform* assessments. Two standard-setting methods were employed to set overall cut scores: the contrasting groups method was used for Grades K–8, and the modified bookmark method was used for high school grades. A national sample of educators contributed to the cut score determination by participating in standard-setting workshops. Curriculum Associates staff performed a variety of critical functions, such as using statistical methods to establish domain cut scores and facilitating vertical articulation across grade levels.

7.2 Introduction

7.2.1 The Basis for Standard Setting

The *i-Ready Inform* assessment is a valuable tool for educators because it provides criterion-referenced placements that form the foundation for many components of our reports. These placement levels are aligned to grade-level expectations. Through a process known as standard setting, we establish a relationship between those expectations and student achievement on the test. During standard setting, educators participate in a systematic process that results in recommended cut scores that map placement levels onto the score scale. This chapter summarizes the standard-setting procedures and subsequent statistical methods used to set cut scores for the *i-Ready Inform* assessments.

The Standards for Educational and Psychological Testing (AERA et al., 2014) include guidelines for setting cut scores and documenting procedural evidence to support the validity argument for score interpretations and uses. Specifically, Standard 5.21 states: "When proposed score interpretations involve one or more cut scores, the rationale and procedures used for establishing cut scores should be documented clearly." For a criterion-referenced assessment such as *i-Ready Inform*, the procedure that links descriptions of student performance to the test's score scale is vital to the overall validity argument and to ensure that placement levels are not arbitrary (Kane, 2013). Thus, cut scores must be determined using a suitable methodology. Note that summative assessment programs use similar procedures to determine performance standards and their associated cut scores.

There were two main guiding principles in our standard-setting procedures:

1. The cut scores must correspond to the specified performance standard, which in our case is the placement level, so that students meeting the expectations of the placement level receive scores within its specified score range.
2. A cut score must be reasonable and align with the purpose(s) of the placement decisions made.

7.2.2 Overview of Methods

Several methodologies are available to set performance standards that are compliant with the guidelines set forth by the Standards for Educational and Psychological Testing (AERA et al., 2014). These methods are often

differentiated by the tasks assigned to panelists during standard-setting workshops. The choice of methodology is typically influenced by the types of items and the characteristics of the assessment as well as how the resulting evidence supports the overall validity argument (Hambleton & Pitoniak, 2006). This chapter offers procedural evidence from the standard setting of the *i-Ready Inform* assessment in support of the overall validity argument.

There are five key placement levels—both at the overall and domain levels—for each subject and grade level: Below Grade Level, Early Grade Level, Mid Grade Level, Late Grade Level, and Above Grade Level. As the labels indicate, Below and Above Grade Level placements are considered *off-grade level placements*, while Early, Mid, and Late Grade Level placements are considered *on-grade level placements*. Note that Below and Above Grade Level placements can be further specified with the number of grade levels below or above. The Early and Mid cuts were set using industry-standard methodology involving panelist judgment. The Late cuts were set by Curriculum Associates staff using content expert judgment and statistical moderation. The Above cuts and all individual cuts in the Below Grade Level and Above Grade Level categories were determined based on heuristics which relate to adjacent grade levels and (for Below Grade Level cuts) account for measurement error of the test.

The Early and Mid cut scores were set through a person-centered standard-setting method called contrasting groups (Livingston and Zieky, 1982) for Grades K–8 and through a modified bookmark standard-setting approach (Lewis et al., 1996) for Grades 9–12. In the contrasting groups method, teachers categorize their students into levels based on their holistic understanding of what each student knows and can do. Cut scores are set by comparing test performance for the panelists' assigned student categorizations and identifying points that best discriminate between students of different proficiency levels. Note that this method relies on teachers' comprehensive understanding of their students' levels of proficiencies. The modified bookmark method was used for Grades 9–12 because there was not an existing sample of high school students suitable for the contrasting groups method. Rather than relying on teachers' judgments of their students' levels of proficiencies, the modified bookmark method relies on panelists' ability to judge the difficulty of an ordered set of items with respect to the performance standards. Cut scores for all placement levels at the overall and domain level can be found in the *i-Ready Diagnostic* K–12 Scale Score Placement Tables document.

7.3 Placement Levels

7.3.1 Goals of Setting Placement Levels

One critical step in developing an assessment program is setting the level of student proficiency that is minimally sufficient for that student to be classified into each placement level. Such classifications make a many-point score scale more readily interpretable by breaking it up into a small number of categories that map regions of the score scale to a meaningful description of performance.

In the case of the *i-Ready Inform* assessments, the placement levels are characterized by carefully articulated descriptions of the performance necessary for that region of the score scale. These descriptions are referred to as PLDs and are developed by teachers, subject-matter experts, and other assessment professionals, and they are the basis upon which these assessments can be considered to be criterion referenced.

7.3.2 Placement-Level Descriptors

7.3.2.1 Policy-Level PLDs

The policy-level PLDs are broad descriptive statements that generally describe the progression of skills and knowledge students gain through learning (Egan et al., 2012). These statements succinctly describe student performance relative to the expectations set forth in CCRS for a grade level and subject. Policy-level PLDs for *i-Ready Inform* also include a recommendation for the level of instruction from which students who have been assigned that placement level would most benefit. The five policy-level PLDs are presented in Table 7.1 below. Note that Table 7.1 is identical to Table 2.1 and is reproduced here for convenience.

Table 7.1. Policy-Level PLDs

Placement Level	Description
Below Grade Level	Remediation focused on below-grade level material is recommended to help fill in gaps in students' foundational knowledge. Students in this level are not close to meeting the expectations of CCRS for their grade level.
Early On Grade Level	Students in this level will benefit from on-grade level instruction to help them meet the expectations of CCRS for their grade level. Students in the early on grade level have only partially met these grade-level expectations.
Mid Grade Level	Students in this level will benefit from instruction in late on-grade level topics. These students have met the minimum requirements for the expectations of CCRS in their grade level.
Late Grade Level	Students in this level will benefit from late on-grade level enrichment and will be ready for instruction focused on topics typically covered in the beginning of the subsequent grade level. Students in the late grade level have successfully met or surpassed the grade-level expectations of CCRS.
Above Grade Level	Students in this category will benefit from above-grade level instruction. Students in the above grade level have successfully met or surpassed all the expectations of CCRS for their grade level as well as some expectations from subsequent grade levels.

7.3.2.2 Standard-Setting PLDs

To help inform instructional recommendations and placement decisions, SMEs and assessment experts developed standard-setting PLDs, which provide grade- and domain-specific descriptions of the knowledge and skills expected of students in each placement level. The development of the standard-setting PLDs was informed by a variety of factors, including various sets of CCRS and the *i-Ready* indicators. They also developed standard-setting PLD summaries for each grade and subject with overall descriptions of what students should know and be able to do across all domains in that grade and subject to be classified into a particular placement level.

The standard-setting PLDs were then reviewed and refined by Sireci Psychometric Services as well as by a panel of 22 educators and SMEs. A typical panelist had more than 15 years of Grades K–8 teaching experience. To illustrate, examples of these PLDs are provided in Table 7.2 and Table 7.3 for mathematics and reading, respectively. Each table provides a domain-specific and a summary standard-setting PLD for Grade 3. Note that the standard-setting PLDs have since been retired and replaced with the range PLDs discussed in Chapter 2, which provide the basis for ongoing item and test development.

Table 7.2. Example Standard-Setting PLDs for Mathematics: Grade 3

PLD Type	Early Placement Level	Mid Placement Level
Number and Operations	Early Grade 3 students can use concepts of place value to round to the nearest multiple of ten, to add and subtract within 1,000 and to multiply a one-digit whole number by a multiple of ten. They understand a fraction represents a part of a whole and can represent a unit fraction using visual models and symbols, and identify them on a number line. They are able to compare unit fractions in relation to a whole.	In addition to the skills listed for Early Grade 3 students, Mid Grade 3 students are able to fluently add and subtract within 1,000 and can apply concepts of multiplication of whole numbers to division. They demonstrate competency in all single-digit multiplication facts. They interpret and represent fractions, including those other than unit fractions, using a number line; apply concepts of fractions to real world contexts; model equivalencies in diagrams and number lines; and use reasoning to identify simple equivalent fractions and compare two fractions that have the same numerator or the same denominator.
Summary	Early Grade 3 students utilize strategies, visual representations, properties of operations and/or place value concepts to compute, interpret meaning of operations, and to solve simple problems. They can identify and compare unit fractions in relation to a whole. They are beginning to apply reasoning skills to identify geometric figures, and can interpret simple graphs and solve basic problems involving measurement.	Mid Grade 3 students can apply strategies to use properties of operations and relationships among operations to fluently make accurate computations, solve problems, and justify solutions. They use reasoning to compare fractions, analyze geometric figures, solve measurement problems, and interpret data and graphs.

Table 7.3. Example Standard-Setting PLDs for Reading: Grade 3

PLD Type	Early Placement Level	Mid Placement Level
Vocabulary	Early Grade 3 students demonstrate understanding of prefixes, suffixes, inflectional endings, base words, and root words and apply that understanding to interpret unfamiliar vocabulary. They demonstrate basic understanding of homophones and rhyme. Independently they demonstrate knowledge of below-grade words in multiple texts and across content areas. With support, they demonstrate knowledge of on-grade vocabulary. With support, they can understand literal and nonliteral	Building upon the skills listed for Early Grade 3 students, Mid Grade 3 students demonstrate good understanding of meaningful word parts, homophones, and rhyme. They consistently demonstrate knowledge of on-grade vocabulary and sometimes apply knowledge of on-grade vocabulary in multiple texts and across content areas. They can understand literal and nonliteral word meanings when presented in context. They can distinguish shades of meaning among related words that describe states of mind or degrees of

	word meanings when presented in context.	certainty (e.g., knew, believed, suspected, heard, wondered).
Summary	Early Grade 3 students understand most consonant and vowel sounds singly and in combination and can apply this knowledge to decode common, on-grade words. They demonstrate understanding of word parts and categories and can understand below-grade level words independently and on-grade level words with support. They demonstrate basic understanding of on-grade level texts and are able to make inferences and connections in below-level and early on-grade level texts.	Mid Grade 3 students have a good understanding of consonant and vowel sounds singly and in combination and can apply that understanding to decode a range of on-grade words. They demonstrate understanding of word parts and categories and can understand many on-grade words, including those with irregular spelling. They demonstrate good understanding of on-grade texts and can make inferences and connections independently. They understand point of view in both on-grade informational and literary texts and key details related to literary elements in on-grade literary texts.

7.4 Methods and Procedures: Grades K–8

7.4.1 Contrasting Groups Standard-Setting Method

Contrasting groups is a person-centered standard-setting method in which students are categorized into groups by an educator who has experience with students' level of proficiency (Livingston & Zieky, 1982). The panelists, who are usually teachers, make judgments based on their comprehensive knowledge about each student with respect to things like class performance, class assignments, and homework and without regard to the score on the test of interest. Typically, panelists do not see how students performed on individual items of the test, nor do they use the results of the tests in their judgments about students.

Panelists read through a series of placement-level or category descriptions and place each student into the category they feel most appropriately represents the student's competency level (Cizek & Bunch, 2007). Each cut score is determined via logistic regression as the test score that best differentiates between students classified in a given placement level or higher and students classified in a lower placement level.

7.4.2 Panelists

The Contrasting groups method requires that (1) there are enough students classified into each placement level by their teachers to fit the statistical models necessary to accurately determine the cut score and (2) that the overall score distribution reflected a broad range of proficiency. To ensure these conditions were met, analysts carefully considered the number of teachers participating as well as how many students were included in the evaluation. The sampling strategy began with finding teachers who had sufficient numbers of students who tested within a time window around the planned standard-setting meetings. Those teachers were asked about their interest and availability. Then, the student score distributions of those teachers who agreed to participate were examined to ensure that the distribution was as expected, and specifically, that the

students to be evaluated would represent a broad range of proficiency levels. Teachers were recruited through school administrator nomination, peer recommendation, or self-nomination.

More than 100 teachers from 24 states who used *i-Ready* with their students in Grades K–8 during the 2013–2014 school year were recruited for the standard-setting panel. Table 7.4 displays the number of teachers by subject and grade band. Note that many Grades K–5 teachers participated in both the reading and mathematics panels. Approximately 5,000 total students were represented by the teachers serving on the standard-setting panel. These students represented a broad range of proficiency levels based on *i-Ready Inform*.

Table 7.4. Number of Teachers in Each Panel by Subject and Grade Band

Subject	Grade Band	Number of Teachers
Mathematics	K–2	38
Mathematics	3–5	48
Mathematics	6–8	15
Reading	K–2	39
Reading	3–5	48
Reading	6–8	14

7.4.3 Process

Curriculum Associates conducted the standard setting via a series of virtual meetings. In the initial meeting, panelists learned about the contrasting groups method and were provided instructions for making their judgments. In a second meeting, panelists learned about PLDs, which were the basis upon which students were categorized into the placement levels.

During the first round of judgments, panelists assigned their students to placement levels based on the PLDs. Panelists were given approximately 10 days to complete their first round of judgments. Panelists each received an Excel® document containing a row for each of their students. Panelists were instructed to indicate which level they would place each student into (i.e., 1 = Below, 2 = Early, 3 = Mid) as well as how confident they were about the placement (i.e., 1 = Not Confident, 2 = Somewhat Confident, 3 = Very Confident). Panelists had the option to fill out an additional column explaining their rating or providing more information about their rating.

After panelists submitted the first round of judgments, students' scores on *i-Ready* were analyzed to determine draft cut scores. Panelists' judgments were merged with the most recent *i-Ready Inform* scores for each student. Logistic regression models were used to find the cut scores that would maximize agreement with the teacher judgments and best distinguish placement levels from one another. Impact data showing the percentages of students that would be classified into each placement level based on the draft cut scores were generated for each grade and subject and presented to panelists in a third virtual meeting. Panelists reviewed their initial cut scores and the impact data and then were allowed to revise their assignment of students into placement levels. After panelists made their second round of judgments, they submitted Excel documents containing updated student placements to Curriculum Associates for analysis. Logistic regression models were again used to determine the revised draft cut scores for each grade and subject. The recommended cut scores were later reviewed, possibly adjusted, and finalized as described at the end of this chapter.

Microsoft Excel® is a registered trademark of Microsoft Corporation.

7.5 Methods and Procedures: Grades 9–12

7.5.1 Modified Bookmark Method

Cut scores for *i-Ready Inform* in Grades 9–12 were set using a modified bookmark standard-setting approach. The modified bookmark method is a popular standard-setting method. In contrast to the person-centered nature of the contrasting groups method used for Grades K–8, the modified bookmark method is an item-centered method (Lewis et al., 1996). The bookmark method requires item difficulty data to create an ordered item booklet (OIB) in which a set of items is arranged in order of ascending difficulty. The OIB is then used by standard-setting panelists to identify the dividing lines between placement levels based on item difficulty.

7.5.2 Panelists

Two panels of teachers—one for mathematics and one for reading—convened to recommend cut scores. In the modified bookmark approach, the role of the panelist is to read the items in order and identify the location (item) in the OIB at which a test-taker just barely in a given placement level has a 67 percent chance of answering the item correctly (Cizek & Bunch, 2007).

7.5.3 Process

Constructing the OIB was the initial step in the process. Items selected for the OIBs were exemplars of—and matched the difficulty level of—each *i-Ready* indicator (see Chapter 2 for a description of the indicator framework).

Panelists then reviewed the OIBs and identified cuts representing the placement level of students ready for grade-level content—the Early cut—and representing the placement level of students performing at a proficient level—the Mid Grade Level cut. Following this procedure, an impact analysis was conducted in which teachers reviewed the set of Can Do statements that would have resulted from those cut scores.

7.6 Extensions

7.6.1 Late Grade Cut Scores

After the Early and Mid Grade Level cut scores were determined, the Late cut scores were set by Curriculum Associates after taking into consideration factors such as:

- Subject-matter expertise
- Connections to norms and Lexiles
- Connections to percent mastery of indicators

SMEs examined the skills a student classified as Late could likely do and ensured that students who placed into the Late placement level had a large percentage of the indicators listed in the Can Do section of the student profile report. Specifically, students with Late placements would be said to be likely to be successful in answering questions aligned to the vast majority of the indicators for that grade level.

7.6.2 Above and Below Grade Cut Scores

Once the Late cut scores were identified, the Above and Below Grade cuts were systematically set. For each grade and subject, the Above Grade range is based on the Late Grade range for one grade one level above. For each grade and subject, the Below Grade range was set in reference to the Early On Grade range for the grade below. There are no limits on Below Grade placements; students can place into any grade below their chronological grade.

7.6.3 Domain Cut Scores

Statistical analyses were used to determine cut scores for the domain scores. Using the recommended placement-level cut scores, we fit a series of logistic regression models to predict classification into placement levels based on student scores in each domain. For example, for Grade 4 mathematics, four different logistic regression models were fit—one each for Algebra, Number and Operations, Geometry, and Measurement and Data. Finally, when necessary, the domain cut scores were adjusted so that they deviated by no more than 5 points from the overall cut score. However, subsequent adjustments due to calibration work in 2016 resulted in some small adjustments to the placements so that in a few cases the differences are now slightly more than 5 points.

For reading, the Comprehension: Overall cut score was added in the 2022–2023 school year and is the average of the two comprehension cut scores (Comprehension: Informational Text and Comprehension: Literature).

7.6.4 Finalizing the Cut Scores

An internal Curriculum Associates team determined the final cut scores by considering the following additional factors:

- Percentages of a national sample that placed above or below the cuts
- Statistical relationship with New York State Common Core Assessment results
- Progression of the cut scores across grades
- Number and percentage of indicators mastered for each proposed placement level
- Changes from current cut scores

A vertical articulation was carried out to ensure coordination of cut scores across grade levels.

Chapter 8: Norms

8.1 Chapter Summary

This chapter describes the methods Curriculum Associates used to develop norms for *i-Ready Inform*, including the stratified sampling approach, key characteristics of the norm samples, and the process of developing percentile ranks associated with each scale score value and calculated for each testing window and grade combination for both reading and mathematics. The chapter describes how norms were updated for 2024–2025 and how they are monitored over time. Finally, the resulting 2024 National Norms are presented in tables along with guidance for their interpretation and use, in accordance with best industry practices (AERA et al., 2014).

8.2 Introduction

The *i-Ready Inform* norms provide a way for educators to evaluate how their students are performing relative to other students across the country. Normative scores, together with other types of scores such as *i-Ready Inform* criterion-referenced scores (i.e., grade-level placements for reading and mathematics and for domains), help educators gain a more complete picture of student performance. Whereas *i-Ready*'s normative scores communicate how students perform compared to other students, criterion-referenced scores communicate what students know and can do against grade-level standards.

Norms are expressed through percentile ranks, which relate each student's performance to a nationally representative sample of students in the same grade level who took *i-Ready Inform* at the same time of year (e.g., fall). Therefore, norm scores indicate how students are performing relative to their peers nationwide. For example, if a student's percentile rank is 90, this means the student scored better than or equal to 90 percent of their national peers from the same grade level at the same time of year.

8.3 Norms Development

The 2024 National Norms were developed based on students who took *i-Ready Inform* during the 2022–2023 school year. To ensure that the norms for each grade closely reflect the national student population at that grade and time of year, students in the norming sample for Grades K–8 were administered *i-Ready* as a regular academic assessment tool. Students generally take *i-Ready Inform* three times during the year (i.e., fall, winter, and spring) and could be included in sampling if they tested during any of those windows. Students in the norming sample for Grades 9 and 10 were administered *i-Ready* as a regular academic assessment tool in the fall, which is the most common time of year for high schools to use *i-Ready*. Included schools were diverse geographically and in terms of demographics and socioeconomic status, such that the samples approximated the 2020–2021 national population according to the National Center for Educational Statistics (NCES). In sum, norms were designed to be available for Grades K–8 in fall, winter, and spring and for Grades 9 and 10 in fall.

The *i-Ready Inform* assessment has three norming windows: fall, winter, and spring. The normative percentile rank associated with a given *i-Ready Inform* scale score differs across norming testing windows. The national norming windows for the *i-Ready Inform* assessment at Grades K–8 are:

- **Fall:** August 1 to November 15
- **Winter:** November 16 to March 1
- **Spring:** March 2 to June 15

The percentile ranks associated with each scale score value were calculated for each testing window, grade, and subject combination for Grades K–8 and for each grade and subject combination for the fall testing window for Grades 9 and 10. The national norms are updated every three to five years but this timing is also dependent on other relevant changes to the assessment program or student population.

8.4 Sample Selection and Stratification Steps

For norm development, a representative sample of the student population is needed. Six steps were taken to ensure that the norm development criteria were satisfied for each grade and subject.

- ***Step 1.*** The pool of eligible schools that administered an *i-Ready Inform* assessment during each testing window (i.e., fall, winter, spring) for Grades K–8 and the fall testing window for Grades 9 and 10 was identified. To be eligible, schools were required to administer the *i-Ready Inform* assessment to at least 90 percent of their student enrollment (as reported by NCES) in two of the three testing windows for Grades K–8 and to at least 70 percent of their student enrollment in the fall testing window for Grades 9 and 10. This ensured that (1) eligible schools used *i-Ready* with the general student population in the school rather than a specific group such as special education or students considered academically at risk, as the latter would have resulted in oversampling low-achieving students and (2) the NCES school-level demographic information could be used as a strong proxy for the demographics of students in the sampling frame.
- ***Step 2.*** Grade-specific population characteristics were defined by school-level NCES data on urbanicity (i.e., city, suburban, rural/town), region (i.e., Midwest, Northeast, South, West), percentages of racial/ethnic groups (i.e., White, Hispanic, African American), and percentage of students eligible for free- and reduced-price lunch (FRL).
- ***Step 3.*** Grade-specific population targets for the characteristics in step 2 were derived from the population of schools comprising the NCES data. These targets are presented in the next section.
- ***Step 4.*** Schools were grouped into strata based on these characteristics (e.g., a school might have region = southern; urbanicity = city; ethnicity percentages = low White, high Hispanic, high African American; percentage FRL = high). For Grades K–8, schools were randomly sampled iteratively at each grade until the resulting sample was within ±5 percent for all region-by-urbanicity combinations and ±5 percent for demographics.
- ***Step 5.*** Percentile ranks were then calculated. For Grades K–8, the percentile ranks were calculated based directly on students in the sample. For Grades 9 and 10, the normalized weight associated with each stratum was calculated using DeBell and Krosnick's (2009) approach to trim extreme weights. The percentile ranks of students in this weighted sample were calculated for each grade and subject for the fall testing window.
- ***Step 6.*** For Grades K–8, steps 4 and 5 were repeated 100 times for each combination of grade, subject, and testing window. The final norms represent the average percentile ranks across the 100 replications.

8.4.1 Sample Targets

Sample targets are presented in Table 8.1 by grade level.

Table 8.1. Grade-Specific Population Targets Used for Sampling

Ethnicity[1]

Grade	WH	HI	AA
K	45%	28%	14%
1	45%	28%	15%
2	45%	28%	15%
3	45%	28%	15%
4	45%	28%	15%
5	45%	28%	15%
6	45%	28%	15%
7	45%	28%	15%
8	46%	28%	15%
9	46%	28%	15%
10	47%	28%	15%

[1]WH = White; HI = Hispanic; AA = African American

Urbanicity

Grade	City	Suburban	Rural/Town[2]
K	31%	39%	31%
1	31%	39%	30%
2	31%	39%	30%
3	31%	40%	30%
4	31%	40%	30%
5	30%	40%	30%
6	30%	39%	31%
7	30%	39%	31%
8	29%	39%	31%
9	30%	39%	31%
10	30%	39%	31%

[2]Rural and town urbanicity categories were combined for sample composition and post-stratification weight development.

Region[3]

Grade	MW	NE	SO	WE
K	21%	15%	39%	25%
1	20%	15%	40%	24%
2	20%	15%	40%	24%
3	20%	15%	40%	24%

4	20%	15%	40%	24%
5	20%	15%	40%	24%
6	20%	15%	40%	24%
7	20%	15%	40%	24%
8	21%	15%	40%	24%
9	21%	15%	40%	25%
10	21%	15%	40%	25%

[3]MW = Midwest; NE = Northeast; SO = South; WE = West

Free/ Reduced Lunch

Grade	Percent
K	56%
1	56%
2	55%
3	55%
4	55%
5	55%
6	54%
7	54%
8	54%
9	49%
10	49%

8.4.2 Grades K–8 Sample Characteristics

For Grades K–8, the total sample size averaged across replications was 6.9 million students for Reading and 10.1 million students for Mathematics, inclusive of all grades and testing windows. Percentages of special education students and English Learners were also examined based on district-level data of sampled schools obtained from NCES.

The average sample sizes, shown in Table 8.2, were large across all grades, subjects, and windows, ranging from approximately 98,000 to 549,000. Overall, the stratified samples not only had strong demographic representativeness, which is shown in Table 8.3, but also represented the performance levels of a large sample of students at each grade and during a specific testing window.

Table 8.2. Average Sample Sizes by Subject, Grade, and Window

Reading

Grade	Fall	Winter	Spring
K	203,174	232,493	221,278
1	291,409	287,703	282,284
2	309,914	307,738	297,931

3	332,290	338,482	327,942
4	264,835	265,547	285,749
5	314,553	315,173	320,310
6	286,770	285,280	266,269
7	106,533	102,490	98,343
8	183,302	182,172	173,584
Total	2,292,780	2,317,078	2,273,690

Mathematics

Grade	Fall	Winter	Spring
K	257,040	288,988	278,263
1	351,097	355,486	346,002
2	492,085	495,072	488,066
3	397,290	404,935	548,875
4	521,035	521,696	512,036
5	521,180	521,530	506,831
6	373,408	373,268	365,848
7	189,027	186,718	177,802
8	206,898	206,572	203,708
Total	3,309,060	3,354,265	3,427,431

Table 8.3. Average Demographic Representation across Grades K–8 by Subject and Window

Sex

Subject	Window	Female	Male
Reading	Fall	49%	51%
Reading	Winter	49%	51%
Reading	Spring	49%	51%
Mathematics	Fall	49%	51%
Mathematics	Winter	49%	51%
Mathematics	Spring	49%	51%

Ethnicity

Subject	Window	WH	HI	AA	AS	AI	HP	OT
Reading	Fall	46%	27%	15%	6%	0%	0%	6%
Reading	Winter	46%	27%	15%	6%	0%	0%	6%
Reading	Spring	46%	27%	15%	6%	0%	0%	6%
Mathematics	Fall	47%	27%	15%	6%	1%	0%	6%

Mathematics	Winter	47%	27%	15%	6%	1%	0%	6%
Mathematics	Spring	47%	27%	15%	6%	1%	0%	6%

Note: WH = White; HI = Hispanic; AA = African American; AS = Asian; AI = American Indian; HP = Hawaiian Pacific Islander; OT = Other

Limited English Proficient and Students with Disabilities

Subject	Window	LEP	SWD
Reading	Fall	10%	14%
Reading	Winter	10%	14%
Reading	Spring	10%	14%
Mathematics	Fall	10%	14%
Mathematics	Winter	10%	14%
Mathematics	Spring	10%	14%

LEP = Limited English Proficient; SWD = Students with Disabilities

8.4.3 Grades 9 and 10 Sample Characteristics

For Grades 9 and 10, the total sample size was more than 188,000 students for Reading and 144,000 for Mathematics (inclusive of both Grades 9 and 10). Like Grades K–8, percentages of special education students and English Learners were also examined based on district-level data of sampled schools obtained from NCES. The sample sizes for Grade 9 were larger (in excess of 109,000 students for Reading and 87,000 for Mathematics) than the sample sizes for Grade 10 (close to 80,000 students for Reading and 60,000 for Mathematics). Sample sizes for both grades were limited due to the requirement that schools test at least 70 percent of their student enrollment. Many high schools administer *i-Ready Inform* to specific populations of students (e.g., special education or students considered to be academically at risk) and were, therefore, not eligible for inclusion in the norming sample. For this reason, post-stratification weights based on the demographic breakdown of the NCES population were applied so that the resulting sample approximated the population. Additionally, the sample represented performance levels of a large group of students at each grade.

8.5 Development of Percentile Ranks

The percentile ranks associated with each scale score value were calculated for each testing window, grade, and subject (Curriculum Associates, 2024a, 2024b). A student's percentile rank indicates that they performed as well as or better than that percentage of students for their testing window, grade, and subject.

When translating these results into tables that provide the *i-Ready Inform* scale score associated with each percentile rank (i.e., Table 8.4–Table 8.11), the following rules were applied:

- Where multiple *i-Ready Inform* scale scores exist for a single percentile rank, use the lowest score corresponding for that percentile rank.
- Where a percentile rank exists with no corresponding *i-Ready* score, use the *i-Ready* score from the next percentile rank higher that does have a score.

Minor irregularities in the percentile ranks across grade levels were adjusted through policy-based smoothing; this occurred rarely, but when observed, it was most often at the lower ends of the percentile range where data were sparse.

8.6 Updates to National Norms for *i-Ready Inform*

The norms reported in various *i-Ready Inform* results reports will reflect the *i-Ready* 2024 National Norms beginning in the 2024–2025 school year. Curriculum Associates will continue to monitor these norms and will release revised norms in a future school year, as appropriate. In addition, Curriculum Associates will continue to monitor high school testing volumes for winter and spring testing windows and will consider adding National Norms for these testing windows in future years, should student counts reach a sufficient level to support accurate and robust norm-referenced interpretations.

8.7 Resulting Norms

The 2024 Reading and Mathematics Norms for Grades K–8 are presented in Table 8.4–Table 8.9. The fall 2024 Reading and Mathematics Norms for Grades 9 and 10 are presented in Table 8.10 and Table 8.11.

Table 8.4. Fall *i-Ready Inform* for Reading Percentile Rank to Overall Score Conversion*

Percentile Rank	Grade K	Grade 1	Grade 2	Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8
1	100–281	100–303	100–328	100–348	100–367	100–382	100–395	100–404	100–413
2	282–287	304–314	329–340	349–364	368–385	383–401	396–420	405–431	414–445
3	288–291	315–322	341–349	365–376	386–397	402–416	421–436	432–452	446–462
4	292–294	323–327	350–356	377–384	398–405	417–426	437–451	453–463	463–476
5	295–296	328–332	357–361	385–390	406–411	427–437	452–460	464–475	477–486
6	297–299	333–335	362–366	391–394	412–419	438–448	461–468	476–482	487–497
7	300–301	336–338	367–370	395–398	420–424	449–456	469–476	483–491	498–505
8	302–303	339–340	371–374	399–402	425–429	457–461	477–481	492–498	506–512
9	304	341–343	375–378	403–405	430–435	462–467	482–487	499–504	513–516
10	305–306	344–345	379–381	406–408	436–442	468–472	488–493	505–511	517–519
11	307–308	346–348	382–384	409–411	443–448	473–477	494–498	512–514	520–522
12	309	349–350	385–387	412–414	449–452	478–480	499–502	515–517	523–527
13	310–311	351–352	388–390	415–418	453–456	481–484	503–507	518–519	528–530
14	312	353–354	391–392	419–421	457–460	485–488	508–511	520–522	531–534
15	313–314	355–356	393–395	422–423	461–462	489–492	512–514	523–525	535–538
16	315	357	396–397	424–425	463–466	493–496	515–516	526–528	539–541
17	316–317	358	398–399	426–427	467–469	497–498	517–518	529–531	542–544
18	318	359–360	400–401	428–430	470–473	499–501	519–520	532–534	545–547
19	319	361	402–403	431–433	474–475	502–504	521–522	535–537	548–550
20	320	362	404–405	434–436	476–478	505–508	523–524	538–539	551–552
21	321–322	363–364	406	437–440	479–480	509–511	525–527	540–542	553–555
22	323	365	407–408	441–443	481–482	512–513	528–529	543–544	556–558
23	324	366–367	409	444–447	483–484	514	530–531	545–547	559–560
24	324	368	410	448–449	485–487	515–516	532–533	548–549	561–562
25	325	369–370	411–412	450–452	488–490	517	534–536	550–551	563–564
26	326	371	413	453–454	491–492	518–519	537–538	552–554	565–567
27	327	372	414–416	455–457	493–495	520	539–540	555–556	568–569
28	328	373	417–418	458–459	496–497	521–522	541	557–558	570–571
29	328	374	419–420	460	498–499	523–524	542–543	559–560	572
30	329	375	421	461–462	500	525–526	544–545	561–562	573–574
31	330	376	422	463–464	501–503	527	546–547	563	575–576
32	330	377–378	423	465–466	504–505	528–529	548–549	564–565	577–578
33	331	379	424	467–468	506–508	530–531	550–551	566–567	579–580
34	332	380	425–426	469–470	509–511	532–533	552–553	568–569	581–582
35	332	381	427	471–472	512	534	554	570	583

36	333	382	428	473–474	513	535–536	555–556	571–572	584–585
37	334	383–384	429–430	475	514	537	557–558	573–574	586–587
38	334	385	431	476–477	515	538–539	559–560	575	588
39	335	386	432–433	478	516	540	561	576–577	589–590
40	335	387–388	434	479–480	517–518	541–542	562–563	578–579	591
41	336	389	435	481	519	543	564	580	592–593
42	336	390	436–437	482–483	520	544–545	565–566	581–582	594
43	337	391	438	484	521	546	567	583	595–596
44	337	392–393	439–440	485–486	522–523	547–548	568–569	584–585	597–598
45	338	394	441–442	487–488	524	549	570	586	599
46	338	395	443–444	489	525	550–551	571	587–588	600–601
47	339	396	445–447	490–491	526–527	552–553	572–573	589	602
48	340	397	448–450	492–493	528	554	574	590	603–604
49	340	398	451–452	494–495	529–530	555–556	575	591–592	605
50	341	399	453–454	496	531	557	576–577	593	606
51	341	400	455–456	497–498	532	558–559	578	594–595	607–608
52	342	401	457–458	499	533–534	560	579–580	596	609
53	343	402	459	500–501	535	561	581	597–598	610
54	343	402	460–461	502	536–537	562–563	582–583	599	611–612
55	344	403	462	503–504	538	564	584	600	613
56	345	404	463	505–506	539	565–566	585	601–602	614
57	345	405	464–465	507–509	540–541	567	586	603	615
58	346	406	466	510–511	542	568	587–588	604	616
59	347	407	467–468	512	543	569–570	589	605	617–618
60	348	408	469	513	544–545	571	590	606–607	619
61	349	409	470–471	514	546	572	591–592	608	620
62	350	410	472–473	515	547	573	593	609	621
63	351	411	474	516	548–549	574–575	594	610	622
64	352	412	475	517	550–551	576	595–596	611	623–624
65	352	413	476–477	518	552	577–578	597	612–613	625
66	353	413	478	519	553–554	579	598	614	626
67	354	414	479–480	520	555	580	599	615	627–628
68	355	415	481–482	521–522	556–557	581–582	600–601	616	629
69	356	416	483	523	558	583	602	617–618	630
70	356	416	484–485	524–525	559–560	584	603	619	631
71	357	417	486	526	561	585–586	604–605	620	632
72	358	418	487	527–528	562–563	587	606	621–622	633
73	359	419	488–489	529	564	588–589	607	623	634–635
74	360	420	490	530	565–566	590	608–609	624–625	636
75	361	421	491	531–532	567	591–592	610	626–627	637
76	361	422–423	492–493	533	568	593	611–612	628	638
77	362	424	494	534–535	569–570	594–595	613	629	639
78	363	425–426	495–496	536	571	596	614–615	630–631	640–641
79	364-365	427	497	537–538	572–573	597–598	616	632	642
80	366	428–429	498–499	539–540	574–575	599	617–618	633	643
81	367	430–431	500–501	541	576	600–601	619	634–635	644–645
82	368–369	432–433	502–503	542–543	577–578	602	620–621	636	646
83	370	434–435	504	544–545	579–580	603–604	622–623	637	647–648
84	371	436–438	505–506	546–547	581–582	605–606	624	638–639	649–650
85	372–373	439–441	507–508	548–549	583–584	607–608	625–626	640	651
86	374–375	442–444	509–510	550–552	585–586	609–610	627–628	641–642	652–653
87	376	445–447	511–512	553–554	587–588	611–612	629–630	643–644	654–655
88	377–378	448–450	513–515	555–556	589–590	613–614	631–632	645–646	656
89	379–380	451–453	516–517	557–559	591–592	615–616	633–634	647–648	657–658
90	381–383	454–456	518–520	560–562	593–595	617–618	635–636	649–650	659–660
91	384–385	457–460	521–523	563–564	596–597	619–621	637–639	651–653	661–663
92	386–389	461–463	524–526	565–567	598–600	622–624	640–641	654–655	664–665
93	390–392	464–467	527–529	568–570	601–603	625–627	642–644	656–658	666–668
94	393–397	468–471	530–533	571–574	604–606	628–631	645–647	659–661	669–671
95	398–401	472–476	534–537	575–578	607–610	632–634	648–651	662–664	672–675
96	402–406	477–484	538–543	579–583	611–615	635–638	652–655	665–668	676–679
97	407–413	485–491	544–550	584–589	616–621	639–643	656–660	669–673	680–685
98	414–423	492–504	551–560	590–598	622–631	644–650	661–667	674–681	686–692

99	424–800	505–800	561–800	599–800	632–800	651–800	668–800	682–800	693–800

Note: *For *i-Ready Inform* assessments taken between the beginning of the school year and November 15. Analyses were based on the 2022–2023 academic year.

Table 8.5. Winter *i-Ready Inform* for Reading Percentile Rank to Overall Score Conversion*

Percentile Rank	Grade K	Grade 1	Grade 2	Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8
1	100–294	100–320	100–342	100–357	100–373	100–387	100–397	100–404	100–413
2	295–302	321–333	343–356	358–375	374–392	388–410	398–425	405–433	414–448
3	303–307	334–340	357–366	376–387	393–406	411–426	426–446	434–457	449–468
4	308–312	341–346	367–373	388–395	407–417	427–442	447–459	458–470	469–482
5	313–317	347–351	374–379	396–401	418–425	443–454	460–470	471–480	483–495
6	318–320	352–356	380–385	402–406	426–432	455–462	471–478	481–490	496–505
7	321–323	357–359	386–390	407–411	433–442	463–471	479–485	491–499	506–513
8	324–326	360–362	391–394	412–416	443–450	472–477	486–493	500–506	514–517
9	327–328	363–365	395–398	417–421	451–456	478–482	494–499	507–512	518–521
10	329–330	366–368	399–402	422–424	457–461	483–487	500–504	513–516	522–526
11	331–332	369–370	403–405	425–428	462–465	488–493	505–510	517–519	527–531
12	333	371–372	406–407	429–432	466–470	494–498	511–513	520–522	532–535
13	334–335	373–375	408–410	433–437	471–474	499–502	514–516	523–526	536–539
14	336	376–377	411–413	438–442	475–478	503–505	517–518	527–530	540–543
15	337	378–379	414–416	443–447	479–481	506–510	519–521	531–533	544–546
16	338–339	380–381	417–419	448–451	482–484	511–512	522–523	534–537	547–549
17	340	382–383	420–422	452–454	485–487	513–514	524–526	538–540	550–553
18	341	384–385	423	455–457	488–491	515–516	527–529	541–543	554–556
19	342	386–388	424–425	458–460	492–494	517–518	530–532	544–545	557–559
20	343	389–390	426–427	461–462	495–497	519–520	533–535	546–548	560–561
21	344	391	428–429	463–465	498–499	521–522	536–537	549–551	562–564
22	345	392–393	430–431	466–468	500–502	523–524	538–539	552–553	565–567
23	346	394–395	432–433	469–471	503–504	525–526	540–542	554–556	568–569
24	347–348	396	434–435	472–473	505–507	527–528	543–544	557–558	570–571
25	349	397	436–437	474–475	508–511	529–530	545–546	559–560	572–573
26	350	398–399	438–439	476–477	512	531–532	547–548	561–562	574–576
27	351	400	440–442	478–479	513	533–534	549–550	563–564	577–578
28	352	401	443–445	480–481	514–515	535–536	551–552	565–566	579–580
29	353	402–403	446–450	482–483	516	537–538	553–554	567–568	581–582
30	354	404	451–453	484–485	517	539–540	555–556	569–570	583–584
31	355	405	454–455	486–487	518–519	541–542	557–558	571–572	585
32	356	406	456–458	488–489	520	543	559–560	573–574	586–587
33	356	407	459–460	490–491	521	544–545	561	575	588–589
34	357	408	461	492–494	522–523	546–547	562–563	576–577	590–591
35	358	409	462–463	495–496	524–525	548	564–565	578–579	592
36	359	410	464	497	526	549–550	566	580–581	593–594
37	360	411	465–466	498–499	527–528	551–552	567–568	582	595–596
38	360	412	467–468	500–501	529	553	569	583–584	597–598
39	361	413	469	502	530–531	554–555	570–571	585	599
40	362	414	470–471	503–504	532	556–557	572	586–587	600–601
41	363	415	472–473	505–506	533–534	558	573–574	588	602
42	364	416	474	507–509	535	559–560	575	589–590	603–604
43	365	416	475–476	510–511	536–537	561	576–577	591	605
44	366	417	477	512	538	562–563	578	592–593	606–607
45	367	418	478	513	539–540	564	579–580	594	608
46	367	419	479–480	514	541	565	581	595–596	609
47	368	420	481–482	515	542–543	566–567	582	597	610–611
48	369	421	483	516	544	568	583–584	598	612
49	370	422	484	517	545	569	585	599–600	613
50	371	423	485–486	518	546–547	570–571	586–587	601	614
51	372	424	487	519	548	572	588	602	615
52	373	425	488	520	549–550	573	589	603–604	616
53	374	426	489	521	551	574	590	605	617–618
54	375	427	490	522	552–553	575–576	591–592	606	619
55	376	428–429	491	523–524	554	577	593	607–608	620

56	377	430	492–493	525	555–556	578	594	609	621
57	377	431	494	526–527	557	579–580	595–596	610	622
58	378	432–433	495	528	558	581	597	611	623–624
59	379	434–435	496	529	559–560	582	598	612	625
60	380	436	497–498	530	561	583–584	599	613	626
61	381	437–438	499	531–532	562–563	585	600–601	614	627
62	382	439–440	500	533	564	586	602	615–616	628–629
63	383–384	441–442	501–502	534	565	587	603	617	630
64	385	443	503	535–536	566	588–589	604	618	631
65	386	444–445	504	537	567–568	590	605–606	619	632
66	387	446–447	505	538	569	591	607	620–621	633
67	388	448	506	539–540	570	592–593	608	622	634
68	389–390	449–450	507–508	541	571	594	609	623–624	635
69	391	451	509	542	572–573	595	610–611	625	636
70	392	452–453	510	543–544	574	596–597	612	626	637
71	393	454	511–512	545	575	598	613	627–628	638
72	394–395	455	513	546	576–577	599	614	629	639
73	396	456–457	514	547–548	578	600–601	615–616	630	640
74	397	458	515–516	549–550	579–580	602	617	631	641–642
75	398	459–460	517	551	581	603	618–619	632	643
76	399	461–462	518–519	552–553	582	604–605	620	633–634	644
77	400	463	520	554–55	583–584	606	621	635	645
78	401	464–465	521–522	556	585	607–608	622–623	636	646–647
79	402	466	523	557–558	586–587	609	624–625	637	648
80	403–404	467–468	524–525	559–560	588	610–611	626	638–639	649
81	405	469–470	526–527	561–562	589–590	612	627–628	640	650–651
82	406	471–472	528	563	591–592	613–614	629	641	652
83	407–408	473–474	529–530	564–565	593	615	630–631	642–643	653–654
84	409	475–476	531–532	566–567	594–595	616–617	632–633	644	655
85	410	477–479	533–534	568	596–597	618–619	634	645–646	656–657
86	411–412	480–482	535–536	569–570	598–599	620	635–636	647–648	658
87	413	483–484	537–538	571–572	600–601	621–622	637	649–650	659–660
88	414	485–486	539–541	573–574	602–603	623–625	638–639	651–652	661–662
89	415	487–489	542–543	575–576	604–605	626–627	640–641	653–654	663–664
90	416–417	490–492	544–546	577–579	606–607	628–629	642–643	655–656	665–666
91	418–419	493–496	547–548	580–581	608–610	630–631	644–645	657–658	667–668
92	420–421	497–500	549–551	582–584	611–612	632–634	646–648	659–660	669–671
93	422–424	501–503	552–555	585–587	613–615	635–636	649–651	661–663	672–674
94	425–428	504–507	556–559	588–591	616–619	637–639	652–653	664–666	675–677
95	429–433	508–511	560–563	592–594	620–623	640–642	654–657	667–670	678–681
96	434–441	512–516	564–567	595–599	624–628	643–646	658–661	671–674	682–685
97	442–451	517–522	568–573	600–605	629–635	647–651	662–666	675–680	686–690
98	452–464	523–531	574–581	606–613	636–643	652–659	667–673	681–687	691–698
99	465–800	532–800	582–800	614–800	644–800	660–800	674–800	688–800	699–800

Note: *For *i-Ready Inform* assessments taken between November 16 and March 1. Analyses were based on the 2022–2023 academic year.

Table 8.6. Spring *i-Ready Inform* for Reading Percentile Rank to Overall Score Conversion*

Percentile Rank	Grade K	Grade 1	Grade 2	Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8
1	100–303	100–333	100–353	100–365	100–379	100–390	100–403	100–407	100–418
2	304–313	334–345	354–368	366–383	380–399	391–415	404–431	408–440	419–454
3	314–322	346–354	369–378	384–394	400–413	416–432	432–453	441–462	455–473
4	323–328	355–360	379–386	395–403	414–424	433–449	454–465	463–476	474–486
5	329–332	361–365	387–393	404–410	425–433	450–459	466–476	477–485	487–499
6	333–336	366–370	394–399	411–417	434–444	460–468	477–483	486–497	500–509
7	337–339	371–374	400–404	418–422	445–452	469–476	484–492	498–504	510–515
8	340–342	375–378	405–409	423–427	453–459	477–482	493–499	505–511	516–519
9	343–344	379–382	410–413	428–432	460–464	483–488	500–505	512–515	520–524
10	345–347	383–385	414–418	433–439	465–470	489–494	506–511	516–519	525–530
11	348–349	386–389	419–422	440–444	471–474	495–499	512–514	520–522	531–535
12	350–351	390–392	423–424	445–450	475–478	500–503	515–516	523–526	536–539

13	352–353	393–395	425–428	451–454	479–482	504–508	517–519	527–530	540–543
14	354–355	396–397	429–430	455–458	483–485	509–512	520–522	531–534	544–547
15	356–357	398–399	431–433	459–461	486–489	513–514	523–525	535–538	548–550
16	358	400–401	434–436	462–465	490–493	515–516	526–528	539–541	551–554
17	359–360	402–403	437–440	466–468	494–497	517–518	529–531	542–544	555–557
18	361	404–405	441–444	469–472	498–500	519–520	532–534	545–547	558–560
19	362	406–407	445–449	473–474	501–503	521–522	535–537	548–550	561–563
20	363–364	408–409	450–453	475–477	504–505	523–525	538–540	551–553	564–566
21	365	410	454–456	478–479	506–509	526–527	541–542	554–556	567–568
22	366–367	411–412	457–459	480–482	510–512	528–530	543–545	557–558	569–571
23	368	413	460–461	483–484	513–513	531–532	546–547	559–560	572–573
24	369	414	462–463	485–486	514–515	533–534	548–549	561–563	574–575
25	370–371	415–416	464–465	487–489	516	535–537	550–552	564–565	576–577
26	372	417	466–467	490–491	517–518	538–539	553–554	566–567	578–580
27	373	418	468–469	492–494	519	540–541	555–556	568–569	581–582
28	374	419–420	470–471	495–496	520–521	542–543	557–558	570–571	583–584
29	375	421	472–473	497–498	522	544–545	559–560	572–573	585–586
30	376–377	422–423	474–475	499–500	523–524	546	561–562	574–575	587–588
31	378	424	476–477	501–502	525–526	547–548	563–564	576–577	589
32	379	425–426	478	503–504	527–528	549–550	565	578–579	590–591
33	380	427	479–480	505–506	529–530	551–552	566–567	580	592–593
34	381	428–429	481–482	507–509	531	553–554	568–569	581–582	594–595
35	382–383	430	483	510–511	532–533	555–556	570	583–584	596–597
36	384	431–432	484	512	534–535	557	571–572	585–586	598
37	385	433–434	485–486	513	536–537	558–559	573	587	599–600
38	386	435–436	487	514	538	560–561	574–575	588–589	601–602
39	387–388	437–438	488	515–516	539–540	562	576–577	590	603
40	389	439–440	489–490	517	541	563–564	578	591–592	604–605
41	390	441–442	491	518	542–543	565	579–580	593	606
42	391	443	492	519	544	566–567	581	594–595	607–608
43	392–393	444–445	493–494	520	545–546	568	582–583	596	609
44	394	446	495	521	547–548	569–570	584	597–598	610
45	395	447–448	496	522–523	549	571	585–586	599	611–612
46	396	449	497	524	550–551	572	587	600–601	613
47	397	450–451	498–499	525–526	552	573–574	588	602	614
48	398	452	500	527	553–554	575	589–590	603	615
49	398	453	501	528	555	576	591	604–605	616
50	399	454	502	529–530	556–557	577–578	592	606	617–618
51	400	455	503	531	558	579	593–594	607	619
52	401	456	504–505	532	559–560	580–581	595	608–609	620
53	402	457–458	506	533–534	561	582	596	610	621
54	403	459	507	535	562	583	597–598	611	622
55	404	460	508	536–537	563–564	584–585	599	612	623–624
56	405	461	509	538	565	586	600	613	625
57	406	462–463	510–511	539	566	587	601	614	626
58	407	464	512	540	567–568	588	602	615–616	627
59	408	465	513	541–542	569	589–590	603–604	617	628–629
60	409	466	514	543	570	591	605	618	630
61	410	467	515–516	544	571	592	606	619	631
62	410	468	517	545–546	572–573	593	607–608	620–621	632
63	411	469–470	518	547	574	594–595	609	622	633
64	412	471	519	548	575	596	610	623–624	634
65	413	472–473	520–521	549–550	576	597	611	625	635
66	414	474	522	551–552	577–578	598–599	612	626	636
67	414	475–476	523	553	579	600	613–614	627–628	637
68	415	477	524	554	580	601	615	629	638
69	416	478–479	525–526	555–556	581–582	602	616	630	639
70	416	480	527	557	583	603–604	617–618	631	640
71	417	481–482	528	558–559	584–585	605	619	632	641
72	418	483	529–530	560	586	606	620	633	642
73	419	484–485	531	561–562	587	607–608	621	634	643–644
74	420	486	532–533	563	588–589	609	622–623	635–636	645
75	421	487–488	534	564–565	590	610	624	637	646

76	422	489	535–536	566	591	611–612	625–626	638	647
77	423–424	490–491	537	567	592–593	613	627	639	648–649
78	425	492–493	538–539	568–569	594	614–615	628–629	640	650
79	426	494–495	540	570	595–596	616	630	641–642	651
80	427–428	496–497	541–542	571–572	597	617–618	631–632	643	652–653
81	429	498–499	543	573	598–599	619	633	644	654
82	430–431	500–501	544–545	574–575	600–601	620–621	634	645–646	655–656
83	432–433	502–503	546–547	576–577	602	622	635–636	647	657
84	434–436	504	548–549	578–579	603–604	623–624	637	648–649	658
85	437–439	505–506	550–551	580	605–606	625–626	638–639	650–651	659–660
86	440–441	507–508	552–553	581–582	607–608	627–628	640–641	652	661–662
87	442–444	509–510	554–555	583–584	609–610	629–630	642	653–654	663
88	445–447	511–512	556–557	585–587	611–612	631	643–644	655–656	664–665
89	448–450	513–514	558–559	588–589	613–614	632–633	645–646	657–658	666–667
90	451–452	515–516	560–562	590–591	615–616	634–635	647–648	659–660	668–670
91	453–455	517–519	563–564	592–594	617–619	636–638	649–651	661–662	671–672
92	456–459	520–522	565–567	595–596	620–621	639–640	652–653	663–665	673–675
93	460–463	523–525	568–570	597–599	622–626	641–643	654–655	666–668	676–678
94	464–467	526–529	571–573	600–603	626–629	644–645	656–658	669–671	679–681
95	468–472	530–533	574–577	604–607	630–633	646–649	659–662	672–675	682–685
96	473–478	534–537	578–581	608–612	634–638	650–653	663–666	676–679	686–689
97	479–487	538–543	582–587	613–618	639–644	654–659	667–671	680–685	690–695
98	488–499	544–551	588–595	619–626	645–654	660–668	672–679	686–693	696–702
99	500–800	552–800	596–800	627–800	655–800	669–800	680–800	694–800	703–800

Note: *For *i-Ready Inform* assessments taken between March 2 and the end of the school year. Analyses were based on the 2022–2023 academic year.

Table 8.7. Fall *i-Ready Inform* for Mathematics Percentile Rank to Overall Score Conversion*

Percentile Rank	Grade K	Grade 1	Grade 2	Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8
1	100–294	100–306	100–324	100–344	100–358	100–370	100–381	100–389	100–393
2	295–298	307–313	325–333	345–353	359–368	371–382	382–393	390–401	394–406
3	299–300	314–317	334–340	354–360	369–375	383–390	394–401	402–410	407–414
4	301–302	318–321	341–345	361–365	376–381	391–395	402–407	411–416	415–421
5	303–304	322–324	346–348	366–369	382–385	396–400	408–412	417–421	422–425
6	305	325–327	349–351	370–372	386–390	401–404	413–416	422–425	426–430
7	306	328–330	352–354	373–374	391–393	405–408	417–420	426–429	431–433
8	307–308	331–333	355–356	375–377	394–396	409–411	421–423	430–432	434–436
9	309	334–335	357–359	378–380	397–398	412–413	424–426	433–434	437–439
10	310	336–337	360–361	381–382	399–401	414–416	427–428	435–436	440–441
11	311	338–339	362–363	383–384	402–403	417–419	429–431	437–439	442–444
12	312	340	364–365	385–386	404–405	420–421	432–433	440	445–446
13	313	341–342	366	387–388	406–407	422–423	434	441–442	447–448
14	314	343	367–368	389–390	408–409	424	435–436	443–444	449–450
15	315	344	369	391–392	410–411	425–426	437–438	445–446	451–452
16	316	345	370	393–394	412	427–428	439	447	453
17	316	346	371	395	413–414	429–430	440–441	448–449	454–455
18	317	347	372	396	415–416	431	442	450–451	456–457
19	318	348	373–374	397	417	432	443–444	452	458–459
20	319	349	375	398	418–419	433–434	445	453–454	460
21	320	350	376	399	420	435	446	455	461–462
22	320	351	377	400	421–422	436	447	456–457	463
23	321	351	378	401–402	423	437–438	448–449	458–459	464
24	322	352	379	403	424	439	450	460	465–466
25	322	353	380	404	425	440	451	461	467
26	323	354	381	405	426	441	452	462	468
27	324	355	382	406	427–428	442	453–454	463	469
28	324	356	382	406	429	443	455	464	470–471
29	325	357	383	407	430	444	456	465	472
30	326	358	384	408	431	445	457	466–467	473
31	326	359	385	409	432	446	458	468	474
32	327	360	386	410	432	447	459	469	475

33	328	361	387	411	433	448	460	470	476–477
34	329	362	388	412	434	449	461	471	478
35	329	363	389	412	435	450	462	472	479–480
36	330	363	390	413	435	451	463	473	481
37	330	364	391	414	436	452	464	474	482
38	331	365	392	415	437	453	465	475	483
39	332	366	392	416	438	454	466	476–477	483
40	332	366	393	417	438	455	466	478	484
41	333	367	394	417	439	455	467	479	485–486
42	334	368	395	418	440	456	468	480	487
43	334	368	395	419	440	457	469	481	488
44	335	369	396	420	441	458	470	482	489
45	336	369	396	421	442	459	471	483	490
46	336	370	397	422	442	460	472	484	491
47	337	371	398	422	443	460	472	484	492
48	337	371	398	423	444	461	473	485	493
49	338	372	399	424	445	462	474	486	494
50	339	372	399	424	445	463	475	487	495
51	339	373	400	425	446	463	476	488	496
52	340	374	400	426	447	464	477	489	497
53	340	374	401	426	447	465	478	490	497
54	341	375	402	427	448	465	479	491	498
55	341	375	402	428	448	466	480	492	499
56	342	376	403	429	449	467	481	493	500
57	343	376	403	430	450	468	482	494	500
58	343	377	404	430	451	469	483	495	501
59	344	378	405	431	452	469	483	496	502
60	344	378	405	432	452	470	484	496	503
61	345	379	406	432	453	471	484	497	503
62	345	379	406	433	454	471	485	498	504
63	346	380	407	433	455	472	486	499	505
64	347	381	407	434	455	472	487–488	499	506–507
65	347	381	408	435	456	473	489	500	508
66	348	381	408	435	457	473	490	501	509
67	348	382	409	436	457	474	490	501	510
68	349	382	409	436	458	474	491	502	511
69	349	383	410	437	459	475	492	503	512
70	350	384	411	437	460	476	493	503	513
71	350	384	411	438	460	477	494	504	513
72	351	385	412	439	461	478	495	505	514
73	352	386	413	439	462	479	496	506	514
74	352	387	414	440	462	480	496	507	515
75	353	387	414	440	463	481	497	508–509	516
76	354	388	415	441	464	482	498	510	517
77	354	389	416	442	464	483	499	511	518–519
78	355	390	417	442	465	483	500	511	520–521
79	356	391	418	443	466	484	501	512	522
80	357	392	419–420	444	467	485	501	513	523
81	358	393	421	445	468	485–486	502	513	524–525
82	359	394	421	445	469	487	503	514	526–527
83	360	394	422	446	469	488	504	515	528–529
84	361	395	423	447	470	489–490	505	516	530
85	362	396	424–425	448	471	491	506–507	517	531–532
86	363	397	426	449	472	492	508–509	518–519	533–534
87	364	398–399	427	450	473	493–494	510	520–521	535
88	365–366	400	428	451	474	495	511	522–523	536–537
89	367	401	429–430	452–453	475–476	496	512	524–525	538–540
90	368	402–403	431	454	477–478	497–498	513	526–528	541–542
91	369–370	404	432	455	479	499–500	513	529–530	543–545
92	371	405–406	433	456	480–481	501	514	531–532	546–547
93	372–373	407–408	434–435	457	482–483	502–503	515–517	533–534	548–550
94	374–375	409–410	436–437	458–459	484–485	504–507	518–520	535–537	551–553
95	376–377	411–412	438–439	460–462	486–488	508–511	521–524	538–542	554–555

96	378–380	413–417	440–441	463–464	489–491	512–514	525–530	543–546	556–558
97	381–383	418–422	442–445	465–468	492–495	515–517	531–535	547–551	559–564
98	384–390	423–428	446–451	469–474	496–502	518–522	536–542	552–557	565–572
99	391–800	429–800	452–800	475–800	503–800	523–800	543–800	558–800	573–800

Note: *For *i-Ready Inform* assessments taken between the beginning of the school year and November 15. Analyses were based on the 2022–2023 academic year

Table 8.8. Winter *i-Ready Inform* for Mathematics Percentile Rank to Overall Score Conversion*

Percentile Rank	Grade K	Grade 1	Grade 2	Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8
1	100–302	100–319	100–339	100–355	100–369	100–380	100–387	100–393	100–398
2	303–307	320–328	340–348	356–365	370–380	381–393	388–400	394–406	399–411
3	308–310	329–334	349–354	366–372	381–388	394–400	401–408	407–414	412–419
4	311–313	335–339	355–359	373–377	389–393	401–406	409–414	415–420	420–425
5	314–315	340–343	360–363	378–381	394–398	407–411	415–419	421–425	426–431
6	316–318	344–346	364–367	382–386	399–402	412–415	420–423	426–429	432–435
7	319–320	347–348	368–369	387–389	403–405	416–419	424–427	430–433	436–438
8	321	349–350	370–372	390–392	406–408	420–421	428–430	434–436	439–441
9	322–323	351–352	373–374	393–395	409–411	422–424	431–432	437–438	442–444
10	324	353–354	375–376	396–397	412–414	425–426	433–435	439–441	445–446
11	325–326	355–356	377–378	398–399	415–416	427–429	436–437	442–443	447–449
12	327	357–358	379–380	400–401	417–419	430–431	438–439	444–445	450–451
13	328–329	359	381	402–403	420–421	432	440–441	446–447	452–453
14	330	360–361	382–383	404	422	433–434	442–443	448–449	454–456
15	331	362	384–385	405–406	423–424	435–436	444	450–451	457–458
16	332–333	363	386	407–408	425	437–438	445–446	452–453	459
17	334	364–365	387–388	409	426–427	439	447–448	454–455	460–461
18	335	366	389	410–411	428–429	440–441	449	456	462–463
19	336	367	390–391	412	430	442	450–451	457–458	464
20	337	368	392	413–414	431	443	452	459–460	465–466
21	338	369	393	415	432	444–445	453–454	461	467–468
22	339	370	394	416	433	446	455	462	469
23	340	371	395	417–418	434	447	456–457	463–464	470
24	341	371	396	419	435	448	458	465	471–472
25	341	372	396	420	436	449	459–460	466	473
26	342	373	397	421	437	450–451	461	467–468	474–475
27	343	374	398	422	438	452	462	469	476
28	344	375	399	423	439	453	463	470	477–478
29	344	375	400	424	440	454	464	471	479
30	345	376	400	424	441	455	465	472	480–481
31	346	377	401	425	442	456	466	473–474	482
32	347	378	402	426	443	457	467	475	483
33	347	378	403	427	444	458	468	476	484
34	348	379	403	428	445	459	469	477–478	485
35	348	380	404	429	446	460	470	479	486–487
36	349	380	405	430	446	461	471	480	488
37	350	381	405	431	447	462	472	481	489
38	350	382	406	431	448	462	473	482	490
39	351	382	407	432	449	463	474	483	491
40	351	383	407	433	450	464	475	484	492
41	352	384	408	433	450	465	476	485	493
42	353	385	408	434	451	466	477	486–487	494
43	353	385	409	435	452	466	478–479	488	495
44	354	386	410	435	453	467	480	489	496
45	354	387	411	436	454	468	481	490	497
46	355	388	411	436	454	469	482	491	498
47	356	388	412	437	455	470	482	492	499
48	356	389	413	438	456	470	483	493	500
49	357	390	413	438	457	471	484	494	500
50	358	390	414	439	457	472	484	495	501
51	359	391	415	439	458	472	485–486	496	502
52	359	392	415	440	459	473	487	496	502

53	360	392	416	441	460	473	488	497	503
54	361	393	417	441	461	474	489	498	504
55	361	393	418	442	461	475	490	499	505
56	362	394	418	442	462	476	490	500	506
57	362	394	419	443	462	476	491	500	507–508
58	363	395	420	444	463	477	492	501	509
59	364	396	421	444	464	478	493	502	510
60	364	396	421	445	464	479	494	502	511
61	365	397	422	446	465	480	495	503	512
62	366	397	423	446	465	481	496	504	512
63	366	398	423	447	466	482	496	504	513
64	367	399	424	447	467	482	497	505	514
65	368	399	425	448	468	483	498	506	514
66	368	400	425	448	468	483	499	507–508	515
67	369	400	426	449	469	484	500	509	516
68	369	401	426	450	469	485	500	510	517
69	370	402	427	451	470	485–486	501	511	518–519
70	371	402	428	451	471	487	502	511	520
71	371	403	429	452	471	488	502	512	521
72	372	404	430	453	472	489	503	513	522
73	372	405	431	453	473	490	504	513	523
74	373	405	431	454	473	491	504	514	524–525
75	374	406	432	454	474	492	505–506	514	526
76	374	407	433	455	474	493	507	515	527–528
77	375	407	433	456	475	494	508	516	529–530
78	375	408	434	456	476	495	509	517–518	531
79	376	409	434	457	477	496	510	519–520	532
80	377	410	435	458	478–479	497	511	521	533–534
81	378	411	436	458	480	498	512	522	535
82	378	412	437	459	481	499	512	523–524	536–537
83	379	413	438	460	482	500	513	525	538–539
84	380	414	439	461	483	501	514	526–528	540
85	380	415–416	439	462	483	502	514	529	541–542
86	381	417	440	463	484–485	503	515	530–531	543–544
87	382	418–419	441	464	486	504	516–517	532–533	545–546
88	383–384	420	442	464	487–488	505–507	518–519	534	547–548
89	385	421–422	443	465	489	508–509	520–521	535–536	549–550
90	386–387	423	444	466	490–491	510–511	522–524	537–539	551–552
91	388	424–425	445	467–468	492	512	525–527	540–541	553
92	389–390	426–427	446–447	469	493–494	513	528–530	542–544	554–555
93	391–392	428	448	470–471	495–496	514–515	531–533	545–546	556–557
94	393–394	429–430	449–451	472–473	497–499	516	534–535	547–549	558–560
95	395–397	431	452–453	474–475	500–501	517–519	536–539	550–552	561–564
96	398–400	432–434	454–455	476–480	502–506	520–522	540–543	553–555	565–568
97	401–404	435–438	456–458	481–485	507–512	523–529	544–548	556–560	569–573
98	405–410	439–442	459–463	486–492	513–518	530–535	549–554	561–567	574–580
99	411–800	443–800	464–800	493–800	519–800	536–800	555–800	568–800	581–800

Note: *For *i-Ready Inform* assessments taken between November 16 and March 1. Analyses were based on the 2022–2023 academic year.

Table 8.9. Spring *i-Ready Inform* for Mathematics Percentile Rank to Overall Score Conversion*

Percentile Rank	Grade K	Grade 1	Grade 2	Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8
1	100–308	100–328	100–347	100–360	100–372	100–384	100–390	100–394	100–398
2	309–315	329–339	348–356	361–370	373–384	385–396	391–403	395–408	399–411
3	316–320	340–345	357–363	371–378	385–392	397–404	404–411	409–416	412–420
4	321–323	346–349	364–368	379–384	393–398	405–410	412–417	417–423	421–426
5	324–327	350–353	369–372	385–389	399–403	411–415	418–422	424–428	427–432
6	328–330	354–356	373–375	390–393	404–408	416–419	423–426	429–432	433–436
7	331–332	357–359	376–378	394–396	409–412	420–423	427–430	433–436	437–339
8	333–335	360–361	379–381	397–399	413–415	424–425	431–433	437–439	440–442
9	336–337	362–364	382–384	400–402	416–418	426–429	434–436	440–441	443–445

10	338–339	365–366	385–387	403–405	419–421	430–431	437–438	442–444	446–448
11	340–341	367–368	388–389	406–407	422–423	432–434	439–440	445–446	449–451
12	342	369	390–391	408–410	424–425	435–436	441–442	447–448	452–453
13	343–344	370	392–393	411–412	426–427	437–438	443–444	449–450	454–456
14	345	371–372	394	413–414	428–429	439–440	445–446	451–452	457–458
15	346	373	395–396	415–416	430–431	441	447–448	453–454	459–460
16	347	374	397	417–418	432–433	442–443	449–450	455–457	461–462
17	348	375–376	398	419–420	434	444	451–452	458–459	463
18	349	377	399	421	435–436	445–446	453–454	460	464–465
19	350	378	401	422–423	437	447	455–456	461–462	466–467
20	351	379	402	424	438–439	448–449	457–458	463	468–469
21	352	380	403	425	440	450	459	464	470
22	353	381–382	404	426–427	441	451–452	460	465–466	471–472
23	354	383	405	428	442–443	453	461–462	467	473
24	355	384	406	429	444	454–455	463	468–469	474–475
25	356	385	407	430	445	456	464	470	476–477
26	357	386	408	431	446	457–458	465–466	471	478–479
27	358	387	409	432	447	459	467	472–473	480
28	359	388	410	433	448	460	468	474	481–482
29	360	389	411	434	449–450	461	469	475–476	483
30	361	390	412	435	451	462	470	477	484
31	362	391	413	436	452	463	471–472	478–479	485–486
32	362	392	414	437	453	464	473	480	487
33	363	393	415	437	454	465	474	481–482	488–489
34	364	394	416	438	455	466	475	483	490
35	365	395	417	439	456	467	476	484	491
36	365	395	418	440	457	468	477–478	485	492
37	366	396	418	440	458	469	479	486–487	493
38	367	397	419	441	459	470	480–481	488	494
39	368	397	420	442	460	471	482	489	495
40	368	398	421	442	461	472	482	490	496
41	369	399	422	443	462	472	483	491	497
42	369	399	422	444	463	473	484	492–493	498
43	370	400	423	445	464	474	485	494	499
44	370	400	424	446	464	475	486–487	495	500
45	371	401	425	446	465	476	488	496	501
46	372	401	425	447	466	477	489	496	501
47	372	402	426	448	467	478–479	490	497	502
48	373	403	427	448	468	480	491	498	503
49	373	403	428	449	469	481	492	499	504
50	374	404	428	450	469	482	493	500	505
51	374	404	429	451	470	482	494	501	506
52	375	405	430	452	471	483	495	501	507–508
53	375	406	431	453	471	483	496	502	509
54	376	406	432	453	472	484	497	503	510
55	376	407	432	454	472	485	498	504	511
56	377	407	433	455	473	486	498	505	512
57	377	408	434	456	473	487	499	506	513
58	378	409	434	456	474	488	500	507	514
59	378	409	435	457	474	489	501	508–509	514
60	379	410	435	458	475	490	501	510	515
61	379	411	436	459	476	491	502	511	516
62	380	411	436	460	477	492	503	511	517
63	380	412	437	461	478	493	503	512	518–519
64	381	413	437	461	479	494	504	513	520
65	382	413	438	462	480	495	505	513	521
66	382	414	439	462	481	496	506	514	522
67	383	415	439	463	481	497	506–507	515	523
68	384	416	440	644	482	497	508	515	524–525
69	384	417	440	464	483	498	509	516	526
70	385	417	441	465	483	499	510	517	527–528
71	386	418	441	465	484	500	511	518	529
72	387	419	442	466	485	501	512	519–520	530

73	388	420	442	467	486	502	512	521	531–532
74	388	421	443	468	487	503	513	522	533
75	389	422	443	469	488	503	513	523–524	534
76	390	423	444	469	489	504	514	525	535–536
77	391	423	445	470	490–491	505–506	515	526–527	537
78	392	424	445	471	492	507	515	528	538–539
79	393	425	446	471	493	508	516–517	529–530	540
80	394	426	447	472	494	509–510	518–519	531–532	541
81	395	427	447	473	495	511	520–521	533	542–543
82	396	428	448	473	496	512	522	534–535	544–545
83	397	429	449	474	497	512	523–524	536	546
84	398	430	450	475–476	498	513	525–526	537–538	547–548
85	399	430	451	477–478	499	514	527–529	539–540	549–550
86	400	431	452	479	500–501	515	530–531	541–542	551
87	401	432	453	480–481	502	515	532	543–544	552–553
88	402	433	454	482	503–504	516–517	533–534	545–546	554–555
89	403	434	455	483	505–507	518	535–537	547–548	556–557
90	404–405	435	456	484–485	508–510	519–520	538–540	549–550	558–559
91	406–407	436–437	457	486–487	511–512	521–522	541–542	551–552	560–562
92	408	438–439	458–459	488–489	513–515	523–526	543–544	553–554	563–565
93	409–410	440–441	460–461	490–491	516	527–529	545–547	555–556	566–568
94	411–412	442	462–463	492–494	517–518	530–532	548–549	557–559	569–571
95	413–415	443–445	464	495–497	519–520	533–535	550–552	560–563	572–574
96	416–418	446–447	465–467	498–500	521–522	536–537	553–555	564–567	575–578
97	419–423	448–451	468–472	501–506	523–526	538–541	556–559	568–572	579–583
98	424–429	452–455	473–478	507–513	527–530	542–548	560–565	573–579	584–590
99	430–800	456–800	479–800	514–800	531–800	549–800	566–800	580–800	591–800

Note: *For *i-Ready Inform* assessments taken between March 2 and the end of the school year. Analyses were based on the 2022–2023 academic year.

Table 8.10. Fall *i-Ready Inform* for Reading Percentile Rank to Overall Score Conversion*

Percentile Rank	Grade 9	Grade 10
1	100–421	100–439
2	422–450	440–463
3	451–466	464–480
4	467–479	481–494
5	480–488	495–508
6	489–498	509–516
7	499–508	517–521
8	509–514	522–527
9	515–518	528–532
10	519–522	533–538
11	523–526	539–542
12	527–531	543–546
13	532–536	547–550
14	537–539	551–554
15	540–543	555–557
16	544–546	558–561
17	547–549	562–564
18	550–552	565–567
19	553–555	568–570
20	556–558	571–573
21	559–561	574–575
22	562–563	576–578
23	564–565	579–580
24	566–568	581–582
25	569–570	583–585
26	571–572	586–587
27	573–574	588
28	575–577	589–591
29	578–579	592–593
30	580–581	594
31	582	595–596
32	583–584	597–598
33	585–586	599–600
34	587–588	601–602
35	589–590	603
36	591	604–605
37	592–593	606–607
38	594–595	608
39	596	609–610
40	597–598	611
41	599–600	612
42	601	613
43	602–603	614–615
44	604	616
45	605–606	617
46	607	618–619
47	608–609	620
48	610	621
49	611	622–623
50	612–613	624
51	614	625–626
52	615	627
53	616	628–629
54	617	630
55	618–619	631
56	620	632
57	621	633
58	622–623	634
59	624	635–636

60	625–626	637
61	627	638
62	628	639
63	629–630	640
64	631	641
65	632	642–643
66	633	644
67	634	645
68	635–636	646
69	637	647–648
70	638	649
71	639	650–651
72	640–641	652
73	642	653
74	643	654–655
75	644–645	656
76	646	657–658
77	647–648	659
78	649	660–661
79	650–651	662–663
80	652	664–665
81	653–654	666–667
82	655–656	668
83	657–658	669–670
84	659	671–672
85	660–661	673–674
86	662–664	675–676
87	665–666	677–678
88	667–668	679–680
89	669–670	681–682
90	671–672	683–684
91	673–675	685–687
92	676–678	688–689
93	679–680	690–692
94	681–684	693–695
95	685–688	696–699
96	689–692	700–704
97	693–697	705–709
98	698–705	710–718
99	706–800	719–800

Note: *For *i-Ready Inform* assessments taken between the beginning of the school year and November 15. Analyses were based on the 2022–2023 academic year.

Table 8.11. Fall *i-Ready Inform* for Mathematics Percentile Rank to Overall Score Conversion*

Percentile Rank	Grade 9	Grade 10
1	100–393	100–400
2	394–406	401–413
3	407–414	414–421
4	415–421	422–428
5	422–425	429–433
6	426–430	434–437
7	431–433	438–440
8	434–436	441–443
9	437–439	444–446
10	440–441	447–448
11	442–444	449–451
12	445–446	452–453
13	447–448	454–455
14	449–450	456–458
15	451–452	459–460
16	453–454	461–462

17	455–456	463
18	457–458	464–465
19	459	466–467
20	460–461	468–469
21	462	470–471
22	463–464	472
23	465	473–474
24	466–467	475–476
25	468	477–478
26	469–470	479–480
27	471	481
28	472	482
29	473–474	483
30	475	484–485
31	476–477	486
32	478–479	487
33	480	488
34	481	489
35	482	490
36	483	491
37	484	492
38	485	493
39	486	494
40	487	495
41	488	496
42	489	497
43	490	498
44	491	499
45	492	500
46	493	500
47	494	501
48	495	502
49	495	503
50	496	503
51	497	504
52	498	505
53	499	506
54	499	507
55	500	508
56	501	509
57	502	510
58	502	511
59	503	512
60	504	513
61	505	514–515
62	506	516
63	507	517
64	508	518
65	509	519
66	510	520
67	511	521
68	512	522–523
69	513	524
70	514	525
71	515	526
72	516–517	527–528
73	518	529
74	519	530
75	520–521	531
76	522	532
77	523–524	533–534
78	525	535
79	526–527	536

80	528	537
81	529	538–539
82	530–531	540
83	532	541
84	533–534	542–543
85	535	544–545
86	536–537	546
87	538–539	547–548
88	540–541	549
89	542–543	550–551
90	544–545	552–553
91	546–547	554–555
92	548–550	556–558
93	551–552	559–560
94	553–555	561–563
95	556–558	564–567
96	559–562	568–572
97	563–567	573–577
98	568–574	578–585
99	575–800	586–800

Note: *For *i-Ready Inform* assessments taken between the beginning of the school year and November 15. Analyses were based on the 2022–2023 academic year.

Table 8.12 shows the 25th, 50th, and 75th percentiles for 2024 fall, winter, and spring scale scores Grades K–8 for Reading and Mathematics. Table 8.13 shows the 25th, 50th, and 75th percentiles for the fall 2024 scale scores for Grades 9 and 10 for Reading and Mathematics.

Table 8.12. The 25th, 50th, and 75th Percentiles for Grades K–8 Scale Scores by Subject and Grade

Reading

Testing Window	Percentile	Grade K	Grade 1	Grade 2	Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8
Fall	25th	325	369	411	450	488	517	534	550	563
Fall	50th	341	399	453	496	531	557	576	593	606
Fall	75th	361	421	491	531	567	591	610	626	637
Winter	25th	349	397	436	474	508	529	545	559	572
Winter	50th	371	423	485	518	546	570	586	601	614
Winter	75th	398	459	517	551	581	603	618	632	643
Spring	25th	370	415	464	487	516	535	550	564	576
Spring	50th	399	454	502	529	556	577	592	606	617
Spring	75th	421	487	534	564	590	610	624	637	646

Mathematics

Testing Window	Percentile	Grade K	Grade 1	Grade 2	Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8
Fall	25th	322	353	380	404	425	440	451	461	467
Fall	50th	339	372	399	424	445	463	475	487	495
Fall	75th	353	387	414	440	463	481	497	508	516
Winter	25th	341	372	396	420	436	449	459	466	473

Winter	50th	358	390	414	439	457	472	484	495	501
Winter	75th	374	406	432	454	474	492	505	514	526
Spring	25th	356	385	407	430	445	456	464	470	476
Spring	50th	374	404	428	450	469	482	493	500	505
Spring	75th	389	422	443	469	488	503	513	523	534

Table 8.13. The 25th, 50th, and 75th Percentiles for Grades 9 and 10 Fall Scale Scores by Subject and Grade

Subject	Testing Window	Percentile	Grade 9	Grade 10
Reading	Fall	25th	569	583
Reading	Fall	50th	612	624
Reading	Fall	75th	644	656
Mathematics	Fall	25th	468	477
Mathematics	Fall	50th	496	503
Mathematics	Fall	75th	520	531

8.8 Interpretation and Use

Norms are represented by percentiles and percentile ranks that help test users compare each student's performance to a nationally representative sample of students in the same grade who took *i-Ready Inform* at the same time of year. For example, a student whose Grade 5 fall *i-Ready Inform* for Mathematics percentile rank is 90 scored as well or better than 90 percent of a nationally representative sample of students who took the Grade 5 fall *i-Ready Inform* for Mathematics.

Normative scores, together with other types of scores such as *i-Ready Inform* criterion-referenced scores (i.e., grade-level placements for reading and mathematics and for domains), are intended to help educators gain a more complete picture of student performance. Whereas *i-Ready*'s normative scores communicate how students perform compared to other students, criterion-referenced scores communicate what students know and can do against grade-level standards.

National Norms corresponding to overall scale scores are available for *i-Ready Inform* for both Reading and Mathematics. The *i-Ready Inform* normative scores are labeled as "percentile ranks" in *i-Ready* and are located in various places on the *i-Ready* platform, such as the Inform Results (Class) report and the Inform Results (Student) report. Percentile ranks are also available in the Inform Results export.

The correspondence between normative percentile ranks and *i-Ready* placement cut scores is presented in the *i-Ready Diagnostic* 2024 National Norms Technical Documentation for Grades K–8 (Curriculum Associates, 2024a) and *i-Ready Diagnostic* 2024 National Norms Technical Documentation for Grades 9–10 (Curriculum Associates, 2024b). The Grades K–8 report includes the average percentage of students in each *i-Ready* placement level across all 100 replications by subject, grade, and testing window. Additionally, both reports include scale scores corresponding with the 25th, 50th, and 75th normative percentile ranks.

Note that, while Curriculum Associates developed national norms, *i-Ready Inform* is not driven by norm-referenced interpretations; the placement levels are criterion referenced and they represent the primary intended interpretation of student performance. Such score interpretations are driven by what content students have mastered and what content they need to work on next to continue their academic growth.

Chapter 9: Student Growth Measures

9.1 Chapter Summary

Chapter 9 describes *i-Ready Inform*'s two growth measures: *Typical Growth* and *Stretch Growth*. Typical Growth measures are based on the median growth of students with a given starting (i.e., fall) placement level and provide a benchmark for evaluating student improvement compared to students who performed similarly in the fall. Stretch Growth measures complement the Typical Growth measures by presenting a trajectory for students to make progress toward proficiency—or beyond—over a designated period that is both attainable and ambitious. This chapter describes the development of each measure, reports the actual growth measures, and provides guidance for interpreting and using each growth measure. Additionally, this chapter presents reliability coefficients for growth scores.

9.2 Introduction

Student growth is a fundamental measure in education, providing key insights into students' academic progress over time. To this end, we leveraged our vertical scale and our vertically articulated placement levels to create growth measures that provide teachers with both realistic and ambitious goals based on students' initial placement levels. First implemented during the 2018–2019 administration year[9], the growth measures were empirically derived based on the performance data of millions of test-takers across all grades and placement levels. The measures were based on two central ideas:

1. Typical Growth measures provide each student with a spring growth goal that is *normative* in nature. It is designed to provide a sense of how much growth is typical for students, given their starting placement level.
2. Stretch Growth measures provide each student with a spring growth goal that is *criterion referenced* in nature. It is designed to provide an ambitious but attainable growth goal that provides a path toward proficiency or beyond.

Table 9.1 shows the placement categories utilized in the different grade ranges. Student scores can be categorized into on-grade level placements (e.g., Early, Mid, Late) or relative placements (e.g., Two Grade Levels Below). See Chapters 2 and 7 for additional information on placement levels.

Table 9.1. Placement Categories for the *i-Ready Inform* Growth Model

Placement Category	Grade K	Grades 1–2	Grades 3–8
On, Mid, or Above Grade Level	✓	✓	✓
Early On Grade Level	✓	✓	✓
One Grade Level Below	✓	✓	✓

[9]The previous growth model was based on median growth in each subject and grade but did not feature differentiation by starting placement level and did not provide a path toward proficiency.

Two Grade Levels Below		✓	✓
Three or More Grade Levels Below			✓

Note: In Grades K and 1, the lowest placement grade level is Emerging K. However, once students reach Grade 2, the lowest placement grade-level designation is Grade K. Therefore, there is no placement category of Two Grade Levels Below for Grade K or Three or More Grade Levels Below for Grades 1 or 2 students.

9.3 Typical Growth

Typical Growth is a descriptive measure of growth based on the observed growth of *i-Ready Inform* test-takers. It is not intended to serve as guidance for what would constitute *sufficient* growth; however, it offers a useful reference point to help identify individuals or groups of students who are growing either much less or much more than their peers.

9.3.1 Development Process

Typical Growth is based on the median growth of students with a given placement level in the fall. It provides a tool for evaluating student improvement compared to the typical student in the same placement level. For each subject, grade, and fall placement level, Curriculum Associates used the following four-step process to create the Typical Growth measures.

1. A sample of students was selected from the 2015–2016 and 2016–2017 school years using the following selection criteria:
    a. The initial *i-Ready Inform* test was taken in the fall testing window for those years
    b. The time between the fall and final *i-Ready Inform* test of the same academic year was at least 25 weeks
    c. Neither the initial nor the final test was flagged for rushing
    d. The student's chronological grade was consistent at the times of first and last tests
2. For each fall placement level, chronological grade, and subject combination, there were three steps followed to determine the raw growth measure:
    a. The gain score was calculated as the last overall test score from the school year minus the fall overall test score
    b. A prorated gain score was computed for each student using 30 weeks as the proration factor
    c. The Typical Growth measure was calculated as the median prorated growth
3. A linear regression model was then used to smooth variation across the measures so that within any grade, they increase monotonically as the placement levels decrease, thereby resulting in smoothed measures.
4. Lastly, any growth measure that was less than four scale score points was set to four scale score points, thus ensuring that all students had a positive growth measure.

9.3.2 Measures of Typical Growth

Table 9.2 and Table 9.3 show the Typical Growth measures by fall placement category for Grades K–8 in Reading and Mathematics, respectively. These Typical Growth measures help educators understand how students are expected to grow based on subject, fall placement category, and grade.

Table 9.2. Typical Growth Measures in Scale Score Points by Grade Level: Reading

Fall Placement Category	K	1	2	3	4	5	6	7	8
On, Mid, or Above Grade Level	43	37	22	17	12	7	4	4	4
Early On Grade Level	44	47	29	22	17	13	9	6	4
One Grade Level Below	49	49	39	26	20	16	12	10	9
Two Grade Levels Below	N/A	54	44	33	23	20	14	12	12
Three or More Grade Levels Below	N/A	N/A	N/A	36	28	26	19	17	18

Note: N/A = Not Applicable Due to Grade Level

Table 9.3. Typical Growth Measures in Scale Score Points by Grade Level: Mathematics

Fall Placement Category	K	1	2	3	4	5	6	7	8
On, Mid, or Above Grade Level	21	21	18	21	19	14	13	11	9
Early On Grade Level	24	26	22	25	23	18	13	12	9
One Grade Level Below	32	29	26	26	23	18	14	12	9
Two Grade Levels Below	N/A	36	29	27	23	18	14	13	10
Three or More Grade Levels Below	N/A	N/A	N/A	30	24	20	15	13	12

Note: N/A = Not Applicable Due to Grade Level

9.3.3 Considerations for Interpreting and Utilizing Typical Growth Measures

To aid in interpreting growth relative to the Typical Growth measure, we compute the *percent progress toward Typical Growth*, which is calculated as the actual growth divided by the measure. For example, a student who had a Typical Growth goal of 25 but had an actual growth of 24 would have a 96% on their percent progress.

The Typical Growth measures may also be used in an accountability framework whose purpose is to determine how much more or less than expected a group of students grew. Educators can use the median percent progress toward Typical Growth to determine how a group of students grew during a school year.

When the median percent progress toward Typical Growth is:

- Substantially less than 100%, then students grew less than expected
- About 100%, then students grew about as expected
- Substantially more than 100%, then students grew more than expected

While these Typical Growth measures are ideally suited for understanding how students are expected to grow based on subject, fall placement category, and grade, it is important to note that if students continue to grow in a manner consistent with their Typical Growth measures, then it is unlikely that they will see increases in their proficiency levels over time.

9.4 Stretch Growth

To challenge students to attain higher levels of proficiency over time, Curriculum Associates developed Stretch Growth measures to complement the Typical Growth measures discussed in the preceding section.

Stretch Growth presents a trajectory for students performing below grade level to make progress toward proficiency over a designated period. For students already achieving proficiency, Stretch Growth provides a path toward higher proficiency levels.

9.4.1 Development Process

In contrast to Typical Growth, which is based on median growth given a starting placement level, Stretch Growth is derived from the actual growth trajectory exhibited by students who achieved grade-level proficiency. It is designed to be ambitious but attainable (i.e., between the 55th and 80th percentiles of growth for students in a given starting placement level). This provision ensures that the attainment of the Stretch Growth measures remains feasible with additional instructional support and focused effort. To formulate the Stretch Growth measures for each subject, grade, and fall placement category, Curriculum Associates adopted the following six-step process.

1. We determined the eventual spring placement-level goals for each fall (i.e., starting) placement, which are summarized in Table 9.4. Note that the amount of time it takes students to achieve these goals may vary by grade level and starting placement level. Details on the timeline(s) associated with each goal will be discussed in detail in the next section.
2. A sample was selected consisting of students who met their spring placement goals as outlined in Table 9.4 in either one or two years, provided that their growth was reasonable, which was defined as being at or lower than the 80th percentile. Students with two years of data were included in the sample if they grew less in the second year than in the first year and did not experience substantial summer learning loss between the spring of year one and the fall of year two, which was defined as twice the SEM of growth between the two administrations.
3. Prorated first-year growth scores were computed separately for students who met their placement goals in one year and for students who did not meet their placement goals until the second year using a 30-week proration factor. The medians of the prorated growth of the two groups of students were designated as one-year and two-year Stretch Growth measures, respectively.
4. Medians were then limited to certain thresholds to ensure the proposed growth measures were both ambitious and attainable. The upper limit was set at the 80th percentile of observed growth for the 2016–2017 academic year, while the lower limit was the higher of (a) four scale points beyond the corresponding Typical Growth measure and (b) the 55th percentile of growth.
5. Next, it was determined whether a Stretch Growth measure should be based on student growth for those students who achieved their proficiency goal in one year by checking whether the median fell within the thresholds from step 4. If not, it was determined whether a Stretch Growth measure should be based on student growth for students who achieved their proficiency goal in two years. If neither is reasonable, then the Stretch Growth measure was set to the 80th percentile of growth for that subject, grade, and placement category.
6. Finally, a judgment-based smoothing process was conducted to ensure that:
    a. Stretch Growth measures increase with decreasing placement levels within a grade
    b. Stretch Growth measures increase with decreasing grade level within a starting placement level

Table 9.4. Placement Goals for Stretch Growth Measures

Fall Placement Level	Eventual Spring Placement Goal
On, Mid, or Above Grade Level	Late On Grade
Early On Grade Level	Late On Grade
One Grade Level Below	Mid On Grade
Two Grade Levels Below	Mid On Grade
Three or More Grade Levels Below	Mid On Grade

These stipulations ensure that Stretch Growth measures for lower-achieving students are greater than or equal to the recommended Stretch Growth measures for higher-achieving students, and that the Stretch Growth measures for students in lower grades with the same placement category are not lower than the Stretch Growth measures for students in higher grades with the same placement category. For example, a student with a placement of One Grade Level Below in Grade 2 should have a Stretch Growth measure no less than a student with a placement of One Grade Level Below in Grade 3.

9.4.2 Measures for Stretch Growth

Table 9.5 and Table 9.6 show the Stretch Growth measures by fall placement category for Grades K–8 in Reading and Mathematics, respectively. Additionally, Curriculum Associates provides a color-coded designation that indicates the length of time associated with each goal:

- Red: A realistic path toward reaching the proficiency goal would likely take more than two years.
- Yellow: A realistic path toward reaching the proficiency goal may take about two years. Note that these values are based on the median first-year 30-week growth of students who met their proficiency goal in two years; some students may not attain this level of growth.
- Green: A realistic path toward reaching the proficiency goal may take about one year. These values are based on the median first-year 30-week growth of students who met their proficiency goals in one year in a realistic manner; some students may not attain this level of growth.

Table 9.5. Stretch Growth Measures in Scale Score Points by Grade Level: Reading

Fall Placement Category	K	1	2	3	4	5	6	7	8
On, Mid, or Above Grade Level	Green (54)	Green (44)	Green (27)	Green (21)	Green (20)	Green (18)	Green (15)	Green (14)	Green (13)
Early On Grade Level	Green (65)	Green (56)	Green (43)	Green (39)	Yellow (27)	Yellow (25)	Yellow (25)	Yellow (23)	Yellow (22)
One Grade Level Below	Green (67)	Green (67)	Green (53)	Yellow (40)	Yellow (36)	Yellow (30)	Yellow (26)	Yellow (25)	Yellow (25)
Two Grade Levels Below	N/A	Yellow (96)	Yellow (81)	Yellow (63)	Red (50)	Red (47)	Red (38)	Red (37)	Red (36)
Three or More Grade Levels Below	N/A	N/A	N/A	Red (79)	Red (62)	Red (61)	Red (51)	Red (50)	Red (50)

Note: N/A = Not Applicable Due to Grade Level; Green = One-Year Goal; Yellow = Two-Year Goal; Red = Two-Plus-Year Goal

Table 9.6. Stretch Growth Measures in Scale Score Points by Grade Level: Mathematics

Fall Placement Category	K	1	2	3	4	5	6	7	8
On, Mid, or Above Grade Level	Green (35)	Green (32)	Green (31)	Green (30)	Green (24)	Green (20)	Green (20)	Green (20)	Green (19)
Early On Grade Level	Yellow (38)	Yellow (36)	Yellow (35)	Yellow (34)	Yellow (33)	Yellow (29)	Yellow (25)	Red (22)	Red (21)
One Grade Level Below	Green (39)	Green (37)	Green (36)	Green (35)	Green (34)	Green (31)	Yellow (26)	Yellow (23)	Yellow (22)
Two Grade Levels Below	N/A	Yellow (57)	Yellow (48)	Yellow (43)	Yellow (41)	Yellow (35)	Red (30)	Red (25)	Red (23)
Three or More Grade Levels Below	N/A	N/A	N/A	Red (55)	Red (47)	Red (41)	Red (35)	Red (33)	Red (31)

Note: N/A = Not Applicable Due to Grade Level; Green = One-Year Goal; Yellow = Two-Year Goal; Red = Two-Plus-Year Goal

Additionally, the Stretch Growth proficiency target shifts from mid to late on grade level for students with a starting placement of early on grade, and thus the pattern for number of years expected to reach this goal also shifts.

9.4.3 Considerations for Interpreting and Utilizing Stretch Growth Measures

The Stretch Growth goal is updated annually for each student based on their most recent fall *i-Ready Inform* assessment. This ongoing adjustment ensures that students on multiyear trajectories are guided by their proficiency at the beginning of each year, reflecting their growth in preceding years. Next, we present some key considerations for implementing Stretch Growth measures.

9.4.3.1 Low Score in Fall Placement Category

Students beginning with a lower score in their fall placement level may find the Stretch Growth measures insufficient to reach their placement goal. For example, Mid On Grade Level for Grade 1 Mathematics ranges from 413 to 454, and Late On Grade Level ranges from 455 to 496. Consider a student who achieved a 413 on the fall test: their placement goal would be Late On Grade Level, and their measure from Table 9.6 would be 32. Achieving the measure would mean that they would receive a spring score of 445, which would not place them in Late On Grade Level. However, attaining these Stretch Growth measures can bring them closer to proficiency than if they merely achieved Typical Growth, which in this example would be 21 points, or a spring score of 434.

9.4.3.2 Summer Learning Loss

Two-year Stretch Growth measures can be influenced by summer learning loss—the phenomenon whereby students' scores decline from the spring test of one year to the fall test of the next. Summer learning loss may hinder students from meeting their proficiency goals within the intended time frame.

9.4.3.3 Prorated Data Basis

The sample includes students who met their proficiency goals based on non-prorated data. The 30-week prorating of their growth was applied to control for varying numbers of instructional weeks between tests.

9.4.4 Not Intended for Use in Accountability Contexts

Given that many students are unlikely to meet these Stretch Growth measures, Curriculum Associates strongly advises against using the Stretch Growth measures in any accountability framework. Unlike the Typical Growth measures, which are approximately the 50th percentile of growth for each subject, grade, and fall placement category, Stretch Growth measures range from the 55th percentile to the 80th percentile of one-year growth. The percentile ranks of the measures are shown in Table 9.7 and Table 9.8.

Table 9.7. Percentile Ranks of Stretch Growth: Reading

Fall Placement Category	K	1	2	3	4	5	6	7	8
On, Mid, or Above Grade Level	55	55	55	55	63	71	72	74	75
Early On Grade Level	73	55	70	75	66	70	77	80	80
One Grade Level Below	69	65	65	67	72	69	69	71	75
Two Grade Levels Below	N/A	80	80	80	80	80	80	80	80
Three or More Grade Levels Below	N/A	N/A	N/A	80	80	80	80	80	80

Note: N/A = Not Applicable Due to Grade Level

Table 9.8. Percentile Ranks of Stretch Growth: Mathematics

Fall Placement Category	K	1	2	3	4	5	6	7	8
On, Mid, or Above Grade Level	75	73	80	70	60	64	65	74	77
Early On Grade Level	78	70	80	67	71	75	76	71	76
One Grade Level Below	60	62	68	66	70	76	80	80	77
Two Grade Levels Below	N/A	77	78	77	79	80	80	80	80
Three or More Grade Levels Below	N/A	N/A	N/A	80	80	80	80	80	80

Note: N/A = Not Applicable Due to Grade Level

Given the heterogeneity of Stretch Growth measures in terms of percentile ranks, aggregating classrooms, schools, or districts in terms of progress toward Stretch Growth will lead to flawed inferences and thus is not recommended.

9.5 High School Growth

The growth model discussed thus far in this chapter has been designed for Grades K–8. Based on years of research and analysis of millions of student *i-Ready Inform* assessments taken, this model advances measurement of growth by including both Typical Growth and Stretch Growth measures differentiated by students' baseline performance (i.e., placement on the fall assessment). Because of the more limited population using *i-Ready Inform* in high school, the same level of research is not currently available for Grades 9–12. Thus, two alternative approaches to measuring growth are available in Grades 9–12: one that evaluates change in grade-level placements and another that uses generalized Typical Growth and Stretch Growth targets to evaluate growth in Grades 9–12.

9.5.1 Change in Grade-Level Placements Approach

One option for evaluating the growth of a student in Grades 9–12 is to examine students' grade-level placements at the beginning of the year and then again later in the year after a subsequent administration of *i-Ready Inform*. One report within *i-Ready* that can be used to evaluate growth in this way is the student-level Inform Growth report, shown in Chapter 11. In the *Placement by Domain* section of the report, growth is evaluated in terms of the number of grade-level placements a student has changed from their baseline placement and provides detailed information for individual domains on *i-Ready Inform* in addition to the change in placement of a student's overall score. A green arrow indicates that the placement has improved

relative to the baseline assessment. The class-level Inform Growth report, also discussed in Chapter 11, can also be used to understand these results.

Importantly, for mathematics, the ability to evaluate a student's placement depends on which mathematics course sequence a student is taking and the level at which growth is being evaluated. There are two primary course sequences across the United States:

- The Traditional Pathway, in which a student takes algebra in Grade 9, geometry in Grade 10, algebra 2 in Grade 11, and a more complex mathematics course such as pre-calculus in Grade 12
- The Integrated Pathway, in which a student receives instruction in algebra, geometry, measurement and data, and number and operations in each grade from Grade 9 to Grade 12

When looking at growth within a year, it is recommended that educators evaluate a student's change in Overall Placement for students taking the Integrated Pathway because the student is exposed to content covering all four *i-Ready* for Mathematics domains that contribute to the student's Overall Placement, but it is not advised to evaluate change in Overall Placement for students taking the Traditional Pathway who are receiving instruction specific to only one *i-Ready* domain (e.g., Algebra and Algebraic Thinking in Grade 9 or Geometry in Grade 10) because the student is not receiving instruction in all domains that contribute to the student's Overall Placement.

When looking at growth within a year for students taking the Traditional Pathway, evaluating growth at the domain level can be done, provided it is within the *i-Ready* domain in which the student is receiving instruction. That is, for a Grade 9 student taking an algebra course, it would be reasonable to evaluate the student's change in placement within the *i-Ready* Algebra and Algebraic Thinking domain, but not within the Measurement and Data or Geometry domains because the student is not receiving instruction in these domains. Because students in the Integrated Pathway usually receive instruction across all four domains, evaluating growth by examining changes in *i-Ready* domain-level placements can be done. Table 9.9 provides a summary of the instances in which changes in placements can be used to make inferences about student growth.

Table 9.9. Recommendations for When to Use the Change-in-Placement Approach to Growth in High School Mathematics

Pathway	Course in Which Student Is Receiving Instruction	Evaluate Growth Using Change in Overall Placements?	Evaluate Growth Using Change in Domain Placements?
Traditional	Algebra 1 or Geometry	No	Yes (For the course in which the student is receiving instruction)
Traditional	Algebra 2 or Later	Yes, assuming the course includes instruction from all four Mathematics domains	Yes
Integrated	Grades 9–12 (Integrated I-IV)	Yes	Yes (Assuming the student is receiving instruction across the domains, as in common of the Integrated Pathway)

9.5.2 Generalized Growth Target Approach

An alternative approach to evaluating growth using change in placements is to use the generalized Typical Growth and Stretch Growth targets from the *i-Ready* growth model. After students complete their first *i-Ready Inform* assessment, it is possible to use the generalized Typical Growth and Stretch Growth targets to begin evaluating growth for every student. Throughout the year, student progress toward these targets can be monitored to better understand how students are growing compared to students like them and how much they need to grow to close the gap to proficiency.

These targets are carried over from Grade 8, so neither the Typical Growth nor the Stretch Growth targets carry the same meaning as they do for Grades K–8; however, they still provide educators with two sets of growth targets that can be used to inform instruction. For this reason, we only recommend this approach for use by educators who are comfortable with *i-Ready*'s growth model and have facility with data analysis and who plan to use growth for "low-stakes" purposes. These targets are not available through our reporting tools, and thus require an educator to conduct some calculations outside of the *i-Ready Connect* system using information provided in this section, but it gives the advantage of providing an approach to measuring growth that is analogous to what's available within *i-Ready* for Grades K–8.

As with the change-in-placement approach to evaluating growth for students in high school discussed in the previous section, the generalized Typical Growth and Stretch Growth targets may not be appropriate for students in the Traditional Pathway for mathematics who are receiving instruction in only one domain. For example, because a student in a Traditional Pathway who is taking an algebra course is likely only getting instruction on algebra content, and because the student's overall *i-Ready* score is based on their performance across all four Mathematics domains (i.e., Algebra and Algebraic Thinking, Measurement and Data, Geometry, and Number and Operations) and not just the domain in which the student is receiving instruction, the student's overall score that is used to measure growth under this approach will likely not fully convey the amount of growth in the domain that the student is experiencing.

Table 9.10 provides guidance on when it is appropriate to use the generalized growth targets. The generalized targets are displayed in Table 9.11–Table 9.14.

Table 9.10. Recommendations for When to Use the Generalized Typical Growth and Stretch Growth Targets Approach to Growth in High School Mathematics

Pathway	Course in Which Student Is Receiving Instruction	Evaluate Growth Using the Generalized Typical Growth and Stretch Growth Targets?
Traditional	Algebra 1 or Geometry	No
Traditional	Algebra 2 or Later	Yes, this method may be appropriate for students in Algebra 2 or later courses that include content across all four Mathematics domains
Integrated	Grades 9–12 (Integrated I–IV)	Yes

Table 9.11. Typical Growth Targets: Mathematics—Grades 9–12

Fall *i-Ready Inform* Placement Level	**9**	**10**	**11**	**12**
On, Mid, Late, or Above Grade Level	9	9	9	9
Early On Grade Level	9	9	9	9
One Grade Level Below	9	9	9	9
Two Grade Levels Below	10	10	10	10
Three or More Grade Levels Below	12	12	12	12

Table 9.12. Typical Growth Targets: Reading—Grades 9–12

Fall *i-Ready Inform* Placement Level	**9**	**10**	**11**	**12**
On, Mid, Late, or Above Grade Level	4	4	4	4
Early On Grade Level	4	4	4	4
One Grade Level Below	9	9	9	9
Two Grade Levels Below	12	12	12	12
Three or More Grade Levels Below	18	18	18	18

Table 9.13. Stretch Growth Targets: Mathematics—Grades 9–12

Fall *i-Ready Inform* Placement Level	**9**	**10**	**11**	**12**
On, Mid, Late, or Above Grade Level	19	19	19	19
Early On Grade Level	21	21	21	21
One Grade Level Below	22	22	22	22
Two Grade Levels Below	23	23	23	23
Three or More Grade Levels Below	31	31	31	31

Table 9.14. Stretch Growth Targets: Reading—Grades 9–12

Fall *i-Ready Inform* Placement Level	**9**	**10**	**11**	**12**
On, Mid, Late, or Above Grade Level	13	13	13	13
Early On Grade Level	22	22	22	22
One Grade Level Below	25	25	25	25
Two Grade Levels Below	36	36	36	36
Three or More Grade Levels Below	50	50	50	50

9.6 Growth Score Reliability

Growth score reliability coefficients help us to understand the reliability of the growth score between two administrations, in this case, the fall and the spring administrations. Briggs (2020) outlined an approach that is based on the methodology described in Willett (1989), which we use here.

The growth score reliability requires scores from two administrations, which in this case will be the fall and the spring administrations, which correspond to the growth measures presented earlier in this chapter. For these analyses, we started with the data used in Chapter 6, but only used students who tested in both the fall and spring windows. Sufficiently high reliability provides evidence in support of the precision of our growth scores. Because these scores take into account measurement error from two administrations, as well as the correlation between the scores and their variances, the reliabilities in this section will necessarily be lower than those presented in Chapter 6.

Because the time in between testing occasions can be related to the amount of growth shown, we use a 30-week prorated score for the second testing occasion. To derive this score, we calculate the number of weeks between the two testing occasions as well as the difference between the scores. We then multiply this difference by 30/*W*, where *W* is the number of weeks between the scores. Finally, we add this number to the first score to get the 30-week prorated spring score, Y_2, which is used in our calculations.

The growth score reliabilities are estimated as:

$$\widehat{\rho_G} = \frac{\sigma_{Y_1}^2 \rho(Y_1) + \sigma_{Y_2}^2 \rho(Y_2) - 2\sigma_{Y_1}\sigma_{Y_2}\rho(Y_1Y_2)}{\sigma_{Y_1}^2 + \sigma_{Y_2}^2 - 2\sigma_{Y_1}\sigma_{Y_2}\rho(Y_1Y_2)}, \text{where} \qquad (9.1)$$

$\widehat{\rho_G}$ is the estimated reliability of the growth score (i.e., the difference between two scores), $\sigma_{Y_1}^2$ and $\sigma_{Y_1}^2$ are variances of the observed scores from the first and second administrations, respectively, $\rho(Y_1Y_2)$ is the correlation between the scores from the first and second administrations, and $\rho(Y_1)$ and $\rho(Y_2)$ are the reliabilities of each of the administrations, respectively, which will be assumed to be equal for the purposes of this section since the reliabilities in Chapter 6 are estimated across all three testing windows.

Analogous to the SEM computation described in Chapter 6, the SEM of the growth scores can be computed as:

$$\widehat{SEM_G} = \sigma_G\sqrt{1 - \widehat{\rho_G}} \qquad (9.2)$$

where σ_G is the observed variance of the 30-week prorated growth scores.

Growth score reliabilities, SEMs, and counts are presented in Table 9.15 for each grade and subject/language combination. All reliability coefficients are at least .70.

Table 9.15. Growth Score Reliability by Subject, Language, and Grade

Mathematics (English)

Grade	Reliability	SEM	Count
K	0.83	7.8	589,264
1	0.80	8.0	739,133
2	0.74	8.2	806,761
3	0.71	8.3	828,817
4	0.70	8.5	813,119
5	0.70	8.6	811,260
6	0.74	8.7	646,347
7	0.78	8.7	548,469
8	0.81	8.8	497,499
9	0.86	8.6	57,107
10	0.86	8.6	34,544
11	0.89	8.7	21,655
12	0.91	8.9	10,896

Reading (English)

Grade	Reliability	SEM	Count
K	0.84	11.0	475,777
1	0.81	11.9	615,971
2	0.73	13.7	670,598
3	0.74	13.6	716,171
4	0.71	14.1	696,005
5	0.70	14.4	687,984
6	0.73	14.4	544,551
7	0.75	14.5	475,200
8	0.76	14.6	452,713
9	0.82	14.8	61,920
10	0.81	15.1	40,080
11	0.83	15.5	24,094
12	0.86	16.1	13,958

Mathematics (Spanish)

Grade	Reliability	SEM	Count
K	0.86	7.7	16849
1	0.83	8.0	20178
2	0.78	8.3	18699

3	0.76	8.5	10326
4	0.73	9.0	7497
5	0.74	9.0	6349
6	0.84	8.2	4127
7	0.84	9.0	3348
8	0.85	8.7	3281
9	0.83	9.6	449
10	0.85	9.9	435
11	0.89	8.2	137
12	**	**	**

Note: * = Omitted Due to a Sample Size of Less Than 100

9.7 Validity Evidence

9.7.1 Growth and *i-Ready Personalized Instruction* Lesson Pass Rates

Curriculum Associates has conducted several research studies that provide validity evidence in support of the use cases of our growth model. The first study was exploratory research investigating the relationship between pass rates on *i-Ready Personalized Instruction* lessons (i.e., the proportion of lessons a student passed) and growth as measured by *i-Ready Inform*. Note that *i-Ready Personalized Instruction* lessons are not a component of our theory of action discussed in Chapter 1 but are used by many districts and require an *i-Ready Inform* placement level. Data were used from both the lessons and *i-Ready Inform* for students in Grades K–8 in the 2020–2021 academic year.

Data were analyzed using a multiple regression model using pass rate to predict the percentage of a student's Typical Growth target they will achieve. The model was run for each subject (i.e., Reading and Mathematics), grade, and fall 2020 *i-Ready Inform* placement-level combination. Marginal estimated means produced from these models indicated that, as pass rate increases, growth on *i-Ready Inform* also tends to increase. This pattern was consistent across subject, grade, and placement level. Due to the data collection design, this study does not support a causal conclusion; it does, however, support conclusions about the ability of *i-Ready Inform* to detect differences in growth among students.

9.7.2 Stretch Growth and Proficiency

The second study used two years of *i-Ready Inform* data to examine to what degree meeting *i-Ready*'s Stretch Growth targets put students on a path toward proficiency. We used data from more than 1.8 million students and more than 1.9 million students who completed fall and spring *i-Ready Inform* for Reading and for Mathematics, respectively, during the 2017–2018 and 2018–2019 school years. Our analyses found that more than 30 percent of students met Stretch Growth in a given school year and approximately 12 percent of students met Stretch Growth in both years. In both reading and mathematics, meeting Stretch Growth targets in two consecutive years put students on a path toward proficiency; approximately 80 percent of these students finished the 2018–2019 academic year at Mid On Grade Level or higher in *i-Ready*, while 95 percent finished at or above Early On Grade Level. Results of this study provide evidence that *i-Ready*'s Stretch Growth

targets are both ambitious and attainable and demonstrate how reaching Stretch Growth in two consecutive years can place students on a path toward proficiency.

The design of the third study was similar to that of the second study, but it used data from after the onset of the COVID-19 pandemic. We used data from more than 3.1 million students and more than 3.8 million students who completed fall and spring *i-Ready Inform* assessments for Reading and for Mathematics, respectively, during the 2021–2022 and 2022–2023 academic years. Our analyses found that more than 25 percent of students met Stretch Growth in a given academic year and approximately 10 percent of students met Stretch Growth in both years. Across grade levels in reading and mathematics, approximately 76 percent of students who met Stretch Growth in both years finished the 2022–2023 academic year at Mid On Grade Level or higher in *i-Ready*, while 93 percent finished at or above Early On Grade Level.

Although these outcomes represent a slight drop compared to the students represented in the pre-pandemic study, the results are directionally consistent and provide evidence that *i-Ready*'s Stretch Growth targets are both ambitious and attainable and demonstrate how reaching Stretch Growth in two consecutive years puts students on a path toward proficiency.

9.7.3 Fairness

A set of analyses was conducted to investigate the associations between several demographic variables (i.e., ethnicity and English Learner (EL) status) and the attainment of Typical Growth measures in grades K–8. To conduct this investigation, a longitudinal dataset was constructed in which students had both fall and spring scores. Two outcome measures were used: 1) the percentage of the Typical Growth measure achieved by each student and 2) whether or not each student achieved their growth measure.

Separate regression models were estimated for each grade level and subject (i.e., mathematics and reading), outcome measure, and predictor (i.e., English Learner status and ethnicity); the initial placement level was used as the control variable. After accounting for the initial placement level, the additional variance explained by the demographic variables was then computed. Across all models, English Learner status and ethnicity accounted for less than one percent and less than two percent, respectively, of the variance in the outcome measures.

Chapter 10: Projected Proficiency

10.1 Chapter Summary

Chapter 10 summarizes how *i-Ready Inform* produces score reports that allow for useful and timely decision making based on students' projected proficiencies. The Projected Proficiency reporting leverages data from linking studies conducted between *i-Ready Inform* and various state summative tests, along with the Typical and Stretch Growth measures, which are discussed at length in the preceding chapter. Data from a sample of states show that *i-Ready Inform* is highly correlated with state summative tests from across the country, with correlations generally exceeding .70. Additionally, linking studies result in strong classification accuracy between *i-Ready Inform* and the state summative tests and strong classification accuracy at aggregate levels.

10.2 Introduction

States put a large focus on measuring student performance using standardized summative assessments at the end of each school year. These tests not only assess the achievement of individual students, but they also assess the performance of schools and districts, which informs resource allocation. Student performance on state assessments is very important, and a common goal is to maximize the number of students who achieve proficiency status on state tests. To that end, it is valuable for educators to have early insights into whether students are on track to achieving proficiency throughout the school year.

The Projected Proficiency reporting in *i-Ready* utilizes the results of linking studies completed between *i-Ready Inform* and various state summative tests, in conjunction with the *i-Ready Inform* growth model. The following sections briefly describe (1) the equipercentile linking procedure used for the linking studies; (2) a summary of *i-Ready Inform*'s growth model; and (3) how the *i-Ready Inform* growth model and the equipercentile linking results are used to project student proficiency on state summative tests. Note that Projected Proficiency reporting is available only for select states for *i-Ready Inform* and is not available for the Growth Monitoring assessments.

10.2.1 Equipercentile Linking Procedure

Curriculum Associates regularly conducts linking studies with state summative tests. As of 2025, we have completed linking studies with 34 state tests, covering 42 states or territories. For states with existing linking studies, we rerun the analyses every three to five years or when there are significant changes to the state test that impact the generalizability of the existing linking study (e.g., change in the score scale or cut scores). For states without an existing linking study, we conduct a linking study once we have data that are sufficiently representative of the state.

Linking studies begin by partnering with school districts to collect student-level state test data. After collecting the data, the Curriculum Associates research team conducts standard data cleaning procedures to remove suspect data and outliers. Next, a stratified sample is selected for each subject (reading and mathematics) and grade level (3–8) such that the percentage of students in each state test achievement level reflects the population values, as reported by the given state for the given academic year. From the stratified

sample, Curriculum Associates then conducts equipercentile linkings for each subject and grade level using the methodology described below.

First, the frequency distributions of both *i-Ready Inform* and the state test are smoothed by fitting to them a series of polynomial log-linear models. A total of 10 models are fit for each subject, grade level, and test (*i-Ready Inform* or the state test), ranging from a one-parameter model to a 10-parameter model. Significance tests are conducted between adjacent models and the most parsimonious model with model fit (i.e., -2 log-likelihood) that is significantly better than the previous model in the series is selected as the model used for smoothing.

Next, percentile ranks are calculated on the smoothed distributions for each subject, grade level, and state test. Within each subject and grade level, *i-Ready Inform* scale scores are linked to state test scale scores with the same percentile rank in the smoothed distributions. The result is a scale-score-to-scale-score equipercentile linking between *i-Ready Inform* and the state test.

Using the equipercentile linking and publicly available cut scores for the state test achievement levels, we translate the state test achievement level ranges to the *i-Ready Inform* scale. As an example, Table 10.1 shows the crosswalk between *i-Ready Inform* scale scores and achievement levels on the Smarter Balanced Assessment (SBA).

Table 10.1. Spring *i-Ready Inform* to SBA Level Crosswalk

Subject	Grade	Level 1	Level 2	Level 3	Level 4
Reading	3	100–495	496–534	535–564	565–800
Reading	4	100–524	525–557	558–587	588–800
Reading	5	100–540	541–573	574–610	611–800
Reading	6	100–549	550–590	591–634	635–800
Reading	7	100–556	557–596	597–641	642–800
Reading	8	100–564	565–607	608–657	658–800
Mathematics	3	100–428	429–447	448–470	471–800
Mathematics	4	100–443	444–470	471–492	493–800
Mathematics	5	100–464	465–489	490–507	508–800
Mathematics	6	100–475	476–502	503–521	522–800
Mathematics	7	100–480	481–509	510–531	532–800
Mathematics	8	100–494	495–521	522–542	543–800

10.3 *i-Ready Inform* Growth Model

The *i-Ready Inform* growth model uses *i-Ready*'s criterion-referenced placement levels to set growth expectations based on subject, grade, and fall placement category. As described in Chapter 9, every student who takes *i-Ready Inform* in Grades K–8 receives two growth measures:

1. Typical Growth is the expected growth based on the median growth achieved historically by students in the same subject, grade level, and fall placement category.
2. Stretch Growth is an aspirational growth target designed to put students on a path toward proficiency.

See Chapter 9 for an in-depth discussion of the Typical Growth and Stretch Growth measures.

10.4 *i-Ready Inform* Projected Proficiency

The *i-Ready* Projected Proficiency report uses the state test achievement-level crosswalks and the *i-Ready* Typical Growth and Stretch Growth measures to project student performance on the EOY state test. Specifically, three projections are calculated based on different levels of projected growth in *i-Ready*:

1. Projection If Students Show No Additional Growth
2. Projection If Students Achieve Typical Growth
3. Projection If Students Achieve Stretch Growth

10.4.1 Projected Proficiency Based on No Additional Growth

The No Additional Growth scenario—available in the fall, winter, and spring testing windows [10]—uses the state test achievement-level crosswalk for the given state test to translate a student's most recent *i-Ready Inform* score into a state test achievement level. This scenario assumes no growth is demonstrated on *i-Ready Inform* between the most recent score and that on the EOY state test.

10.4.2 Projected Proficiency Based on Typical Growth

The Typical Growth scenario—available in the fall and winter testing windows—uses the Typical Growth measure to project a student's spring *i-Ready* score, then uses the state test achievement-level crosswalk to translate the projected spring *i-Ready* score into a state test achievement level. For *i-Ready Inform* assessments administered in the fall testing window, the entire Typical Growth measure is used to project the spring *i-Ready* score. For *i-Ready Inform* assessments administered in the winter testing window, the student's Typical Growth measure is prorated to 50% to project the spring *i-Ready* score.

10.4.3 Projected Proficiency Based on Stretch Growth

The Stretch Growth scenario—available in the fall and winter testing windows—follows the same logic as the Typical Growth scenario, except that the Stretch Growth measure is used in place of the Typical Growth

[10]The testing windows used for Projected Proficiency are the same as those used for *i-Ready*'s national norms: Fall: Beginning of year–November 15; Winter: November 16–March 1; Spring: March 2–end of year

measure. Table 10.2 summarizes how the Projected Proficiency model calculates projections across the fall, winter, and spring testing windows.

Table 10.2. Fall, Winter, and Spring Student-Level Proficiency Projections

Projection Scenario	Fall	Winter	Spring
Proficiency If Students Show No Additional Growth	i-Ready score translated to state test achievement level	i-Ready score translated to state test achievement level	i-Ready score translated to state test achievement level
Projection If Students Achieve Typical Growth	i-Ready score plus Typical Growth translated to state test achievement level	i-Ready score plus half of Typical Growth translated to state test achievement level	N/A
Projection If Students Achieve Stretch Growth	i-Ready score plus Stretch Growth translated to state test achievement level	i-Ready score plus half of Stretch Growth translated to state test achievement level	N/A

For any reporting group (e.g., class, school, district), aggregate projections are generated by calculating the frequencies in each state test achievement level over all students in the reporting group. Additionally, projected percent proficient is simply the percentage of students whose projected state test achievement level is in the range identified as "proficient" by the given state education agency.

10.5 Supporting Research

Projected Proficiency is supported by research conducted since 2016 demonstrating strong relationships between *i-Ready Inform* and more than 20 state summative tests. Table 10.3 summarizes the sample sizes and correlations between the spring *i-Ready Inform* administration and the state tests for a sample of state tests with completed linking studies as of 2023. The complete set of state test linking studies may be found on the Curriculum Associates website.

Table 10.3. *i-Ready* Linking Study Sample Size and Correlation Summary

***N* Count Range across Grades**

State Test (Year)	ELA	Mathematics
ACAP (2021–2022)	11,979–13,870	10,189–14,802
ILEARN (2018–2019)	1,669–3,337	1,713–4,494
MCAS (2021–2022)	2,463–3,099	2,776–3,398
SBA (2021–2022)	32,740–39,838	33,987–42,914

Note: ACAP = Alabama Comprehensive Assessment Program; ILEARN = Indiana Learning Evaluation Assessment Readiness Network; MCAS = Massachusetts Comprehensive Assessment System; SBA = Smarter Balanced Assessments (SBA states in study: CA, DE, OR, and WA)

Average Correlation across Grades (Range)

State Test (Year)	ELA	Mathematics
ACAP (2021–2022)	.84 (.80–.86)	.83 (.77–.87)

ILEARN (2018–2019)	.80 (.77–.81)	.88 (.86–.90)
MCAS (2021–2022)	.81 (.80–.82)	.88 (.86–.90)
SBA (2021–2022)	.83 (.81–.84)	.88 (.85–.89)

Note: ACAP = Alabama Comprehensive Assessment Program; ILEARN = Indiana Learning Evaluation Assessment Readiness Network; MCAS = Massachusetts Comprehensive Assessment System; SBA = Smarter Balanced Assessments (SBA states in study: CA, DE, OR, and WA)

10.5.1 Student-Level Classification Accuracy Analyses

To evaluate the ability of *i-Ready Inform* scores to classify students as Proficient or Not Proficient on the state test, classification accuracy analyses were conducted. Students were classified as *Proficient* or *Not Proficient* based on both their state test achievement level (observed proficiency) and their *i-Ready* score (projected proficiency). Based on these two classifications, students fit into one of the cells in Table 10.4.

Table 10.4. Student Classification Table

Projected Proficiency Based on *i-Ready Inform*	**Observed Proficiency on State Test is Not Proficient**	**Observed Proficiency on State Test is Proficient**
Not Proficient	True Negative	False Negative
Proficient	False Positive	True Positive

The following statistics were calculated based on students' projected and observed proficiencies:

- **Classification Accuracy:** The proportion of all students whose Projected Proficiency status from *i-Ready Inform* matched their Proficiency status on the state test
- **Sensitivity:** The proportion of students who were projected to be Proficient using *i-Ready Inform* to those who scored Proficient on the state test
- **Specificity:** The proportion of students who were projected to be Not Proficient using *i-Ready Inform* to those who scored Not Proficient on the state test

Table 10.5 summarizes the classification accuracy, sensitivity, and specificity analyses across subjects and grade levels for a sample of states with completed linking studies using the spring *i-Ready Inform* scores as of 2023.

Table 10.5. *i-Ready* Linking Study Classification Accuracy Summary

Mean Classification Accuracy across Grades (Range)

State Test (Year)	**ELA**	**Mathematics**
ACAP (2021–2022)	.85 (.84-.87)	.88 (.85–.91)
ILEARN (2018–2019)	.82 (.80-.84)	.87 (.86–.88)
MCAS (2021–2022)	.83 (.80-.84)	.87 (.86–.88)
SBA (2021–2022)	.84 (.83-.85)	.87 (.86–.88)

Mean Sensitivity across Grades (Range)

State Test (Year)	ELA	Mathematics
ACAP (2021–2022)	.85 (.82–.88)	.80 (.73–.87)
ILEARN (2018–2019)	.82 (.79–.86)	.86 (.79–.91)
MCAS (2021–2022)	.78 (.76–.81)	.84 (.81–.88)
SBA (2021–2022)	.84 (.83–.85)	.84 (.81–.87)

Mean Specificity across Grades (Range)

State Test (Year)	ELA	Mathematics
ACAP (2021–2022)	.85 (.79–.87)	.91 (.84–.95)
ILEARN (2018–2019)	.82 (.76–.87)	.87 (.82–.92)
MCAS (2021–2022)	.85 (.80–.87)	.89 (.87–.92)
SBA (2021–2022)	.83 (.79–.86)	.89 (.84–.91)

10.5.2 Aggregate Accuracy Analyses

To evaluate the accuracy of the school-level percent proficient projections based on the spring *i-Ready Inform* assessment, the projected percent proficient was calculated at the school by subject by grade level, and those aggregate projections were compared to the actual percent proficient among the same sample of students. Because this analysis uses *i-Ready* scores and state test scores for a matched sample of students, results may indicate greater aggregate classification accuracy than what one would observe when comparing the *i-Ready* Projected Proficiency report to the report released by the state education agency, where the two samples may contain different sets of students.

Table 10.6 shows the percentages of schools by subject by grade-level groups that were under-, accurately, and over-projected for a sample of states with completed linking studies using the spring *i-Ready Inform* scores as of 2023.

Table 10.6. *i-Ready* Linking Study Aggregate Projected Proficiency Accuracy Summary

Number of Schools

State Test (Year)	ELA	Mathematics
ACAP (2021–2022)	259	261
ILEARN (2018–2019)	200	261
MCAS (2021–2022)	60	65
SBA (2021–2022)	746	780

Percentage Under-Projected

State Test (Year)	ELA	Mathematics
ACAP (2021–2022)	4%	5%
ILEARN (2018–2019)	11%	13%

MCAS (2021–2022)	5%	4%
SBA (2021–2022)	9%	5%

Percentage Accurately Projected

State Test (Year)	ELA	Mathematics
ACAP (2021–2022)	89%	86%
ILEARN (2018–2019)	83%	80%
MCAS (2021–2022)	82%	84%
SBA (2021–2022)	84%	88%

Percentage Over-Projected

State Test (Year)	ELA	Mathematics
ACAP (2021–2022)	7%	9%
ILEARN (2018–2019)	7%	7%
MCAS (2021–2022)	12%	12%
SBA (2021–2022)	7%	7%

Note: *Projections were defined as Accurately if the percent projected to score Proficient by *i-Ready* was within +/- 10 points of the actual percent proficient. Under-projections were cases in which the *i-Ready* projection was more than 10 points too low. Over-projections were cases in which the *i-Ready* projection was more than 10 points too high.

i-Ready's Projected Proficiency reporting leverages data from linking studies conducted between *i-Ready Inform* and various state summative tests, along with *i-Ready*'s Typical Growth and Stretch Growth measures. The result is projections of student-, class-, school-, and district-level state test performance that emphasize how students' performance on the state summative test will be highly dependent on the growth they show between the time of their *i-Ready* administration and the spring state summative test administration.

10.6 Summary

Data from the exemplar states show that *i-Ready Inform* is highly correlated with state summative tests from across the country, with correlations generally exceeding .70. Additionally, the linking studies result in strong classification accuracy between *i-Ready* and the state summative tests, with classification accuracy of at least 80 percent in states where Projected Proficiency reporting is enabled. Finally, the accurate student-level classifications tend to result in strong accuracy at aggregate levels, with 80 percent or more of school-level projections within +/-10 percentage points of the observed percent proficient.

Chapter 11: Score Reports

11.1 Chapter Summary

Score reports are critical components of the overall assessment system because they are the primary means through which the meaning and intended use of scores is communicated to the various stakeholders who support learning. This chapter presents some of the reports for the *i-Ready Inform* and Growth Monitoring assessments that are available to teachers and administrators in the online reporting system. These reports provide a wealth of descriptive information about students' current levels of proficiency as well as prescriptive information about how to enhance learning. This chapter also discusses the interpretive guidance provided to users within the reports as well as resources external to the reports to support valid inferences from assessment results.

11.2 Introduction

Just as test developers must have clear procedures and rationale for constructing interpretable score scales, the Standards for Educational and Psychological Testing (AERA et al., 2014) require test developers to incorporate clear explanations of intended score uses and interpretations throughout reporting practices. The Standards identify in Standard 5.1 the importance of clearly explaining test scores to users, including the "characteristics, meaning, and intended interpretation of scale scores, as well as their limitations."

11.2.1 Levels of Reporting

The *i-Ready* reporting platform presents all data in practical, intuitive reports that educators can access in real time through their *i-Ready* dashboard. The dashboard allows reports to be printed—though we will display the online versions here—and may provide data exporting functionality. Depending on the nature and purpose of the report, they are accessible at one or more of the following levels:

- **Student-level reports** are designed primarily to communicate the results of an individual student to a teacher, although administrators may access them as well. These reports are detailed and contain clickable features to enhance understanding. Detailed, easy-to-read descriptions of the scores, placement levels, and other indicators of achievement are provided. The For Families report is specifically designed to put student results into lay terms to help facilitate teacher–family conversations as well as to provide appropriate instructional resources.
- **Class-level reports** are designed for teachers and other administrators to understand the collective results of a class. Thus, the main objective of these reports is to provide class-level aggregations of student-level results. Additionally, these reports provide instructional resources that may be tailored to the class-level results.
- **School- and district-level reports** are designed for administrators to understand the collective results of a grade, school, or district. Administrators can view various combinations of aggregations that are of interest, thereby facilitating their instructional and resource planning. Additionally, these reports provide instructional resources that may be tailored to the specified level of results.

While this chapter is not intended to be an exhaustive repository of our score reports, we will present the core components of our main reports in the sections that follow. To organize this chapter, score reports are grouped into three categories: *i-Ready Inform* results reports, reports based on multiple administration, and reports providing specific action steps. We begin by showing portions of the student, class, school, and district levels for the *i-Ready Inform* results reports to show readers the various components of the reports and levels of aggregations. However, in subsequent sections, we will limit the presentation to the student-level (where applicable) and the class-level reports for the sake of readability.

11.3 *i-Ready Inform* Results Reports

11.3.1 Student Level

The student-level report presents criterion-referenced information about the student's performance for the content area and corresponding domains as well as norm-referenced information about the student's overall performance. Additionally, the report provides skill-based information that is intended to inform instruction.

The report is delivered online and is dynamic, thereby allowing teachers to easily navigate from one student report to another using a dropdown list. Teachers can also navigate across content areas or administrations by selecting a different subject or *i-Ready Inform* administration for a given student. Scores and other reported quantities as well as accompanying ancillary information in the report will be discussed next.

11.3.1.1 Purpose of the Report

Upon completion of *i-Ready Inform*, the in-depth Inform Results report provides scores, placements, and detailed insights into student performance. This report is available at the student, class, and school/district levels.

At the student level, the purpose of the report is to provide the teacher with a comprehensive view of each student's performance for a given *i-Ready Inform* administration. At the class and school/district levels, the purpose of the report is to summarize the distribution of overall and domain placements.

This report is intended to be used by teachers and administrators; students have access to their scale scores for each completed *i-Ready Inform* assessment in their *i-Ready* dashboard, but they are not shown their placement level. Curriculum Associates intends for teachers to share assessment results with students in the context of a "data chat" (see the Theory of Action in Chapter 1).

11.3.1.2 Overall Scale Score and Growth Measures

The overall scale score on *i-Ready Inform* quantifies student proficiency in a content area and locates the student on the Grades K–12 vertical scale (see Chapter 5 for details) for the respective content area. This report, a portion of which is displayed in Figure 11.1, presents student performance on *i-Ready Inform* both numerically and graphically. The graphical representation shows the overall scale score achieved for the assessment in the form of a bar graph so that it can be viewed in relation to the range of scores constituting on-grade level performance, the Mid Grade Level cut, and the Typical Growth and Stretch Growth measures established for the academic year based on the student's baseline *i-Ready Inform* score, which is generally the first valid *i-Ready Inform* assessment completed within the school year. With the completion of subsequent *i-*

Ready Inform assessments, the scale score chart also presents the change in scale score since the baseline *i-Ready Inform* assessment.

Adjacent to the graphical presentation of the overall scale score is a box with the definition of Typical Growth and Stretch Growth (see Chapter 9 for details). Teachers can click on the information icons for additional information. The teacher can also follow a hyperlink to Curriculum Associates' *i-Ready Central* website to learn more about *i-Ready*'s approach to measuring student growth, which further supports appropriate interpretations of test results.

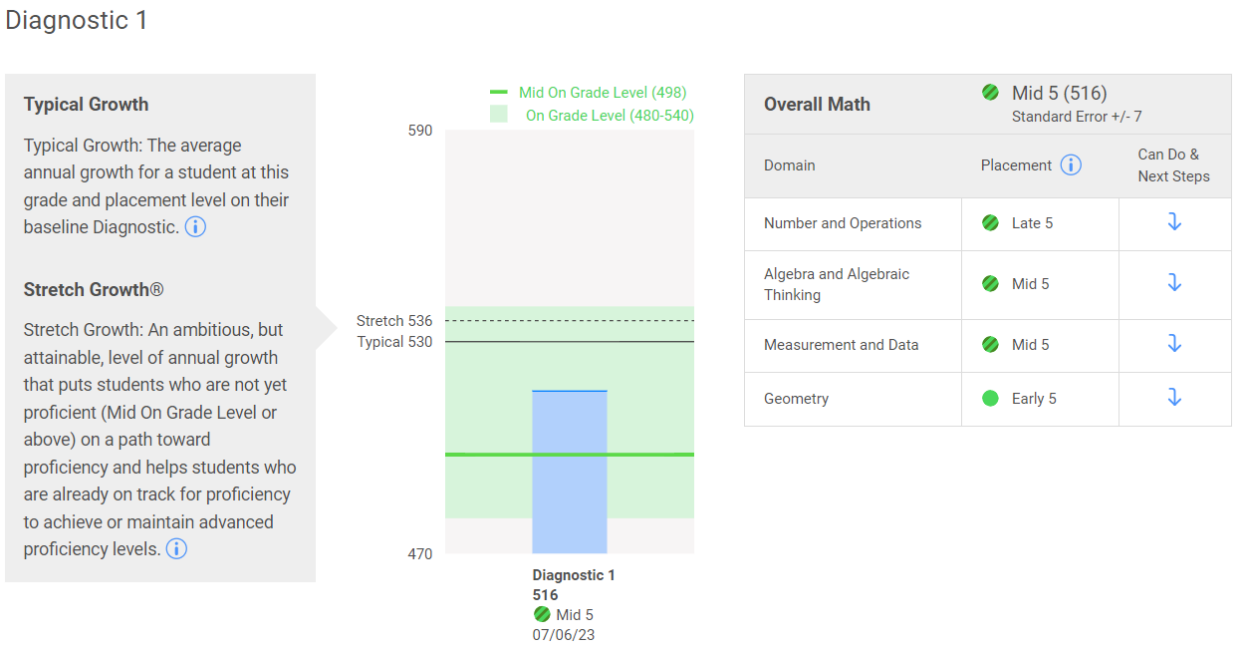


Figure 11.1. Inform Results Report, Student Level: Scale Score and Placements

11.3.1.3 Overall Placements, Domain Scale Scores, and Domain Placements

Students' overall and domain placement levels are reported in a table at the top of the report, as shown on the right of Figure 11.1. Adjacent to the overall placement level is the overall scale score (identical to the value in the bar chart) and its SEM (see Chapter 6 for a detailed discussion on this statistic).

The placement levels are criterion referenced and serve as a means of interpretation of student performance relative to grade-level expectations based on the content standards for the grade. Students whose overall and domain scores fall into the grade-level range for their grade will be assigned one of the on-grade level placements. Alternatively, students scoring at a level above or below their grade will be assigned one of those placement-level designations for the grade-level range in which their score is located. While there are many above- and below-grade level distinctions, they can be collapsed into Above Grade Level or Below Grade Level or into the number of grade levels above or below. The placement-level table also provides interpretive support via various clickable features.

In the lower portion of the report, domain placement levels are reported alongside their corresponding domain scale scores, as shown in Figure 11.2. An information icon next to the domain titles provides a pop-up box that explains the meaning of these terms in the placement-level column and how this informs instruction.

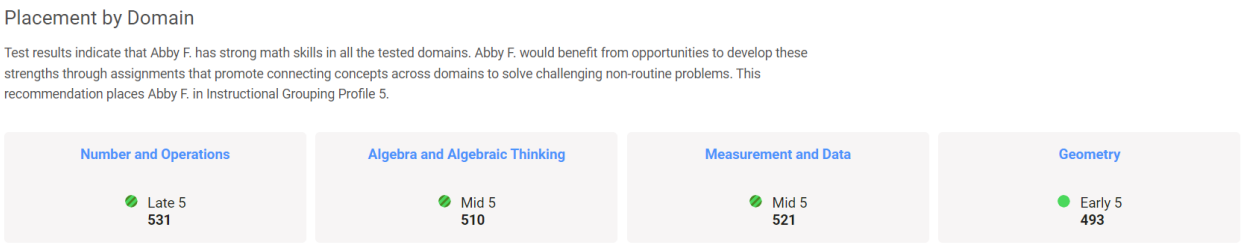


Figure 11.2. Inform Results Report, Student Level: Domain Scores and Placements

11.3.1.4 National Norms

In addition to the scale scores and criterion-referenced placement levels, reports include normative information in the form of percentile ranks. Those values reflect students' overall scale scores relative to a nationally representative sample of peers who took *i-Ready Inform* at the same time of year and in the same grade level, as shown on the left of Figure 11.3. The teacher can click on the information button next to the percentile score to view a pop-up box with an explanation of the meaning of the percentile score. Refer to Chapter 8 for more details on the development of the national norms.

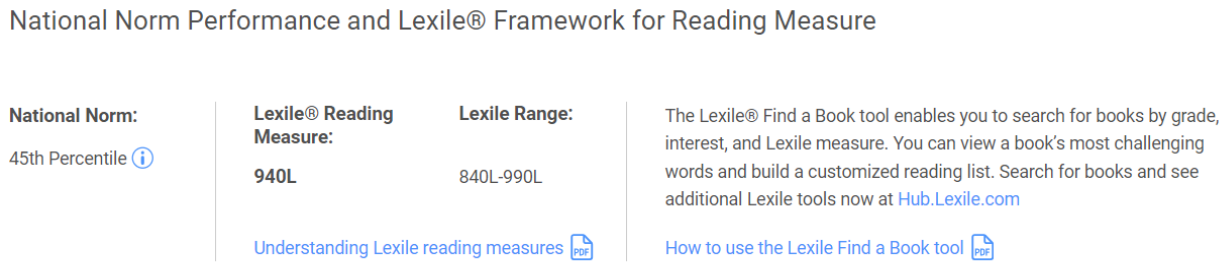


Figure 11.3. Inform Results Report, Student Level: Norm Percentile and Lexile (Reading Only)

11.3.1.5 Lexile® and Quantile® Measures

i-Ready Inform also provides Lexile[11] measures for reading and Quantile[12] measures for mathematics, which are developed by MetaMetrics, and are shown in Figure 11.3. Note that Figure 11.3 shows a sample report in reading, so only Lexile results are displayed. Lexile scores provide an estimation of a test-taker's reading ability given the complexity of a reading passage. Lexiles can help teachers, parents, and students locate appropriate reading materials. Quantile scores, which are also developed by MetaMetrics, are indicators of the

[11]Lexile® measures and the Lexile Framework for Reading are trademarks of MetaMetrics, Inc., and are registered in the United States and abroad. Copyright © 2025 MetaMetrics, Inc. All rights reserved.
[12]Quantile® measures and the Quantile Framework for Mathematics are trademarks of MetaMetrics, Inc., and are registered in the United States and abroad. Copyright © 2025 MetaMetrics, Inc. All rights reserved.

mathematics skills and concepts a student has likely learned and those which require further instruction. Both the Lexile and Quantile scores are reported as a single value and a corresponding range.

Within the online report, the teacher can click on hyperlinks to view a one-page reference guide explaining the meaning and purpose of Lexile and Quantile scores or to link to MetaMetrics's hub for more resources regarding the interpretation and use of Lexiles and Quantiles to support instruction.

11.3.1.6 Can Dos and Next Steps

Can Do statements ("Can Dos") and Next Steps provide actionable information at the skill level to immediately inform instruction, as shown on the right side of Figure 11.6. *Can Dos* are statements about what skills a student can likely do based on their domain proficiency level. *Next Steps* are recommended focus areas for skills that may not yet have been mastered. They are based on the finer-grained skills that are typically, although not exclusively, associated with a subject–grade–domain combination; these skills are referred to as *indicators* (see Chapter 2).

The difficulties of the items associated with each indicator are used to construct indicator characteristic curves (i.e., a test characteristic curve for only the items associated with the indicator) that relate student proficiency to the expected number of items answered correctly for each indicator. These curves are used to establish the difficulty of an indicator. See Figure 11.4 and Figure 11.5 for the current (i.e., academic year 2024–2025) distribution of indicator difficulties by subject, grade, and domain.

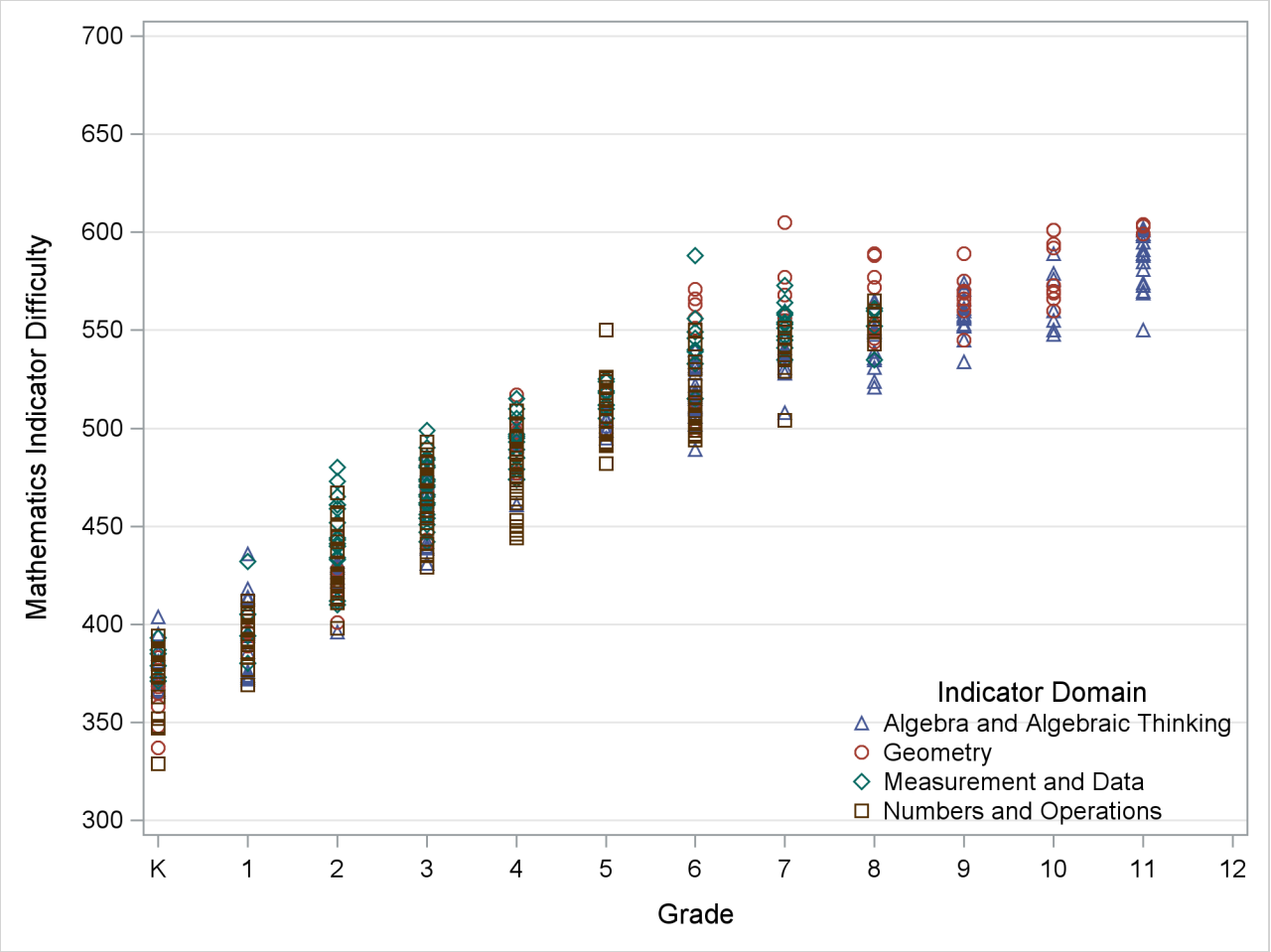


Figure 11.4. Mathematics Indicator Difficulty Distribution by Grade and Domain

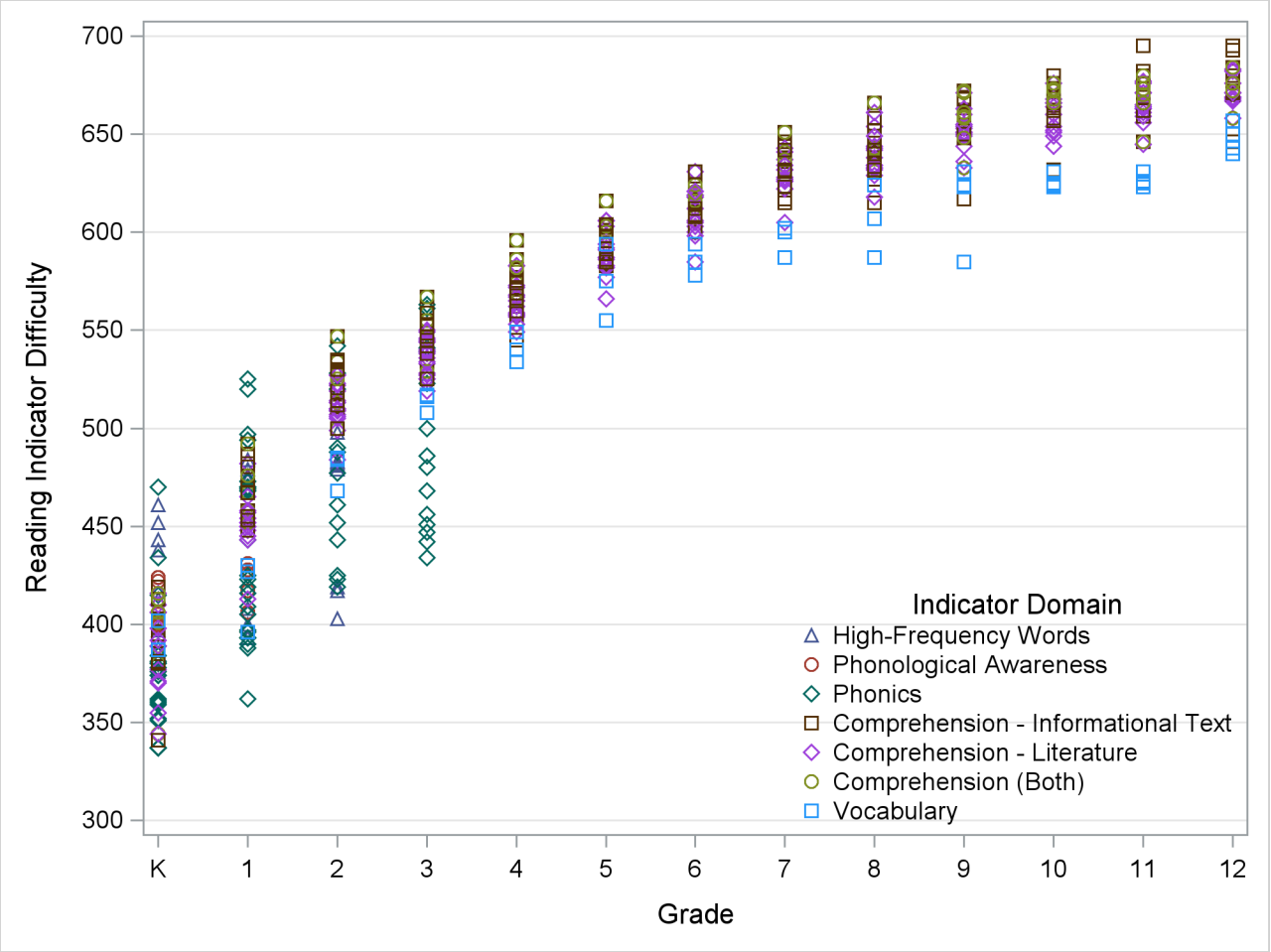


Figure 11.5. Reading Indicator Difficulty Distribution by Grade and Domain

Students' domain placements are compared to the difficulty of the indicator, provided that the number of items assigned to the indicator is adequate, to determine whether it is probable that they can demonstrate the skill represented by the indicator. If a student's domain proficiency is sufficiently high relative to the indicator difficulty, the skill may be considered one that a student likely can do and may be labeled as such on the score report. Conversely, if a student's domain proficiency does not meet this criterion, the skill may be labeled as a Next Step. Note that skills are prioritized based on grade and placement level, so the eligible skills for Can Dos and Next Steps reporting will vary.

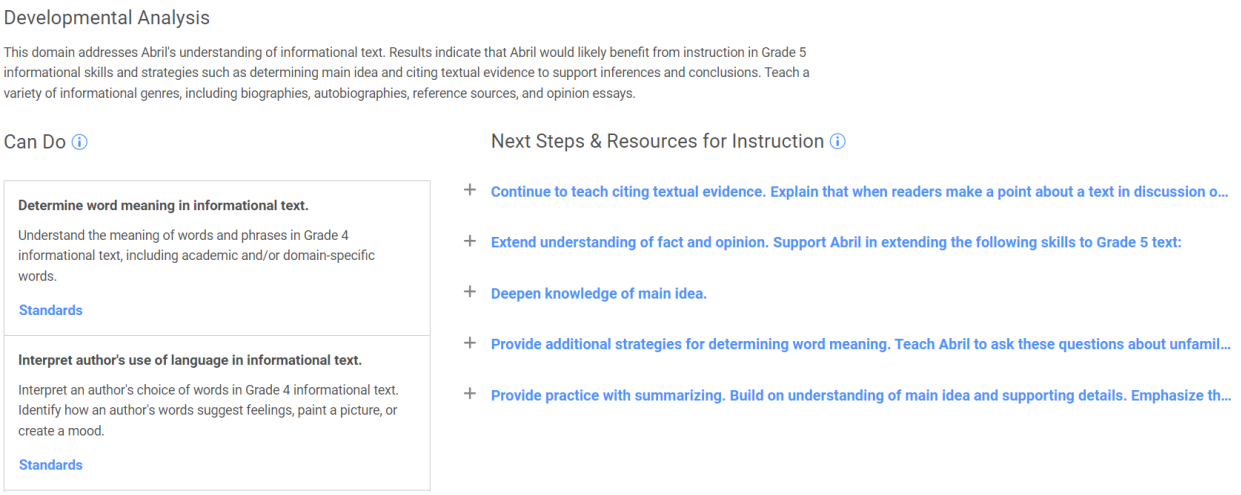


Figure 11.6. Inform Results Report, Student Level: Can Do Statements, Next Steps, and Additional Resources

11.3.1.7 Instructional Guidance and Resources

Instructional guidance based on the student's performance in relation to the grade-level standards for a domain is accessed by clicking on the title for each domain in the Placement by Domain section of the online report. The instructional guidance for each domain provides a Developmental Analysis summarizing the student's performance in the domain and areas of instructional readiness or needs.

The Can Dos and Next Steps column in the placement-level table (Figure 11.1) provides a down-arrow hyperlink that takes the user farther down in the report to the Developmental Analysis section that provides individualized Can Dos and Next Steps and resources for instruction. These links help ensure that report users are connected to helpful instructional guidance.

11.3.2 Class Level

The class-level report is available to teachers and administrators, and it is designed to summarize the performance of students in a class. The top of the report (Figure 11.7) presents a view of the class's overall performance, which shows the number and percentage of students in each placement level according to their overall performance.

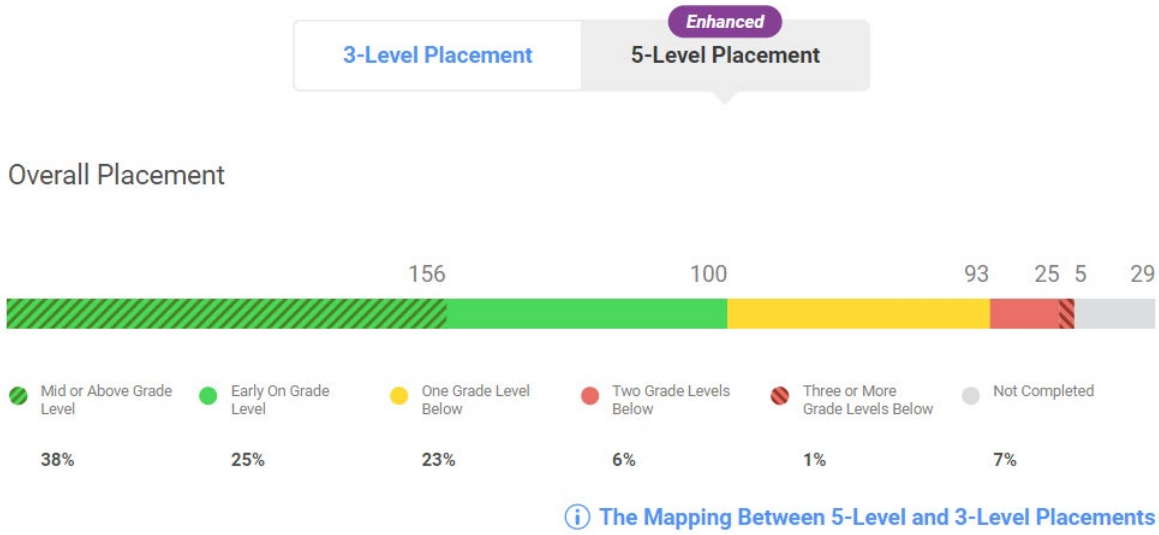


Figure 11.7. Inform Results Report, Class Level: Overall Placement Level Distribution

Next, placement-level distribution information is displayed at the domain level, as shown in Figure 11.8.

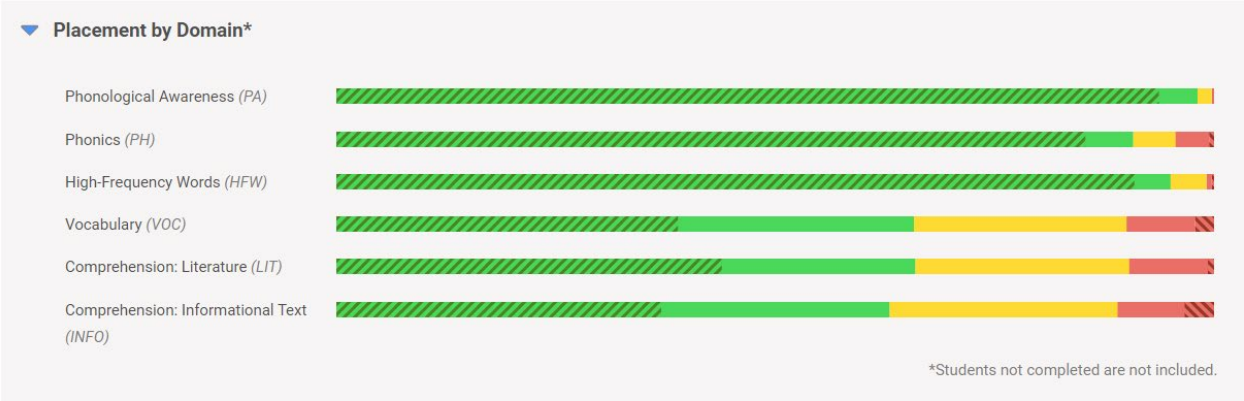


Figure 11.8. Inform Results Report, Class Level: Domain Placement-Level Distribution

Finally, there is a dynamic view, shown in Figure 11.9, in which placements by domain are shown for each student along with several columns in which the quantity to display can be specified by the user (e.g., Typical Growth and Stretch Growth goals, national percentile, or Lexile/Quantile, and date of the *i-Ready Inform* administration). Users can also click on a yellow or red Rush alert for a pop-up box with an explanation of the alert.

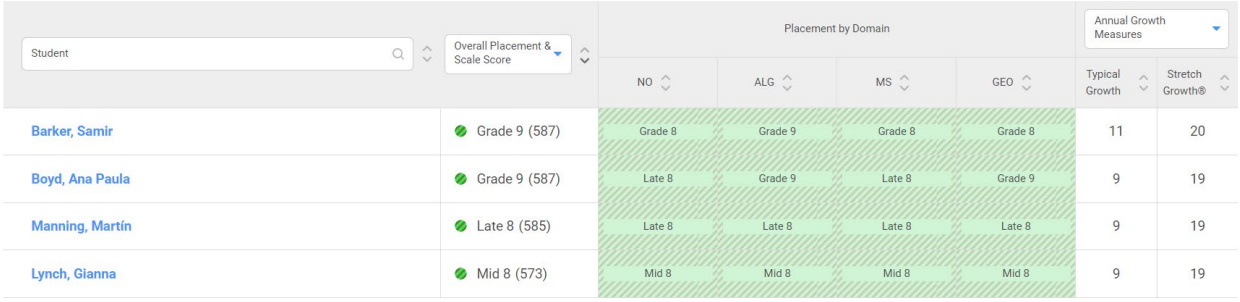


Figure 11.9. Inform Results Report, Class Level: Dynamic and Customizable Table of Student Results

11.3.3 School and District Levels

Administrators can view results aggregated to the school and district levels. Figure 11.10 shows a sample school-level report in which results are aggregated to the grade level. Overall and domain placement-level distributions across all grades, similar to those shown in Figure 11.7 and Figure 11.8, are also available.

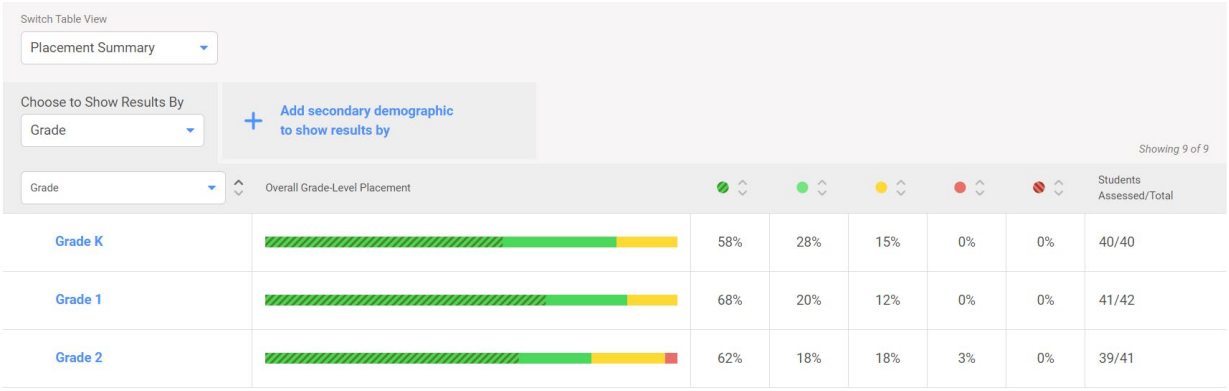


Figure 11.10. Inform Results Report, School Level: Dynamic and Customizable Table of Results

District-level reports, like that shown in Figure 11.11, are also available. These reports aggregate results to the school level. Overall and domain placement-level distributions across all schools in the district, similar to those shown in Figure 11.7 and Figure 11.8, are also available.

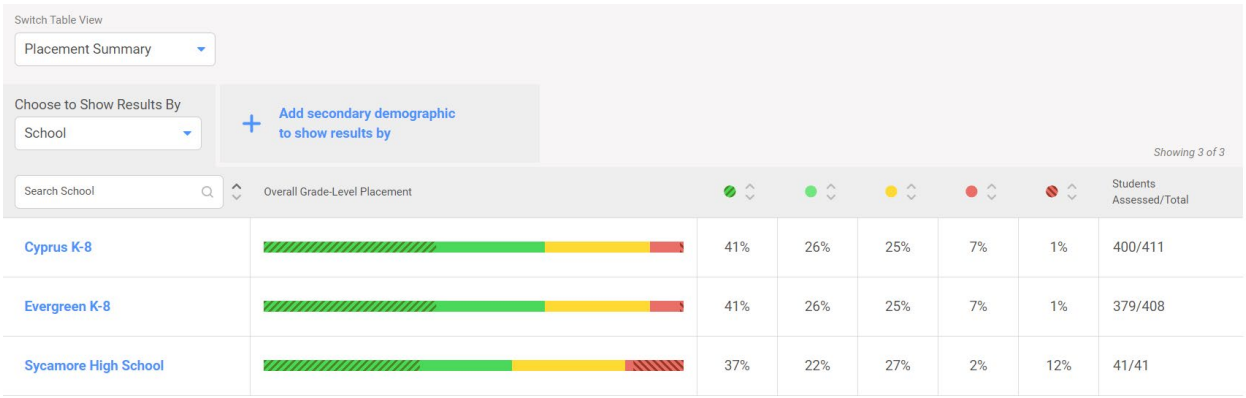


Figure 11.11. Inform Results Report, District Level: Dynamic and Customizable Table of Results

11.4 Reports Highlighting Growth

11.4.1 For Families Report

The For Families report presents a summary of a student's performance on *i-Ready Inform* taken within a school year. The report assists teachers when having conversations about a student's performance in the context of the specific domains and skills the student is mastering or needs support, the student's goals, and actions for families to support the student at home.

The report is available in both English and Spanish. Teachers can print the For Families report and use it to share *i-Ready Inform* scores with parents. Optionally, the For Families report can be enabled on the student dashboard, thereby allowing districts to distribute the report efficiently.

Figure 11.12 and Figure 11.13 show the key features of the For Families reports. In Figure 11.12, the summary graph provides school year-to-date *i-Ready Inform* performance for up to three tests as well as the grade-level band for appropriate contextualization. For each administration, the scale score and placement levels are presented. Student growth from the baseline *i-Ready Inform* assessment to each subsequent *i-Ready Inform* assessment is also shown. The domain table on the right of Figure 11.12 lists details by domain assessed for up to two *i-Ready Inform* assessments—the baseline and most recent administrations—thereby allowing users to see the domains in which the student is excelling or needs additional support as well as their growth by domain during the year.

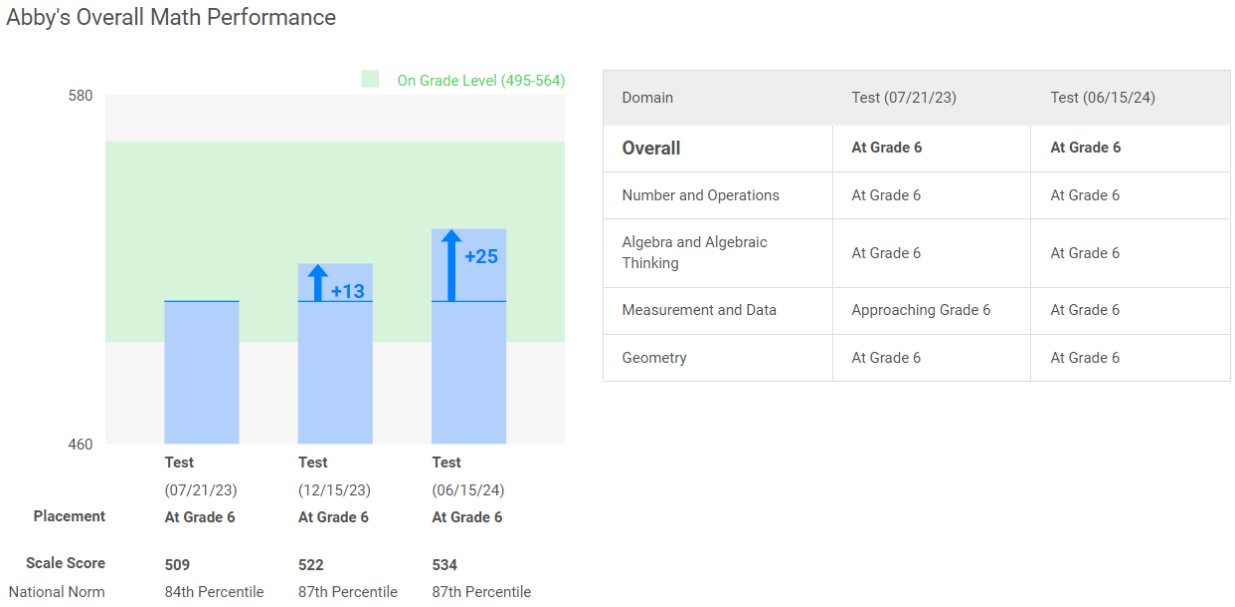


Figure 11.12. Partial View of Page 1 of the For Families Report in Reading

The report ends with a section titled Skill Progress and More Information (Figure 11.13). This section shows placement levels by domain accompanied by additional information, which may include specific examples of skills the student likely has mastered and what should be focused on in the future.

Skill Progress and More Information

↑ Placement Improved from Baseline

Domain	Test (07/21/23)	Test (06/15/24)	More Information
Number and Operations	At Grade 6	At Grade 6	At placement levels 6-8 this domain addresses operations with whole numbers, fractions, decimals, and positive and negative rational numbers, as well as exponents. Abby's score indicates an appropriate understanding of rational number concepts. Abby may be ready to compute with rational numbers and to use absolute value in real-world situations.
Algebra and Algebraic Thinking ↑	At Grade 6	At Grade 6	At placement levels 6-8 this domain addresses ratios and proportional relationships, expressions, equations and inequalities, and functions. Abby's score indicates an appropriate understanding of basic ratio, expression, and equation concepts. Abby may be ready to use expressions, equations, and inequalities to represent and solve problems.
Measurement and Data ↑	Approaching Grade 6	At Grade 6	At placement levels 6-8 this domain addresses probability concepts and statistical analysis of data. Abby's score indicates a basic understanding of angle measurement, line plots, and area. Test results indicate Abby may be ready for further instruction and practice converting between customary and metric units, determining measures of center and variability, and displaying and interpreting data in dot plots.

Figure 11.13. Partial View of Page 2 of the For Families Report in Mathematics

11.4.2 *i-Ready Inform* Growth Report

11.4.2.1 Student Level

The student-level report, which is available to teachers and administrators, provides detailed information about how the student has performed on each *i-Ready Inform* assessment and how they are progressing toward EOY Typical Growth and Stretch Growth measures.

Figure 11.14 shows the portion of this report pertaining to overall results. On the left, progress toward the growth measures are provided in terms of a percentage of the growth measure realized at the time of the latest administration. In addition, the report provides an estimate of the expected time to proficiency goals for the student to attain their Stretch Growth measure. On the right, the results for up to three administrations are shown. The student's observed score gains from the first to the second and third scores, if applicable, are shown in relation to Typical Growth and to Stretch Growth measures.

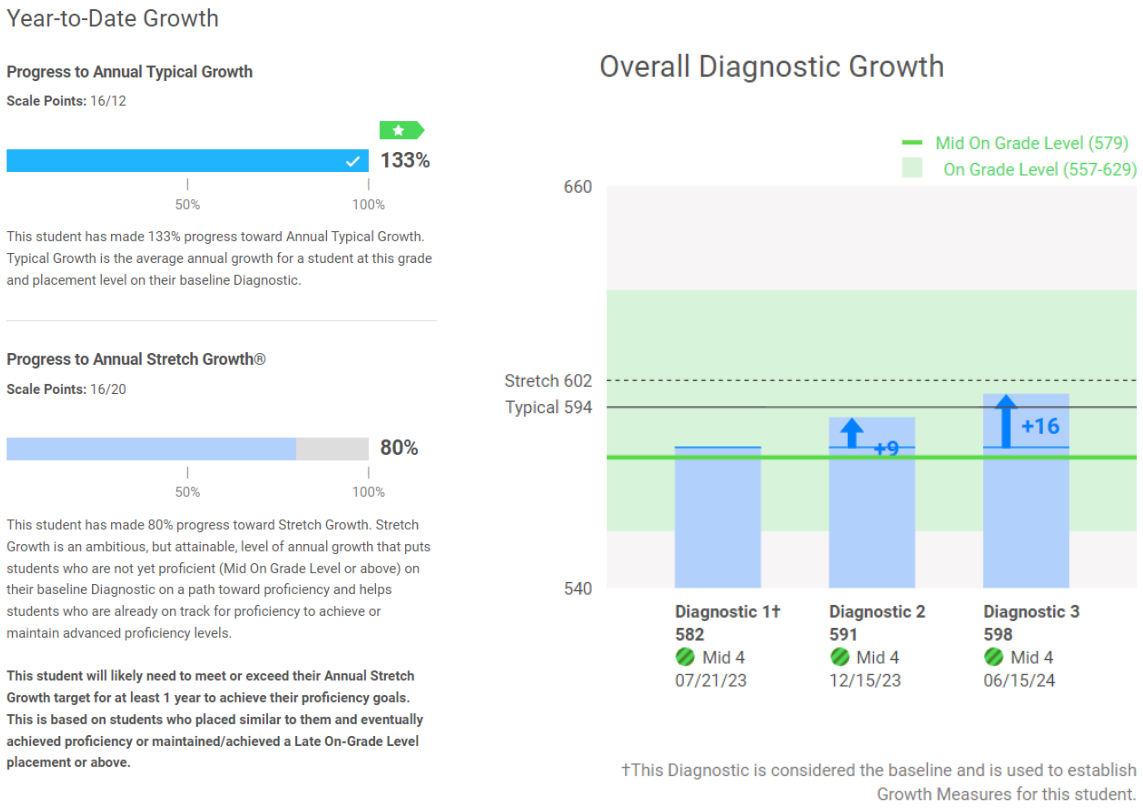


Figure 11.14. Inform Growth Report, Student Level: Overall Results

Figure 11.15 shows the overall and domain placement-level designations for up to three administrations. Domains in which there has been improvement over the baseline are denoted with a green upward-facing arrow.

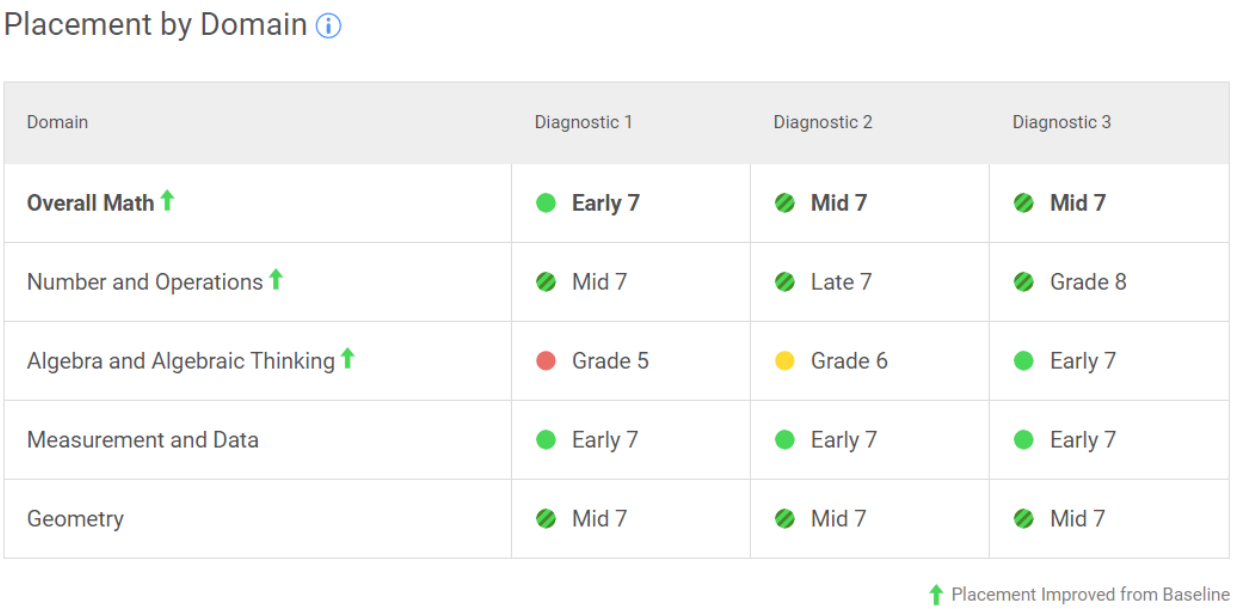


Figure 11.15. Inform Growth Report, Student Level: Domain Results

11.4.2.2 Class Level

The class-level reports, available to teachers and administrators, provide aggregate summaries of growth results at the class level. Administrators can also see aggregations at the grade, school, and district levels. Figure 11.16 shows the class's progress toward Typical Growth calculated as the median of the students' progress toward Typical Growth as well as the current placement-level distribution. Figure 11.17 shows the portion of the report that describes the class's progress toward Typical Growth and Stretch Growth as the distribution of progress toward the growth measure grouped by levels of progress. Figure 11.18 shows the portion of the report that presents each student's results in a table. Student-level reports (see Section 11.4.2.1) can be accessed by clicking on students' names.

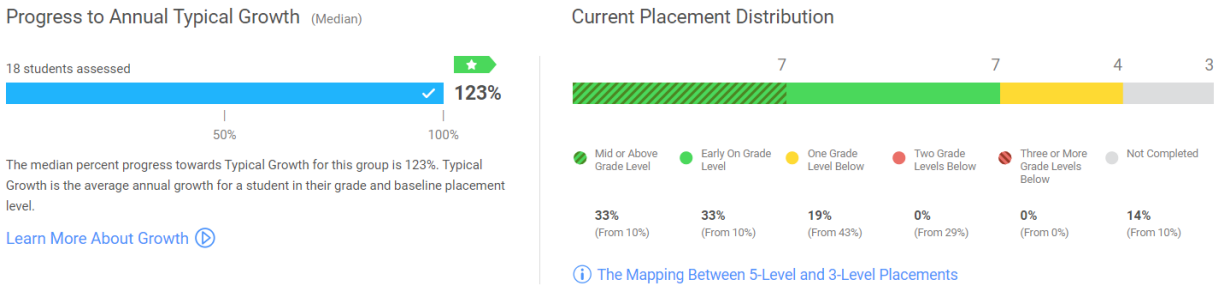


Figure 11.16. Inform Growth Report, Class Level: Growth and Placement Information

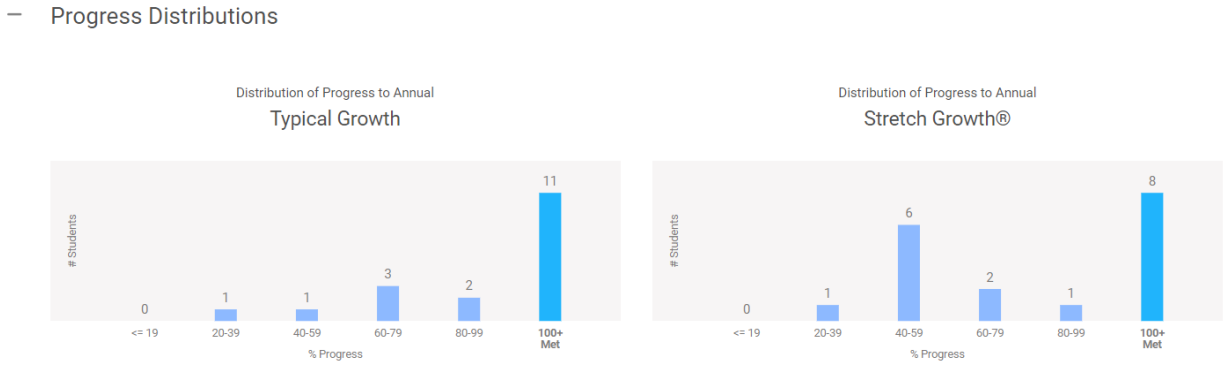


Figure 11.17. Inform Growth Report, Class Level: Growth Distributions

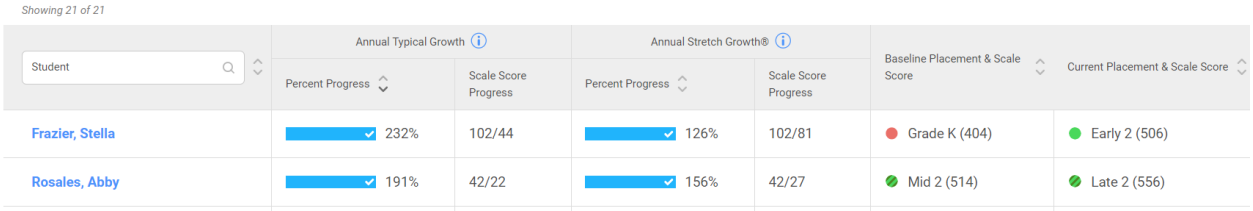


Figure 11.18. Inform Growth Report, Class Level: Student Detail

11.4.3 Growth Monitoring Report

11.4.3.1 Student Level

Growth Monitoring assessments are mini-assessments in the *i-Ready* suite that can be administered multiple times between administrations of *i-Ready Inform* to allow educators to monitor the progress of students to determine whether they are on track to meet Typical Growth and Stretch Growth targets by the end of the year. The Growth Monitoring assessments are adaptive and use a subset of the *i-Ready Inform* item bank. By using the data available in Growth Monitoring reports, teachers can adjust their instructional strategies or change the level of intervention to ensure students stay on track to meet their growth targets.

As students complete the *i-Ready Inform* and Growth Monitoring assessments during the year, *i-Ready* analyzes how their growth is trending and projects progress toward growth measures and the level of achievement students may reach by the end of year. Student-level reports, pictured in Figure 11.19, show the likelihood of a student meeting 100 percent of the Typical Growth and Stretch Growth measures by the end of the year and the projected EOY score in scale score points based on the student's score trend compared to the scale-score point increase needed to be on track.

The class-level report shows the number of students in the class who are likely, somewhat likely, or somewhat unlikely to meet the annual Typical Growth or Stretch Growth goals by the end of the year. The teacher can then drill down into each student's report.

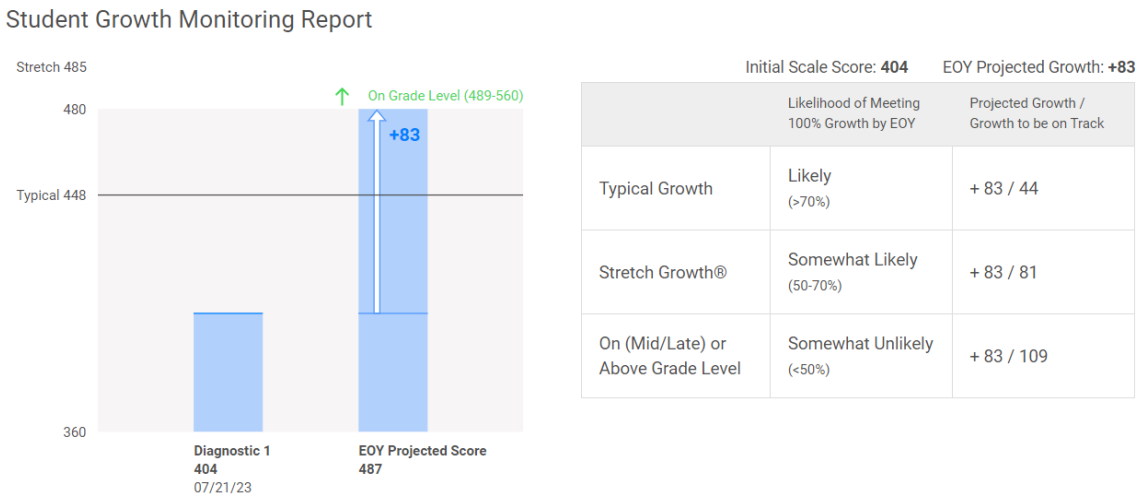


Figure 11.19. Growth Monitoring Report: Student Level

11.4.3.2 Class Level

The class-level Growth Monitoring reports show the distribution of the likelihood of growth measure attainment (Figure 11.20). Figure 11.21 shows a table view of individual student-level results, which are accessible by clicking on the student names.

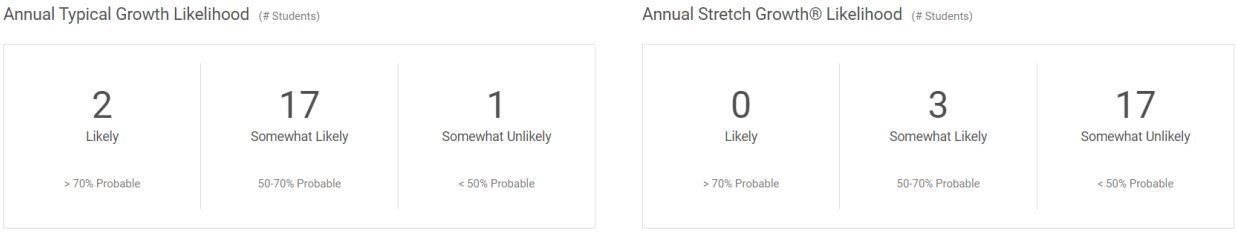


Figure 11.20. Growth Monitoring Report: Class Level

Showing 21 of 21

Student	Annual Typical Growth		Annual Stretch Growth®		On (Mid/Late) or Above Grade Level	
	Likelihood	Projected Growth/ Growth Measure	Likelihood	Projected Growth/ Growth Measure	Likelihood	Projected Growth/ Growth Measure
Combs, Natalie	Somewhat Likely	33/33	Somewhat Unlikely	33/63	Somewhat Unlikely	33/91
Do, Brian	Likely	43/26	Somewhat Likely	43/40	Somewhat Unlikely	43/56
Farrell, Alvaro	Likely	52/26	Somewhat Likely	52/40	Somewhat Unlikely	52/56

Figure 11.21. Growth Monitoring Report: Class Level—Individual Student Results

11.4.4 Projected Proficiency Reports

i-Ready's available Projected Proficiency reporting feature (for Grades 3–8) leverages the strong association between *i-Ready Inform* and a state's summative assessment scores as well as *i-Ready*'s growth model (see Chapter 9).

Projected Proficiency reports provide estimates of the percentages of students who are expected to achieve each level of proficiency on the state assessment based on their performance on the fall or winter *i-Ready Inform* assessment. These reports are available for states in which a linking study between *i-Ready Inform* and the state assessment has been conducted (see Chapter 10). Projected Proficiency reports at the beginning and middle of the year provide estimates of students' likely levels of proficiency on the state's summative assessment at the end of the year based on three scenarios:

- Students attain No Additional Growth after the *i-Ready Inform* assessment
- Students achieve Typical Growth after the *i-Ready Inform* assessment
- Students achieve Stretch Growth after the *i-Ready Inform* assessment

Class-level reports at the beginning and middle of the year provide estimates of each student's likely level of proficiency on the state's summative assessment at the end of the year, given the selected growth scenario.

School- and district-level reports provide estimates of the approximate percentage of students who are likely to place in each state test performance level based on the three scenarios of growth described above, which are displayed in Figure 11.22, Figure 11.23, and Figure 11.24, respectively. Leaders can configure reports to examine aggregates by grade level across schools or within a school and by growth scenario.

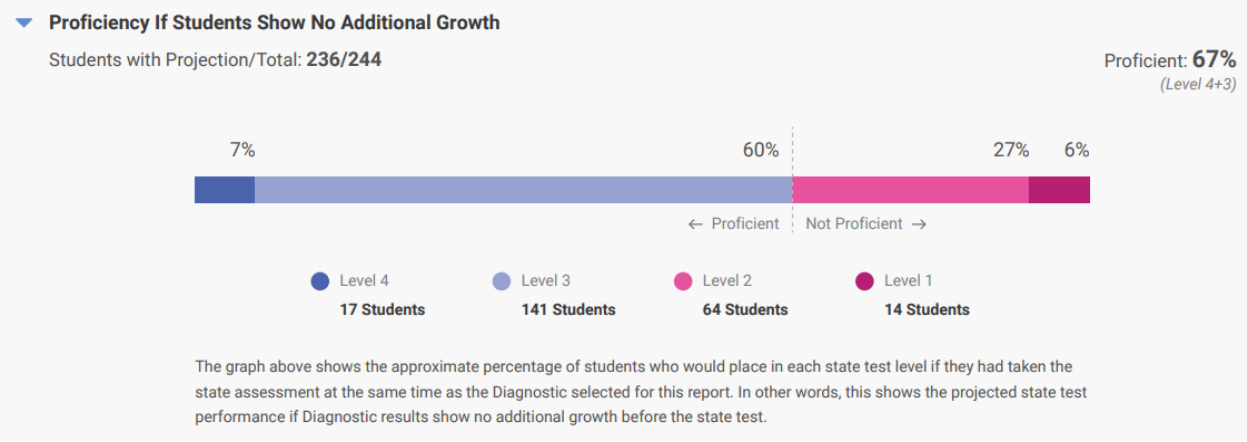


Figure 11.22. Projected Proficiency Report: No Additional Growth

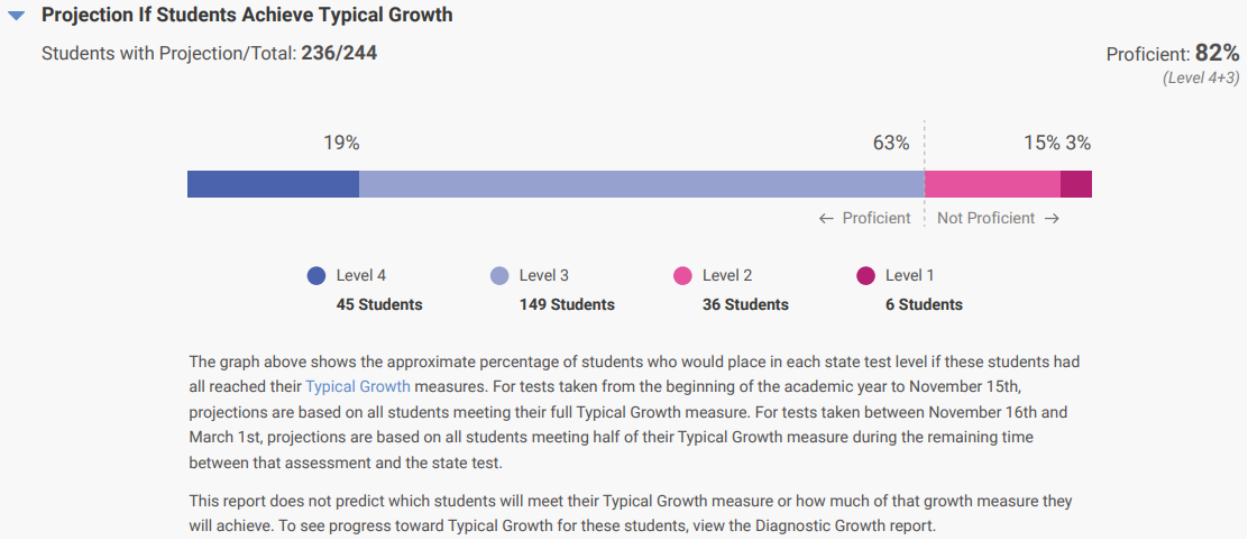


Figure 11.23. Projected Proficiency Report: Achieve Typical Growth

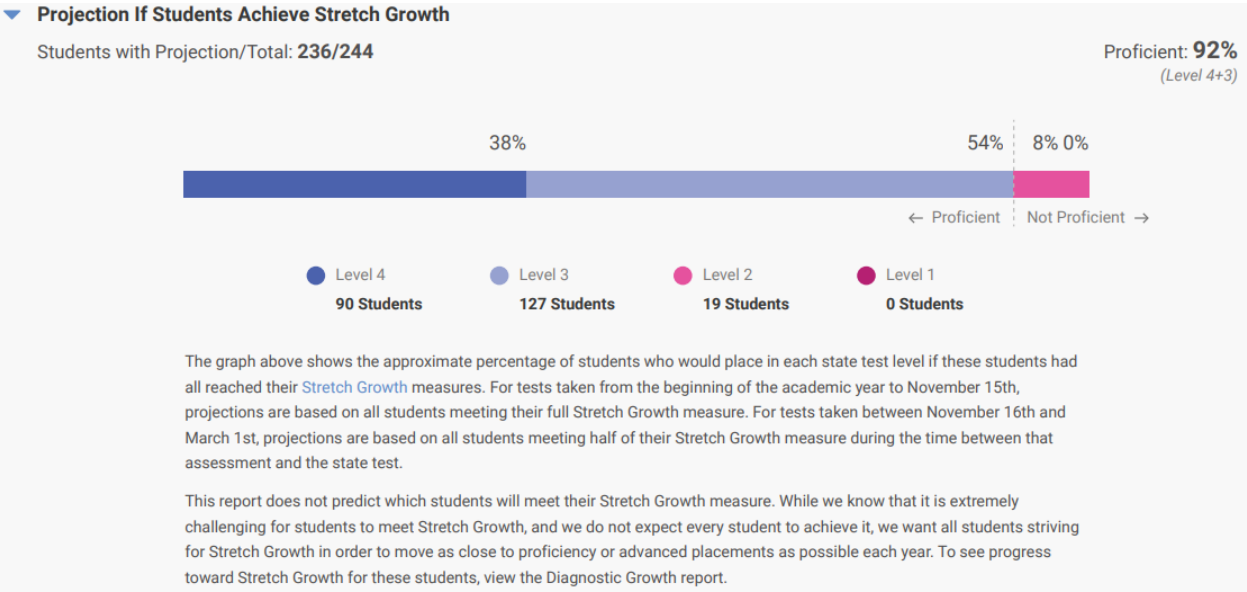


Figure 11.24. Projected Proficiency Report: Achieve Stretch Growth

11.5 Other Reports

11.5.1 Standards Performance Report

The Standards Performance report is a state-specific report that provides a high-level overview of students' likely mastery of the state's content standards based on *i-Ready Inform* performance. For some states, this will be the Common Core State Standards. For others, it will be the state-specific standards based on their alignments to the *i-Ready Inform* indicators. The intended interpretation and use of the report is to provide additional understanding of the strengths and needs of students in a class relative to the local standards and to inform instruction accordingly.

Using aggregate reports, district and school administrators can monitor class, school, and district performance against the applicable standards and to plan accordingly.

Students receive one of three designations based on their performance on *i-Ready Inform* in relation to their state's standards:

- Green Checkmark: Students have likely mastered the *i-Ready* skills aligned to the state standard.
- Clear Checkmark: Students likely have mastered only some—but not all—of the *i-Ready* skills aligned to the standard.
- Grey X: Students likely do not have sufficient mastery of the *i-Ready* skills aligned to the standard.

See Figure 11.25 for an example of a class-level report for Massachusetts, which shows the number of students who received each designation based on their *i-Ready Inform* performance. Note that higher-level reports look similar but are aggregated across more students. By clicking on each standard, the list of students and their designation for the standard appears.

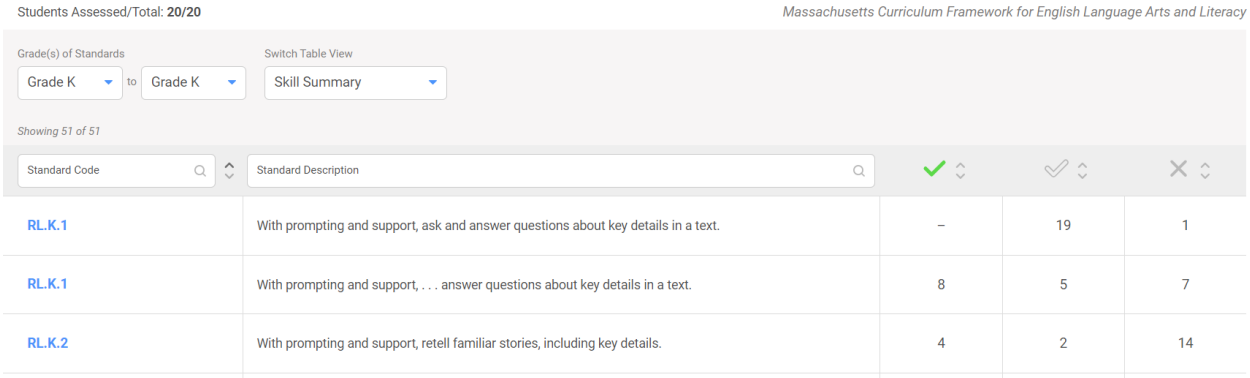


Figure 11.25. Standards Performance Report: Class Level

Clicking on a student's name brings up a table like the one shown in Figure 11.26 in which the standard-by-standard performance for the student can be viewed.

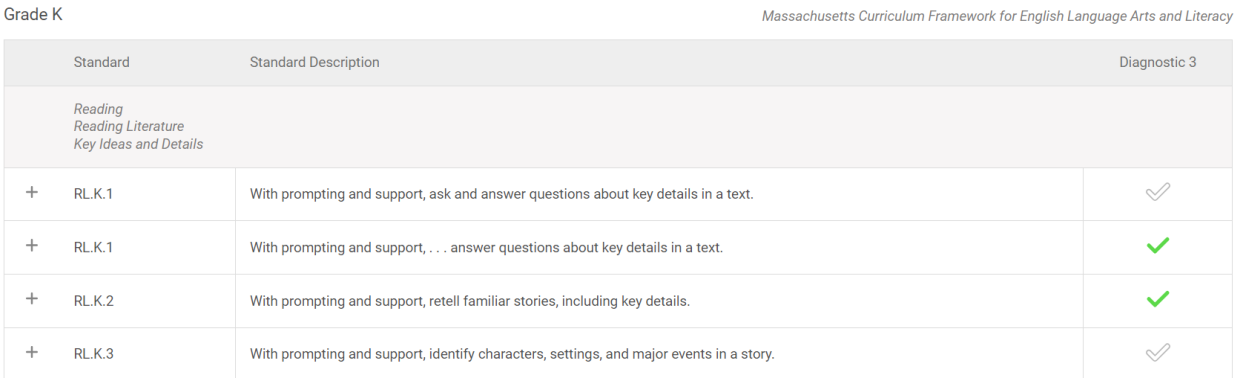


Figure 11.26. Standards Performance Report: Student Level

11.5.2 Instructional Groupings

The Instructional Groupings report is a key feature of the *i-Ready Inform* reporting functionality and the theory of action, as it provides extensive, detailed information to teachers to address students' specific knowledge gaps and guide them in providing impactful, differentiated instruction to students based on their individual needs.

The report groups students with similar instructional needs (Figure 11.27) and, for each group, provides overall and domain-level performance accompanied by a description of the group (Figure 11.28). Feedback on differentiation and instructional priorities is presented for each group (Figure 11.29), followed by a comprehensive list of instructional resources (Figure 11.30 provides a partial view for a group). Administrators can also view schoolwide student groups.

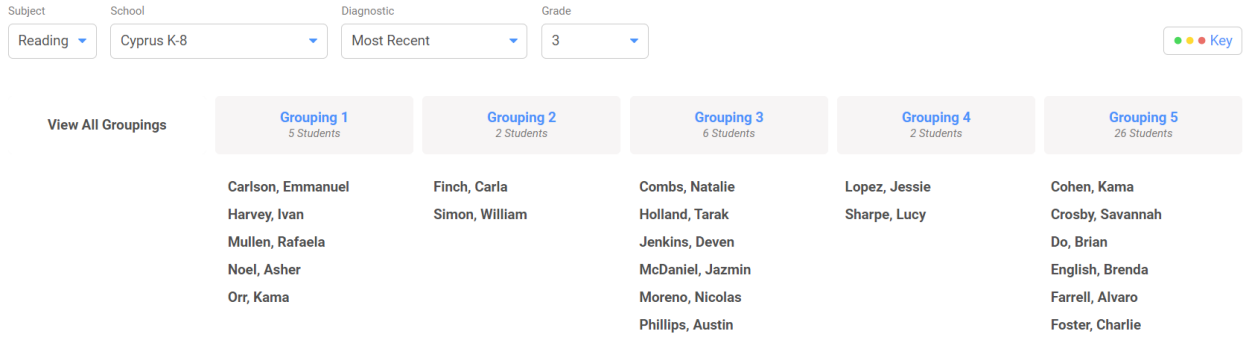


Figure 11.27. Instructional Groupings Report: Class-Level Groups

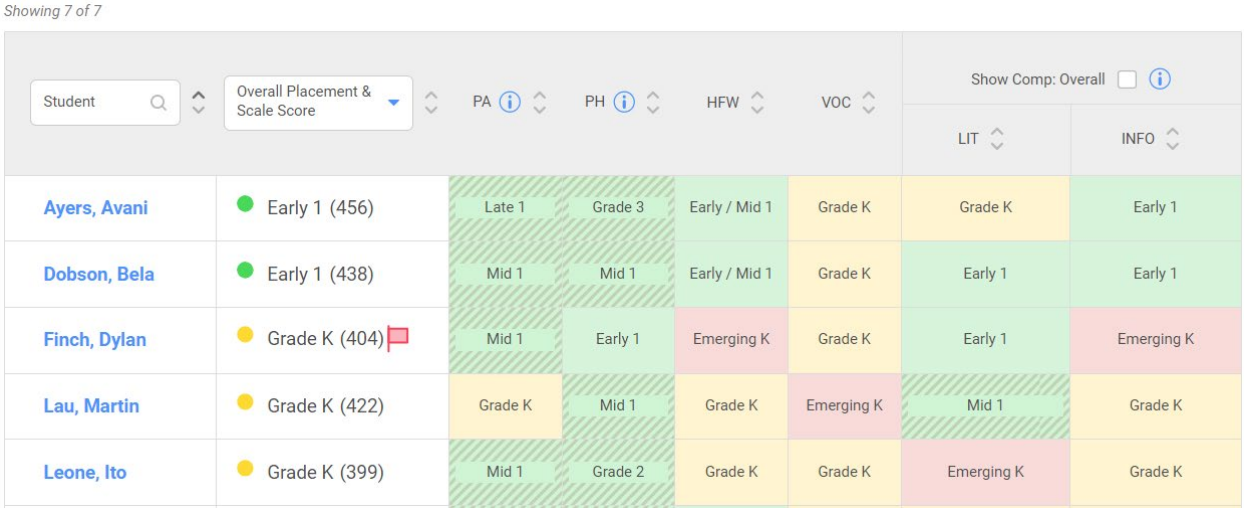


Figure 11.28. Instructional Groupings Report: Sample Group Composition

*** Student Needing Additional Differentiated Instruction**

Results indicate that these students are considerably below level in Vocabulary. They will need more intensive instruction in this area.

For more information about differentiating instruction to meet their needs, see their individual Student Profiles.

Instructional Priorities

Vocabulary

Students in this profile are likely to have difficulty not only with word meanings but also with the background knowledge required by grade-level literary and informational text. Thus, one focus of small-group instruction should be the meanings of individual words, as well as word relationships, word parts, and other word-learning strategies. These students may show some gains in background knowledge as Vocabulary improves. Also integrate instruction of Vocabulary in comprehension activities that focus on drawing meaning from texts.

Comprehension

These students will benefit from more support in Comprehension, as they can already decode accurately. Making Comprehension a focus of small-group instruction will support the reading development of all students in this profile, even those who may be scoring on or above level. Provide explicit strategy instruction that includes modeling, guided practice, and independent application.

Close

Figure 11.29. Instructional Groupings Report: Instructional Differentiation and Priorities

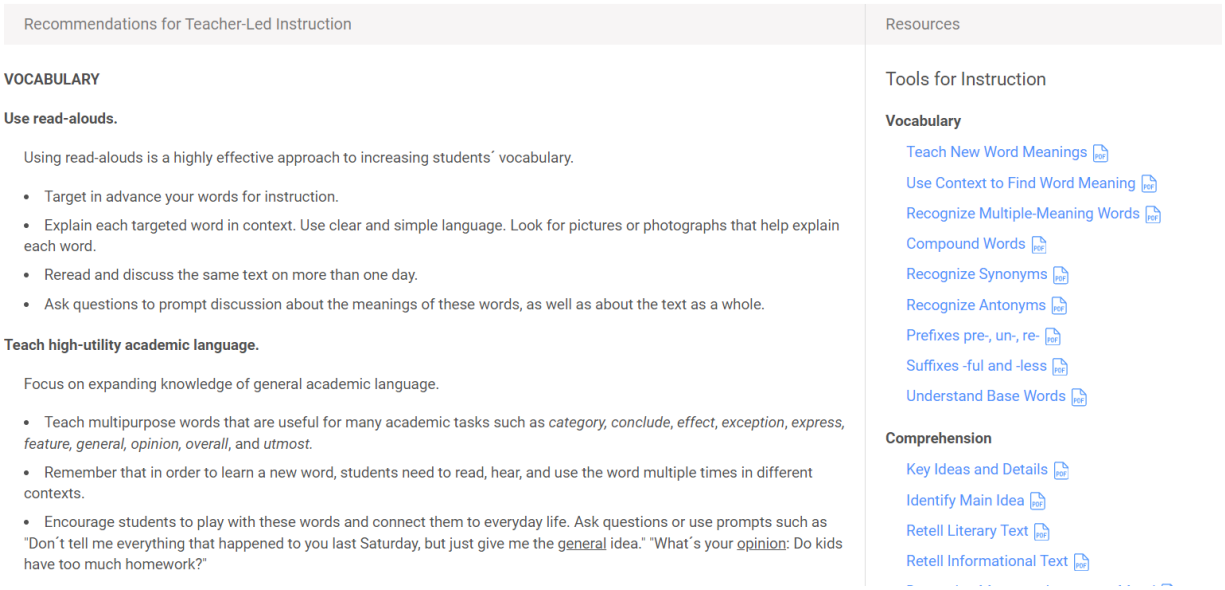


Figure 11.30. Instructional Groupings Report: Partial View of Group-Specific Instructional Resources

11.5.3 Grade-Level Planning Reports

Teachers can use information in the Grade-Level Planning (Scaffolding) report for Reading and Grade-Level Planning (Prerequisites) report for Mathematics to make instructional decisions about how to best support students during an upcoming *i-Ready* lesson. These reports indicate which previous lessons are connected to the current unit and provide guidance to educators about which content is important to review to address

students' unfinished learning from those previous *i-Ready* lessons. The reports also link to unit and lesson support resources.

Teachers select the skill they are preparing to teach in core instruction. The Grade-Level Planning (Scaffolding) or Grade-Level Planning (Prerequisites) report uses data from the most recent *i-Ready Inform* assessment to assign students into groups according to how prepared they are to learn the selected skill. Data are reported for individual students, small groups, and overall common needs in the whole class.

The Grade-Level Planning (Scaffolding) report for Reading, portions of which are shown in Figure 11.31 and Figure 11.32, helps educators understand learning needs for upcoming reading comprehension instruction and identify resources to help students access grade-level texts. The report also recommends Reading Buddies, which are matched, mixed-level recommended reading pairs that will support students in accessing grade-level texts.

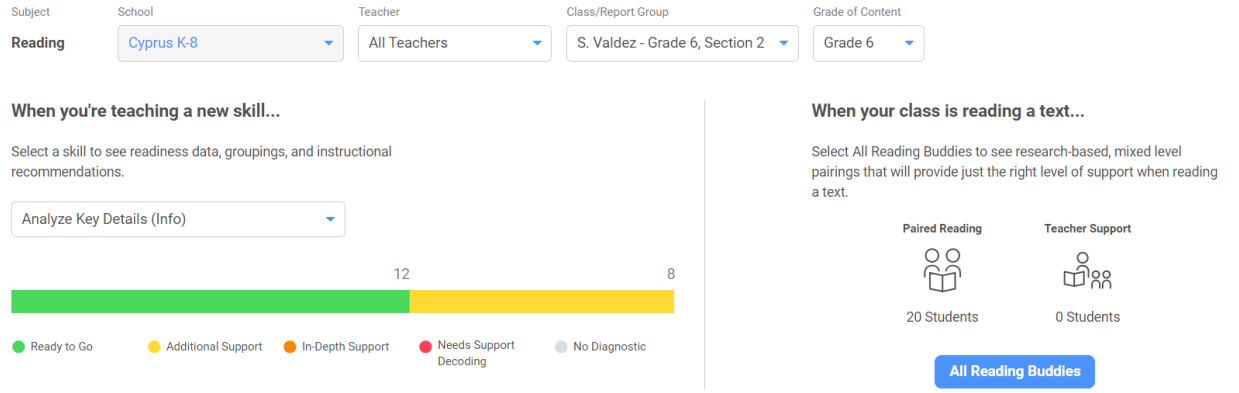


Figure 11.31. Grade-Level Planning (Scaffolding) Report: Class-Level Results

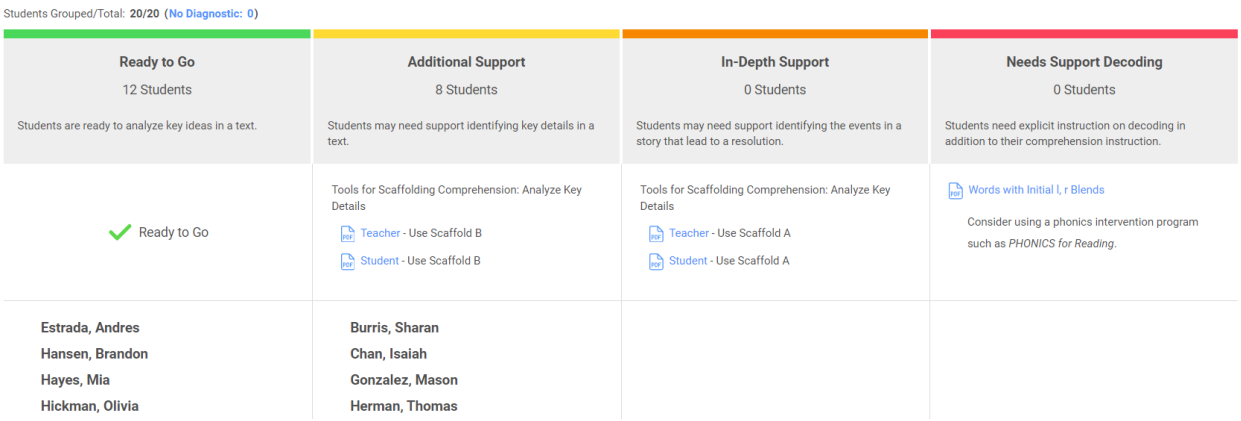


Figure 11.32. Grade-Level Planning (Scaffolding) Report: Student Groups and Support Materials

The Grade-Level Planning (Prerequisites) report for Mathematics, portions of which are shown in Figure 11.33 and Figure 11.34, provides a data-driven method of quickly and precisely addressing concepts and skills most likely to make grade-level instruction successful. This report informs unit/lesson planning by identifying essential prerequisites, potential student groupings, and recommended resources to be used during classroom instruction.

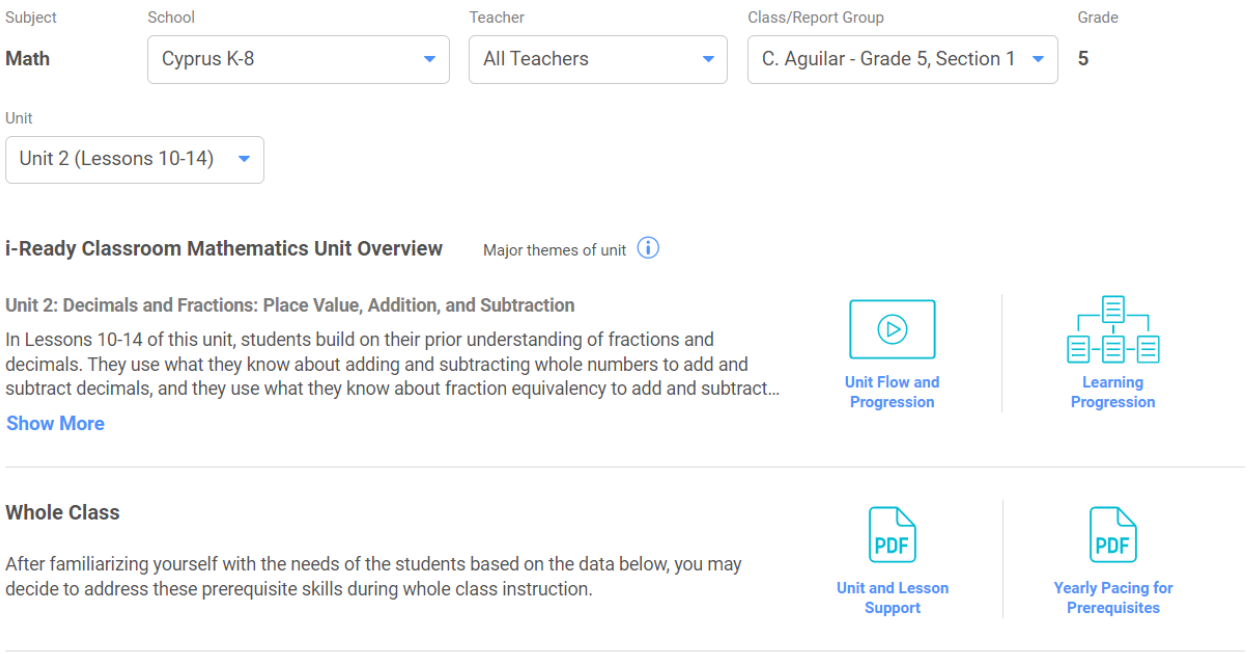


Figure 11.33. Grade-Level Planning (Prerequisites) Report: Unit Overview

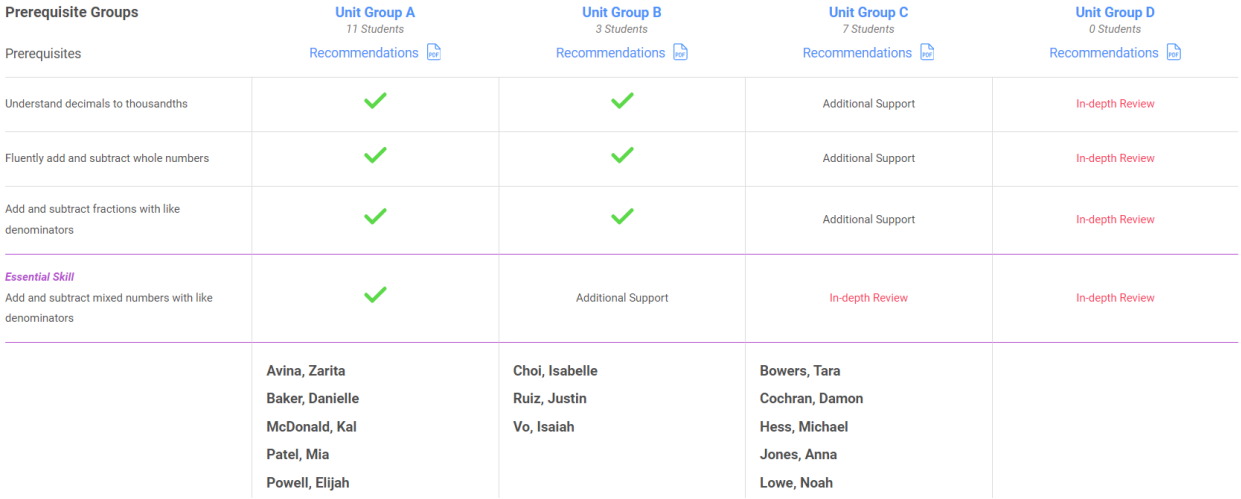


Figure 11.34. Grade-Level Planning (Prerequisites) Report: Partial View of Group-Specific Instructional Resources

11.5.4 Additional Interpretive Guidance

Curriculum Associates equips educators with high-quality, clear, and actionable information about *i-Ready Inform*, how it works, and how to interpret results. In addition to the interpretive guidance and resources included on each report, users and stakeholders have access to an extensive public library of Frequently Asked Questions and *i-Ready Central* resources to support accurate interpretations of *i-Ready Inform* results and using reported information in ways consistent with the intended uses. Also, each *i-Ready* user is assigned an *i-Ready Partners* support team, which is composed of several members, including dedicated professional learning team members and partner success managers, who are the points of connection to a network of experts solely focused on making implementation successful. Depending upon account size, partner success managers may additionally offer districts real-time reviews of achievement data after every assessment window (i.e.,

beginning, middle, and EOY). These reviews include detailed analyses of student performance, growth, and trends in student success to empower data-driven practices in classrooms. The goal is to create an action plan using insights gained from custom district analysis and to gain new perspective with state and local data.

In accordance with the Standards (AERA et al., 2014)—especially Standard 12.15—Curriculum Associates provides professional learning, one-on-one support, and training courses through our Online Educator Learning platform to promote sound use and valid interpretations of the data by stakeholders.

Guidance and support documents are created for the intended audiences with the full range of grade levels and student performance levels in mind. Interpretive and support materials provide guidance in using *i-Ready Inform* results to differentiate instruction to meet the needs of students across the entire performance spectrum.

i-Ready provides extensive instructional resources within the platform, directly accessible from student and class results. These additional resources provide guidance in accessing and using the online instructional support resources.

11.6 Summary

For ease of reference, Table 11.1 lists and briefly describes the standard *i-Ready Inform* reports discussed above, which are available in the online system *i-Ready Central*.

Table 11.1. Summary of Selected Reports

Report	Description
Inform Results (K–8)	• **Student Level:** Presents *i-Ready Inform* results, including scale score, placement level, norm scores, Lexile/Quantile measure, Typical Growth and Stretch Growth measures, with what the student "Can Do" and instructional next steps Reference Sheet: Diagnostic Results (Student): Reading Reference Sheet: Diagnostic Results (Student): Mathematics • **Class, School, and District Level:** Provides student performance at class/school/grade levels, thereby enabling administrators to inform intervention strategies and resources Reference Sheet: Diagnostic Results (Class): Reading Reference Sheet: Diagnostic Results (Class): Mathematics
For Families (K–8)	• **Student Level:** Provides families with their student's overall performance, scale scores, placement levels, Typical Growth and Stretch Growth goals (Grades K–8), and definitions of terms Understanding Your Student's *i-Ready Diagnostic* Results
Standards Performance Results (K–8)	• **Student and Class Level:** Evaluates how the student/class is performing against the standards based on their performance on *i-Ready Inform*. Includes point-of-use descriptions of the aligned standard(s) and related skill. Reference Sheet: Standards Performance (Class)
Inform Growth (K–8)	• **Student Level:** Shows Typical Growth and Stretch Growth measures and progress toward each measure for each student Reference Sheet: Diagnostic Growth (Student) • **Class, School, and District Level:** Shows how students in a class/school/district are progressing toward growth measures along with the median percentage toward Typical Growth for a grade by class, school, or district Reference Sheet: Diagnostic Growth (Class)

Growth Monitoring Results (K–8)	• **Student Level:** Provides information about how much growth a student should be making and whether the student is on track Reference Sheet: Growth Monitoring Results (Student) • **Class, School, and District Level:** Monitors likelihood of a class/school/district meeting their Typical Growth, Stretch Growth, and on-grade level measures Reference Sheet: Growth Monitoring Results (Class)
Projected Proficiency Reports	• **Class Level:** Estimates likely state test level performance given assumptions of growth on the *i-Ready Inform* • **School and District Level:** Provides estimates of approximate percentages of students who would likely place into each state test performance level
Instructional Groupings (K–8)	• **Class and School Level:** Groups students in each class/grade level so those who need support with the same skills can get the most out of small group instruction. It provides instructional priorities and resources. Reference Sheet: Instructional Groupings
Grade-Level (Planning) Scaffolding (3–8) Reading Only	• **Class Level:** Helps teachers prepare students for an upcoming skill they are about to teach with suggested reading buddies, instructional groupings, and standards-based instructional scaffolds Reference Sheet: Grade-Level Planning (Scaffolding)
Grade-Level Planning (Prerequisites) (K–8) Mathematics Only	• **Class Level:** Informs unit/lesson planning by identifying essential prerequisite skills, potential student groupings, and recommended instructional resources in relation to upcoming grade-level content Reference Sheet: Grade-Level Planning (Prerequisites)

Chapter 12: Validity

12.1 Chapter Summary

In this final chapter, we review the relevant evidence provided in the preceding chapters of this technical manual as well as additional information not contained elsewhere in this document to support an overall validity argument rooted in the claims outlined in Chapter 1 and guided by the Standards for Educational and Psychological Testing (AERA et al., 2014). We first show how the evidence contained in the preceding chapters supports the claims we make about *i-Ready Inform*. In the remainder of the chapter, we summarize this evidence and organize it into the five sources of validity evidence described by the Standards.

12.2 Introduction

In this chapter, we discuss the evidence presented in this technical manual and how it supports the validity argument for *i-Ready Inform*. As discussed in Chapter 1, *validation* is a process of developing and substantiating a validity argument (Kane, 2006, 2013). The process is focused primarily on the interpretations and uses of the scores. Thus, *validation* entails gathering ample evidence that supports the intended interpretation of scores, and *validity* is the degree to which evidence and theory support the interpretations of test scores for intended uses (AERA et al., 2014). The chapter provides a narrative analysis of the evidence presented in earlier chapters of this manual as well as additional evidence from other sources.

12.2.1 Five Sources of Evidence

The Standards (AERA et al., 2014) outline five categories under which validity evidence can be organized to develop a strong validity argument. Using this framework as a guide, we organize the chapter around these types of evidence, which we will now briefly describe.

Evidence based on test content. This type of evidence demonstrates that the assessment is indeed measuring the breadth of the construct of interest through the content represented in the test specifications. Item writing protocols, content organization schemata, alignment studies, and administration standardization procedures are examples of supporting evidence.

Evidence based on response process. This type of evidence supports the inferential link between the assessment activity (i.e., the item) and the construct of interest. The degree of inferencing can vary depending on the nature of the activity and the construct; in cases where only the correctness of the response is of interest, there is no inferential link (AERA et al., 2014). Cognitive laboratory studies and response time analyses are common types of evidence in this category.

Evidence based on internal structure. This type of evidence demonstrates that the assessment's statistical qualities are consistent with the nature of the construct being measured and are of sufficient quality to support the interpretations and uses of the scores. Of particular interest in this category are reliability coefficients, classification performance, and DIF analyses.

Evidence based on relationships with other variables. The evidence in this category can be divided into two categories: convergent/divergent relationships with other measures and test–criterion relationships (AERA et

al., 2014). Convergent and divergent relationships show that the instrument correlates strongly with other instruments measuring similar constructs, and that the instrument does not correlate strongly with instruments measuring dissimilar constructs, respectively. Test–criterion relationships show a relationship between the instrument and an external criterion. Linking studies and standard-setting procedures are examples of evidence in this category.

Evidence based on consequences of testing. This type of evidence shows that the consequences of testing—both intended and unintended—are aligned with the claims and are justifiable. Because *i-Ready Inform* is a component in the theory of action (see Chapter 1) whose outcomes drive reporting and instructional resources, evidence in this category may depend on the fidelity of the implementation of the other components of the theory of action.

12.2.2 Locating the Evidence

To organize and evaluate the validity evidence within this manual, we start by exploring the intersection of the claims with the chapters in this manual. In each cell of Table 12.1, we report the source of evidence presented by chapter and claim. In the remaining sections of this chapter, we provide a brief narrative of this evidence organized according to its type.

Table 12.1. Sources of Validity Evidence by Chapter and Claim

Claims of the *i-Ready Inform* Assessment: Interpretation of Scores

Chapter	Title	Measures Domains in Mathematics and Reading	Reports a Variety of Estimates of Accurate Student Proficiency	Contextualizes Student Proficiency	Provides Growth Measures and Projected Proficiency
1	Introduction	T			
2	Test Design and Development	T R		T	
3	Test Administration	T	T		
4	CAT and Test Flows	T R	T		
5	Calibration and Scaling	I	I	I	
6	Reliability				
7	Setting Standards	T C	T C	T C	T C
8	Score Reports		C	C	C
9	Norms			C	
10	Growth Model				C
11	Projected Proficiency			E C	E C

Note: T = Test Content; I = Internal Structure; R = Response Process; E = Relationships with External Variables; C = Consequences

Claims of the *i-Ready Inform* Assessment: Use of Scores

Chapter	Title	Identify and Monitor Students	Provide Feedback on What Students Can Likely Do and Should Work on Next	Group Students to Provide Differentiated Instruction	Route Students into the Appropriate Personalized Instruction Module
1	Introduction				
2	Test Design and Development				
3	Test Administration				
4	CAT and Test Flows				
5	Calibration and Scaling			I	I
6	Reliability			I	I
7	Setting Standards	T C		T C	TC
8	Score Reports	C	C	C	C
9	Norms	C			
10	Growth Model	C			
11	Projected Proficiency				

Note: T = Test Content; I = Internal Structure; R = Response Process; E = Relationships with External Variables; C = Consequences

12.3 Evidence Based on Test Content

Evidence related to test content concerns the relationship between the assessment's content and the construct it is intended to measure. Test content refers to the subject matter, wording, and the format of items as well as their administration and scoring. Test specifications describe the test content in detail and classify the content areas and the types of items. This evidence, which is a combination of procedural documentation and empirical alignment studies, supports the intended interpretation of the scores.

12.3.1 Assessment Design and Development

Chapters 1 and 2 describe the foundations of *i-Ready Inform* and our principled approach to assessment design and development, including:

- A description of the target population
- How the test constructs (i.e., reading and mathematics) are operationalized through construct mapping, indicators, and PLDs
- How items are specified, written, reviewed, and aligned to content and skills
- How items are field tested (also discussed in Chapter 5) to gather empirical evidence of psychometric item characteristics prior to inclusion in operational item pools
- What types of items are administered

- How content is sampled (i.e., test flows, which are also discussed in Chapter 4) and the cognitive processes elicited by items
- How we ensure that students have full access to the assessment content, including a discussion of how accessibility supports and accommodations are incorporated into the assessment design as well as the availability of a Spanish-language version of the *i-Ready Inform* for Mathematics assessment

The outcomes of the content development processes provide evidence supporting the validity argument for the *i-Ready Inform* assessment, thereby ensuring that the links between the construct and the content on the assessment are well articulated and supported by a clear rationale and accompanying evidence. In sum, the evidence provided in Chapter 2 establishes the content-based rationale for ensuring that *i-Ready Inform* assesses student achievement in reading and mathematics to support the intended uses and interpretations of scores.

12.3.2 Test Administration and Accommodations

Although neither a characteristic of the design nor the content, the administration conditions of an assessment are of paramount importance to ensure score comparability across students over time. Student responses are the basis of the score reports as well as the psychometric components of *i-Ready Inform* (i.e., scaling, equating, maintenance of the score scales, norms, growth measures, and projected proficiency scores), which are well documented in various chapters.

Chapter 3 provides evidence describing the administration conditions and supports available to students. Clearly defined monitoring strategies and test security protocols promote the consistent application of administration practices designed to elicit item responses that best reflect student knowledge and ensure that the scores obtained are an accurate reflection of student proficiency. The chapter describes steps that test proctors and administrators must take to prepare for, conduct, and complete an administration of *i-Ready Inform*. Testing windows for each administration (i.e., fall, winter, and spring) and recommended testing times are also provided. In addition, the chapter describes how the *i-Ready Inform*'s accessibility features and accommodations provide all students—including students with disabilities and Multilingual Learners—access to *i-Ready Inform* by minimizing construct-irrelevant barriers to student performance.

12.3.3 The Adaptive Algorithm and Test Flows

Chapter 4 details how content specifications are embedded into the item selection algorithm. First, *i-Ready Inform*'s algorithm leverages adaptive assessment methodology to optimize score precision by selecting items that are closely matched to students' proficiency levels. In doing so, the algorithm enhances reliability—and reduces SEM—compared to a linear (i.e., non-adaptive) test of the same length (assuming all other factors are held constant), and thus supports the appropriate interpretations and the effective uses of the scores.

Additionally, the test flows are incorporated into the algorithm such that appropriate content is displayed to students based on grade level and performance on earlier test items. Furthermore, the chapter summarizes simulation research that Curriculum Associates has conducted to refine processes and improve the effectiveness of *i-Ready Inform* using a simulator that evaluates the efficiency of decisions regarding item selection, proficiency estimation, and item development priorities to support bank maintenance.

12.3.4 Criterion-Referenced Placement Levels

Placement levels are the foundation for criterion-referenced score interpretations. Chapter 2 describes the placement levels and Chapter 7 describes the standard-setting methodology used to determine the cut scores between the placement levels. PLDs provide a detailed framework according to which content is mapped to items. Instructional groupings and Personalized Instruction—two of the key intended uses of scores—are both based on placement levels and are used to deliver appropriate instructional content to students.

12.3.5 Domain Profile Analysis

We conducted research to evaluate the relationship between overall Reading or Mathematics placement level on *i-Ready Inform* and placement level for each domain within the content area. The study examined whether placement levels for the overall *i-Ready Inform* assessment were more or less likely to match the placement level for each domain as the year went on. The study also evaluated yearly trends in domain placement profiles across testing windows during the 2017–2018 school year. Results indicated that, while dominant profiles could be identified in some cases, there was large variability in student placement profiles, including for students with the same overall placement. This provides evidence that the domain scores and placement levels represent distinct information about student skills beyond that provided by the overall score and placement level.

Patterns of profiles suggest that domain skills often develop in predictable sequences. For example, vocabulary skills develop prior to comprehension skills, and literary comprehension develops ahead of informational comprehension. In mathematics, students' proficiency in number and operations and measurement skills develop ahead of geometry skills. Given the variability of domain profiles and the mostly consistent yearly trends in profiles, the study supports the intended use of *i-Ready Inform* to route students to individualized instruction, especially for students who need to make more than Typical Growth during the academic year to move from below-grade level to on-grade level proficiency.

12.3.6 Alignment Studies

In addition to evidence presented in earlier chapters, as summarized in the preceding sections in the current chapter, Curriculum Associates routinely requests independent alignment studies to examine the relationship between the assessment and the content standards as operationalized through test specifications and item writing.

12.3.6.1 Alignment of Items to Content Standards

In 2017, Curriculum Associates collaborated with Odell Education and WestEd® for alignment studies. Odell Education's analysis of a representative sample of items from *i-Ready Inform* in Grades K–8 for Mathematics and for Reading confirmed strong alignment in categorical concurrence, DOK, and range of knowledge (Webb, 1997, 1999, 2007).

WestEd® is a registered trademark of WestEd.

Committed to enhanced alignments, Curriculum Associates engaged WestEd to evaluate the quality and alignment of test items in Grades 3 and 7 in Reading and in Mathematics with CCRS. From a sample of 800 items, findings indicated that 74 percent of items reviewed met criteria for structural quality, 94 percent of

items reviewed met criteria for content quality, and 92 percent of items reviewed met criteria for fairness. Furthermore, they reported that 90 percent of items strongly or partially met criteria for categorical concurrence, 98 percent of items aligned for cognitive complexity, and 93 percent of items aligned for processing complexity.

12.3.6.2 Examples of State-Specific Alignment Studies

In 2018, Odell Education completed two alignment studies that examined the *i-Ready Inform* Grades K–8 alignment to the South Carolina College- and Career-Ready Standards for Mathematics and for ELA from 2015. Expert raters reviewed representative, stratified samples from the *i-Ready Inform* item banks for Mathematics and Reading. The studies confirmed strong alignment for all Grades K–8 for the three criteria examined: categorical concurrence, DOK, and range of knowledge (Webb, 1997, 1999, 2007).

Also in 2018, Odell Education conducted a study on the alignment between the *i-Ready Inform* for Reading assessments and the Oklahoma Academic Standards for ELA. The study used an adaptation of the Webb (1997, 1999, 2007) process with content specialists analyzing the Oklahoma Academic Standards for ELA and aligning a stratified sample of the *i-Ready Inform* item bank for Grades K–3 ELA to the content standards. In most cases, results indicated that the Grades K–3 *i-Ready Inform* items align to the content standards with respect to the three Webb criteria.

12.3.7 Comparability Study between the *i-Ready Inform* for Mathematics in English and the *i-Ready Inform* for Mathematics in Spanish

In order to use and interpret the scores on the *i-Ready Inform* for Mathematics in Spanish relative to the *i-Ready Inform* for Mathematics in English, the assessments should be comparable and measure the same underlying construct. In fall 2019, Curriculum Associates conducted a comparability study to evaluate the performance of students who took the *i-Ready Inform* for Mathematics assessment using either the English- or Spanish-language version. The study included data from more than 1,300 students who took the test in Spanish, who were statistically matched using prior Mathematics and Reading scores to students who took the test in English. There were no statistically significant differences in the mean performance of students between the English- and Spanish-language versions of the test when using the matched scores.

These results provide evidence that scores from the English- and Spanish-language versions of the *i-Ready Inform* for Mathematics are comparable across languages. Note that this study only examined point-in-time scores for the matched sample student distribution and did not evaluate comparability of the *i-Ready* growth model across languages. This research is ongoing.

12.4 Evidence Based on Response Process

Evidence related to response processes focuses on the cognitive processes in which students engage while taking the assessment. Evidence can include theoretical and empirical analyses of the response processes to provide insight into the fit between the construct and the cognition elicited by the test items in cases where a substantial inferential link is required.

12.4.1 Measures of Cognitive Complexity and Difficulty

The cognitive complexity, or cognitive demand, of an item is an area of specification used when developing *i-Ready Inform* items. Each content area uses an appropriate definition of cognitive complexity and a specific approach to guide item development, as described in Chapter 2. For reading, this includes Webb's (1997, 1999) DOK. For mathematics, this includes combining Achieve's (2019) definition of AOR and level of complexity in a matrix approach that leads to a single classification that reflects the cognitive demand of the item. The chapter elucidates how these measures are defined and how methods are employed to drive item development in line with intended complexity.

Chapter 4 further elaborates on the application of cognitive complexity, particularly in the selection of appropriate *i-Ready Inform* items for each student. Test flows specify how the item selection algorithm ensures that all students are exposed to the appropriate spectrum of cognitive demand. The *i-Ready Inform* CAT simulator evaluates changes to the item bank, item selection methods, or test flows to maintain optimal item selection.

12.4.2 Cognitive Laboratory Studies

In addition to what has been presented in this technical report, evidence regarding response process also comes from cognitive laboratory studies (i.e., "cog labs"). A series of cognitive interviews were conducted to gain an enriched understanding of students' perceptions of specific *i-Ready Inform* items and the cognitive processes students use when responding to items. Interviews were conducted at two different sites in Florida with elementary and middle school students.

During the 2011 item pilot, researchers used think-aloud protocols and cognitive interview methods to observe students as they took the test and document their observations. Thirteen elementary and middle school students in Grades K–8 were interviewed in a computer lab setting, which involved responding to computer-based test questions and periodically reflecting on their thought processes with the interviewer. Common queries targeted their understanding of and approach to the presented problems.

Overall, the findings suggested that most items invoked the intended cognitive processes. Multimedia enhancements did not add cognitive strain, and all participants comfortably navigated the digital environment.

12.4.3 Monitoring Response Process Times

To ensure that the test's constructs are being appropriately measured, Curriculum Associates tracks individual item response times and patterns. Varied response times for different item types can provide valuable insights into student engagement. Students who are not likely using the intended response processes to respond to items—as evidenced by a pattern of response times that is too short—may be flagged for "rushing" behavior, which can compromise the integrity of the reported score.

12.5 Evidence Based on Internal Structure

The internal structure of an assessment refers to the interrelationships among the items and components of the test. A consistent internal structure supports the notion that test items measure the same construct or related constructs.

12.5.1 Measurement Characteristics

Chapter 5 presents the psychometric theory and supporting research underlying the *i-Ready Inform* score scales and proficiency estimation. Evidence in Chapter 5 includes the application of classical test theory to monitor the psychometric properties of items and flag items when their properties are unexpected or unfavorable, and the application of IRT for calibration, equating, and vertical scaling. Embedded field testing allows for the refreshing of the *i-Ready Inform* item bank and maintaining a single item bank across administrations within each subject area. Chapter 5 also includes calibration studies to evaluate and recalibrate the vertical scale.

The evidence presented in Chapter 5 demonstrates that we apply best-practice approaches to maintain the *i-Ready Inform* score scale, which provides a meaningful metric rooted in criterion-referenced placement levels (Chapter 7), in conjunction with the PLDs and norms (Chapters 2 and 8), to interpret student knowledge in all content domains and allow for the uses and interpretations of scores consistent with the overall validity argument.

12.5.2 Reliability

Marginal reliability estimates offer insights into the precision of scores and are computed for the overall assessment as well as the corresponding domains. Marginal reliability estimates are also computed for some student groups. Overall, reliability analyses indicate the relative cohesiveness and integrity of the test's internal structure, which, in conjunction with evidence of content alignment, supports the notion that the test items consistently measure the intended construct.

Chapter 6 shows that the overall marginal reliability estimates are above .90 for the full population and most student groups. In addition, domain marginal reliability estimates are generally above .80 in Mathematics and .70 in Reading. Corresponding SEMs are presented as well. Test–retest reliability estimates, which in this case are characterized by a delay between administrations, exceeded .80 across all grades except Grade K, where it exceeded .70, demonstrate the stability of scores across administrations.

12.5.3 Classification Accuracy and Consistency

Classification accuracy and consistency analyses support the validity of placement levels by demonstrating sufficient reliability of classification. Because many of the intended uses of scores are based on the placement levels, evidence that students are accurately and consistently classified provides validity evidence for the use cases that depend on placement levels. The results in Chapter 6 show that strong classification performance is observed for both each individual cut score as well as all five placement levels together.

12.5.4 Differential Item Functioning

Under the Rasch IRT model, the probability of a correct response to an item is only dependent on the item difficulty and the person's ability level. DIF analyses are used to understand if an item favors one group of students over another after adjusting for any proficiency differences between the groups. Items that display DIF, which is purely a statistical procedure, can be reviewed by SMEs to determine the reason for the differential performance, which may result in determining that the item is biased. Note that the presence of DIF alone does not necessarily indicate that an item is biased.

Curriculum Associates periodically runs DIF analyses to ensure that items are operating properly and to identify items that need to go through additional review by SMEs and key stakeholders to determine the appropriate course of action. Items with moderate to large DIF are subjected to this extensive review to identify the potential causes. A determination is then made as to whether each item should remain in the item bank, be removed from the item bank, or be revised and resubmitted for field testing. DIF analyses and subsequent item reviews are important quality assurance procedures to support the validity of the scores, which ultimately depend on the items in the bank.

12.5.4.1 Data Source

One DIF analysis included a random sample (10%) of students from the 2015–2016 *i-Ready* operational data. The following demographic categories were compared (focal group versus reference group): Female versus Male; African American and Hispanic versus Caucasian; South, West, and Northeast versus Midwest; English Learner versus non–English Learner; Special Education versus General Education; and Economically Disadvantaged versus Not Economically Disadvantaged. In each pairwise comparison, estimates of item difficulty for each category in the comparison were calculated. Table 12.2 below presents the total number and percentage of students included in the DIF analysis [13].

Table 12.2. Sample Size and Percentage of Students Included in the DIF Analysis

Category	Variable	Reading *N**	Reading Percentage	Mathematics *N**	Mathematics Percentage
Gender	Male	258.4	52.0	267.2	52.0
Gender	Female	238.8	48.0	247.0	48.0
Ethnicity	Caucasian	129.2	36.6	126.4	34.1
Ethnicity	African American and Hispanic	224.2	63.4	244.1	65.9
Region	Midwest	43.2	8.6	43.5	8.4
Region	South, West, and Northeast	457.5	91.4	474.1	91.6
English Learner Status	Non-English Learner	250.8	81.2	262.7	80.8

[13]It should be noted that not all students have individual demographic information. Among the six DIF groups, region is the only category that's not dependent on the student-level information, and the sum of the *N* counts in the region comparison is the total number of students. But the total of the two exclusive groups in the other categories does not necessarily add up to the total of region.

English Learner Status	English Learner	58.2	18.8	62.4	19.2
Special Education	Not Special Education	165.8	85.7	181.0	85.1
Special Education	Special Education	27.6	14.3	31.6	14.9
Socioeconomic Status	Not Economically Disadvantaged	177.8	69.0	192.1	67.1
Socioeconomic Status	Economically Disadvantaged	80.0	31.1	94.1	32.9

Note: * = Sample Size in Thousands

Active items in the bank as of the 2016–2017 school year were included in the DIF analysis. The total numbers of items were 3,649 for Reading and 3,103 for Mathematics, representing the majority of the total items in the bank at the time of analysis. In order to ensure that the DIF comparison is robust and had adequate power, the joint degrees of freedom of both focal and reference groups based on the Welch–Satterthwaite approximation of the normal distribution is also considered (Satterthwaite, 1946).

12.5.4.2 DIF Analysis

WINSTEPS (Linacre, 2012) was used to conduct the calibration for the DIF analyses. We compared the item difficulties for two demographic categories in a pairwise comparison through a combined calibration analysis. The essence of this methodology is to investigate the interaction of the person groups with each item, while fixing all other item and person measures to those from the combined calibration. The method used to detect DIF is based on the Mantel–Haenszel procedure (MH), and the work of Linacre and Wright (1989) and Linacre (2012). To help interpret the results, the Educational Testing Service (ETS) criteria using the delta method was used to categorize DIF (Zwick et al., 1999). Table 12.3 below provides the general guidelines for the ETS DIF categories using WINSTEPS.

Table 12.3. ETS DIF Category with Rasch Model [14]

ETS DIF Category	Definition
A (Negligible)	$\|DIF\| < 0.43$
B (Moderate)	$0.43 \leq \|DIF\| < 0.64$
C (Large)	$\|DIF\| \geq 0.64$

Note: B- and C- suggests DIF against focal group;
B+ or C+ suggests DIF against reference group

The number and percentage of items exhibiting DIF for each of the demographic categories are reported in Table 12.4 and Table 12.5 for Reading and Mathematics, respectively. The majority of both Reading and Mathematics items studied show negligible DIF (from 90.7 to 99.5 percent). About one to four percent of the items show moderate DIF (Category B), and very few items (less than two percent) show large DIF (Category C).

Table 12.4. DIF Results: Reading

ETS DIF Category	Gend *N*	Gend %	Eth *N*	Eth %	Reg *N*	Reg %	EL N	EL %	SE N	SE %	SS N	SS%

[14]Table adapted from http://www.winsteps.com/winman/table30_1.htm.

A	3,400	96.7	3,417	97.7	3,175	98.8	2,922	95.8	2,914	96.7	3,371	99.5
B+	47	1.3	32	0.9	18	0.6	55	1.8	45	1.5	13	0.4
B-	41	1.2	38	1.1	17	0.5	52	1.7	46	1.5	2	0.1
C+	23	0.7	8	0.2	0	0.0	11	0.4	3	0.1	1	0.0
C-	4	0.1	3	0.1	3	0.1	10	0.3	6	0.2	2	0.1
Total	3,515	100	3,498	100	3,213	100	3,050	100	3,014	100	3,389	100

Note: Gend = Gender, Eth = Ethnicity, Reg = Region, EL = English Learner, SE = Special Education, SS = Socioeconomic Status

Table 12.5. DIF Results: Mathematics

ETS DIF Category	Gend *N*	Gend %	Eth *N*	Eth %	Reg *N*	Reg %	EL N	EL %	SE N	SE %	SS N	SS%
A	2,975	96.5	2,756	90.7	2,640	98.9	2,460	95.0	2,253	94.3	2,827	98.5
B+	62	2.0	100	3.3	13	0.5	53	2.0	29	1.2	20	0.7
B-	26	0.8	130	4.3	15	0.6	64	2.5	86	3.6	16	0.6
C+	14	0.5	24	0.8	1	0.0	6	0.2	2	0.1	3	0.1
C-	7	0.2	30	1.0	1	0.0	7	0.3	19	0.8	4	0.1
Total	3,084	100	3,040	100	2,670	100	2,590	100	2,389	100	2,870	100

12.5.5 Demographic analysis

An analysis was conducted to investigate the associations of ethnicity and English Learner (EL) status with the attainment of Typical Growth measures in grades K–8. To conduct this investigation, a longitudinal dataset was constructed in which students had both fall and spring scores. Two outcome measures were used: 1) the percentage of the Typical Growth measure achieved by each student and 2) whether each student achieved their growth measure.

Separate regression models were estimated for each grade level and subject (i.e., mathematics and reading), outcome measure, and predictor (i.e., English Learner status and ethnicity); the initial placement level was used as the control variable. After accounting for the initial placement level, the additional variance explained by the demographic variables was then computed. Across all models, English Learner status and ethnicity accounted for less than one percent and less than two percent, respectively, of the variance in the outcome measures.

12.6 Evidence Based on Relationships to Other Variables

This section describes validity evidence grounded in the empirical relationships between *i-Ready Inform* scores and those arising from other instruments that purport to measure a similar construct, such as academic outcomes or scores from a different test. Such a relationship would demonstrate *convergent* validity evidence. For example, the *i-Ready Inform* for Reading scores should be positively correlated with scores from a state summative ELA test as well as grades in ELA classes.

12.6.1 State Summative Assessment Linking Studies

Chapter 10 summarizes the extensive work that Curriculum Associates has done to explore how *i-Ready Inform*'s Projected Proficiency reporting leverages data from linking studies conducted between *i-Ready Inform* and various state summative tests. These studies, which relied on an equipercentile linking procedure, resulted in concordance tables relating the scales of *i-Ready Inform* and the state test, as well as tables that link spring *i-Ready Inform* scale score ranges with state performance levels.

Data from a sample of states show that *i-Ready Inform* is highly correlated with state summative tests from across the country, with correlations generally exceeding .70. Additionally, linking studies result in strong classification accuracy rates between *i-Ready Inform* and the state summative tests. The chapter describes each measure, how it was developed, and provides research evidence of the classification accuracy at both the student and aggregate levels.

12.6.2 Additional Linking Studies

Additional evidence based on relations to other variables comes from the linking studies that Curriculum Associates has conducted in collaboration with MetaMetrics. These studies link *i-Ready* performance to the Lexile and Quantile Frameworks across Grades K–12 and also provide convergent validity evidence for measures administered close in time. In Grades K–8, students took a standalone Lexile or Quantile form from MetaMetrics in addition to completing *i-Ready Inform*. Correlations between *i-Ready* for Reading scores and Lexile measures ranged from .88 to .89, while those between *i-Ready* for Mathematics scores and Quantile measures varied from .59 to .79. For high school grades, rather than administering separate tests, linking items were embedded within the *i-Ready Inform* assessment to link the scales. These results reinforce the test's strong external validity and help to establish a relationship between the *i-Ready* scale and the Lexile and Quantile scales.

12.6.3 Correlation Studies

Annual correlation studies aim to investigate the relationships between *i-Ready Inform* scores and scores on state summative assessments in Reading and Mathematics. Two primary research questions guide these studies:

1. Are *i-Ready Inform* scores correlated with the scores on the ELA/Literacy and Mathematics assessments administered to students in the given year?
2. Can *i-Ready Inform* scores be used to accurately predict students' proficiency on the ELA/Literacy and Mathematics assessments?

These studies examine correlations for individual grades and average correlations across grade levels. Generally, the correlations between *i-Ready Inform* scores and state assessments range from about .75 to .85 in both ELA and Mathematics for Grades 3–8. In summary, the strong and consistent correlations between *i-Ready Inform* and other key measures provide evidence in support of convergent and predictive forms of validity.

12.7 Evidence Based on Test Consequences

Consequential validity evidence focuses on an evaluation of the consequences of the uses of test scores. This manual provides information about test content and technical quality with extensive research to support the intended uses and interpretations of scores, much of which is provided through our score reports, which are detailed in Chapter 11. In support of intended uses of test scores, which are outlined in Chapter 1, Chapter 7 details the standard-setting processes, which have implications for student placement and the corresponding resources supplied to teachers. Additionally, the norm development covered in Chapter 8 and the growth models discussed in Chapter 9 inform the interpretation of results, and hence, the decisions surrounding teaching strategies and resource allocation.

12.7.1 Growth Measures

Chapter 9 provides evidence of how *i-Ready Inform* scores can be used to understand and track students' growth over time. The chapter describes the *i-Ready Inform* Typical Growth and Stretch Growth measures. Typical Growth is based on the median growth of students within a starting placement level and provides a benchmark for evaluating student improvement compared to typical students in the same placement level. Stretch Growth complements the Typical Growth measures by presenting a trajectory for students performing below grade level to make progress toward proficiency or beyond, depending on the starting placement, over a designated period.

The chapter describes the development of each measure, reports numerical growth measures, and provides guidance for interpreting and using each measure. Growth score reliability coefficients are presented, and two validity studies are summarized. The first study shows that *i-Ready Personalized Instruction* lesson pass rates are positively associated with growth on *i-Ready Inform*, and the second shows that students who attain their Stretch Growth measures are more likely to be identified as proficient in the future. The vertical scale is instrumental in measuring within- and across-year student growth. Thus, the evidence supports the use of these growth measures as a tool to help teachers understand if students are making adequate progress.

12.7.2 Efficacy Studies

While the focus of this manual has generally been about the technical qualities and composition of the *i-Ready Inform* and Growth Monitoring assessments themselves, we reiterate that *i-Ready Inform* is situated within our comprehensive theory of action, which includes specific intervention strategies to improve student learning. Thus, it is worthwhile to discuss a few interesting findings in that stream of work.

Curriculum Associates conducts efficacy research on an ongoing basis to evaluate the long-term impact and outcomes of the *i-Ready* program, which is the final component of our theory of action, shown in Figure 1.1. This type of research serves as validity evidence for the two left-most boxes on the bottom row of Figure 1.2.

Some of our efficacy research has evaluated the impact of the *i-Ready Personalized Instruction* program, which is a frequently used intervention tailored to students based on their placements from *i-Ready Inform*. To measure the efficacy of the programs and provide validity evidence, our research has generally sought to examine whether the usage of *i-Ready Personalized Instruction* is associated with higher performance on an academic outcome measure (usually either the spring *i-Ready Inform* assessment or a state summative test).

As an example of some of this work, Curriculum Associates collaborated with Century Analytics, WestEd, and Human Resources Research Organization on a research study that looked at students who used *i-Ready Personalized Instruction* regularly and performed well on the lessons. They found that such students scored significantly higher on the spring *i-Ready Inform* assessment than their peers who did not meet these criteria. We also found that students who used Personalized Instruction materials outperformed those who did not use them on major summative assessment programs (e.g., MCAS and SBAC Mathematics Summative Assessments). Finally, we observed that regular *i-Ready* users performed better than their counterparts on locally used tests (e.g., Acadience® Reading and Student Assessment of Growth and Excellence). Additional efficacy research can be found on our website at CurriculumAssociates.com/Research-and-Efficacy/i-Ready-Evidence-Impact - ESSAStudy2.

Acadience® is a registered trademark of Acadience Learning Inc.

12.7.3 National Norms

Norms for *i-Ready Inform* are intended to reflect the achievement level of students relative to a nationally representative sample of test-takers. Thus, norm sample representation and size are essential for supporting such interpretations. Chapter 8 describes the norming process undertaken by Curriculum Associates, including methods used to develop norms for *i-Ready Inform*, the stratified sampling approach, key characteristics of the norm samples, and the process of developing percentiles ranks associated with each scale score and calculated for each combination of testing window and grade for both Reading and Mathematics.

The chapter describes the methodology used to conduct the most recent updates to the norms and how the norms are monitored over time. The resulting 2024 National Norms are presented in tables along with guidance for their interpretation and use, in accordance with best industry practices (AERA et al., 2014). Norm development further adds to *i-Ready Inform*'s value by providing relative meaning to the scores, which aids teachers, administrators, and students in evaluating class performance against a broader educational landscape. Norms may also be used for identification for gifted and talented programs as well as for reclassification purposes.

12.7.4 English Learner Reclassification

Curriculum Associates recommends following state- or district-specific guidance for English Learner reclassification. In some cases, states may offer flexibility in the methodology for determining English Learner reclassification. In those cases, we recommend using multiple sources of evidence for making decisions about English Learner reclassification. *i-Ready Inform* data can be used as one source of evidence to determine if students are sufficiently proficient in reading the English language to participate effectively in a curriculum designed for pupils of the same age whose native language is English. Two determinations based on *i-Ready Inform* performance for their district as a whole, as well as individual students, must be made:

1. What is the English reading proficiency of average native English speakers in our district?
2. Which English Learners have demonstrated proficiency in reading English comparable to our average native English speakers?

To answer these questions, we recommend that spring *i-Ready Inform* for Reading scores be used as one of several evidence components in making decisions about reclassifying English Learners.

12.7.5 Algebra Readiness

To identify the most essential skills for algebra readiness, Curriculum Associates content experts extensively reviewed a number of sources, including the *i-Ready* middle school construct map, the Student Achievement Partners' progression to algebra, and others. As a result, a scale score of 541 was determined to be an appropriate recommendation for an algebra readiness performance-level standard. This scale score of 541 represents a Mid On Grade Level placement for a Grade 8 student, and it reflects sufficient command of crucial Algebra 1 prerequisites, including rational number arithmetic, modeling and solving real-world and mathematical problems, and interpreting and analyzing quantitative relationships.

While Curriculum Associates recommends a scale score of 541 [15] as an indicator of algebra readiness, individual districts may find that adjusting this number up or down results in a performance-level standard that is better suited to their local conditions. In making this determination, districts may want to consider:

- Their Algebra 1 scope and sequence and time allotted for remediation of pre-algebra skills
    - If the Algebra 1 curriculum incorporates a review of prerequisites, it may be reasonable to include students who are not yet proficient in those skills by applying a lower performance-level standard. However, if little time is available for remediation, it may be advisable to apply a higher, more exclusive performance-level standard.
- The performance of past students in Algebra 1 and their associated *i-Ready Inform* scores
    - Over time, districts can accumulate enough of this data to identify the performance-level standard that best predicts success in their Algebra 1 course.

12.7.6 Using *i-Ready Inform* to Address Gifted Identification Needs

Three important identification questions that *i-Ready Inform* scores are used to answer are: "Is this student a high-performing learner?," "Where along the continuum of the curriculum is the student located?," and "What are the instructional next steps for the student?" This information is provided as part of *i-Ready Inform*'s results for every student and can be used as evidence to support the evaluation of high-performing learners for their eligibility to be included in gifted programs.

i-Ready Inform has a vast body of validity evidence to support decisions about two things. First, as a standardized test of academic achievement, the total scale score can be used to describe a student's proficiency with respect to other learners who are on the same learning path. Second, *i-Ready Inform* results are designed to provide domain placement categories that describe the status of a student with respect to having achieved the objectives of Grades K–12 learning (in reading or mathematics). With this type of information, important decisions about instruction and classifications of students can be supported.

i-Ready Inform has two rigorous score types that can be used to gauge a student's readiness for gifted instruction. One score type is *i-Ready Inform*'s criterion-referenced grade-level placements. Because *i-Ready*

[15] Additional information about this recommendation is available upon request.

Inform's reports provide an excellent summary of what students who achieve a given *i-Ready Inform* placement level (e.g., Late On Grade Level, One Grade Level Above) can likely do after a mathematics or reading assessment, a student's potential to succeed in gifted instruction can be gauged based on their performance on *i-Ready Inform*.

The second score type provided by *i-Ready Inform* is norm-referenced scores. *i-Ready Inform* provides a representative national comparison group against which the ability of any single learner or the ability of a group of learners can be gauged using *i-Ready Inform*'s national percentile ranks. For more local decisions, data from district, school, or classroom implementations can be used to develop customized, local reference groups for norming purposes. Different state regulations may specify the use of either national or more local reference groups as part of their gifted identification procedures. *i-Ready Inform* can supply the information that meets the needs of these requirements.

12.8 Discussion

This chapter concludes the validity discussion that begins in Chapter 1 with our theory of action. In summary, the validity evidence presented in this technical manual supports the assessment claims outlined in Chapter 1. We used the five sources of validity evidence (AERA et al., 2014)—depicted in Figure 1.2—as the guiding framework to organize our validity evidence.

We also note that *i-Ready Inform* is situated within a theory of action, in which the end goal is to improve student outcomes. Although we acknowledge that this research extends beyond the boundaries of *i-Ready Inform* by including the use of our Personalized Instruction product, this is a frequent result of an *i-Ready Inform* administration and it uses the scores as inputs. To that end, we ended this chapter with a discussion of the kind of efficacy research we conduct on an ongoing basis and offered a few examples. As such, we offer evidence in support of the other types of validity evidence shown on the bottom row of Figure 1.2.

Chapter 13: References

Achieve, Inc. (2019). *A framework to evaluate cognitive complexity in mathematics assessments*. https://achievethecore.org/content/upload/SAP_Focus_Math_8.pdf#page=2

American Educational Research Association (AERA), the American Psychological Association (APA), and the National Council on Measurement in Education (NCME) Joint Committee on Standards for Educational and Psychological Testing. (2014). *Standards for educational and psychological testing*. AERA.

Briggs, D. (2020). *Memo on growth-score reliabilities* [White paper]. University of Colorado Boulder.

CAST (2024). *Universal design for learning guidelines version 3.0*. https://udlguidelines.cast.org.

Cizek, G. J., & Bunch, M. B. (2007). *Standard setting: A guide to establishing and evaluating performance standards on tests.* Sage.

Curriculum Associates. (2023). *Accessibility and accommodations update.*

Curriculum Associates. (2024a). *i-Ready Diagnostic 2024 national norms technical documentation for grades K–8* [Technical Report].

Curriculum Associates. (2024b). *i-Ready Diagnostic 2024 national norms technical documentation for grades 9–10* [Technical Report].

DeBell, M., & Krosnick, J. A. (2009). *Computing weights for American National Election Study survey data*. ANES Technical Report series, No. nes012427. American National Election Studies. https://electionstudies.org/wp-content/uploads/2018/04/nes012427.pdf

Egan, K. L., Schneider, M. C., & Ferrara, S. (2012). Performance-level descriptors: History, practice, and a proposed framework. In G. Cizek (Ed.), *Setting performance standards* (2nd ed., 79–106). Routledge.

Hambleton, R. J. & Pitoniak, M. J. (2006). Setting performance standards. In R. L. Brennan (Ed.), *Educational measurement* (4th ed., pp. 433–470). Praeger Publishers.

International Test Commission. (2017). *The ITC Guidelines for Translating and Adapting Tests* (Second edition). [www.InTestCom.org]

Kane, M. T. (2006). Validation. In R. L. Brennan (Ed.), *Educational measurement* (4th ed., pp. 17–64). American Council on Education/Praeger.

Kane, M. T. (2013). Validating the interpretations and uses of test scores. *Journal of Educational Measurement, 50* (1), 1–73.

Laitusis, C. C. & Karvonen M. (2024, April). *A case for reimagining universal design in assessment systems*. Paper presented at the annual meeting of the National Council on Measurement in Education, Philadelphia, PA.

Lewis, D. M., Mitzel, H. C., & Green, D. R. (1996). Standard setting: A bookmark approach. In D. R. Green (Chair), *IRT-based standard setting procedures utilizing behavioral anchoring*. Symposium conducted at the Council of Chief State School Officers National Conference on Large-Scale Assessment, Phoenix, AZ.

Linacre, J.M. (2012). *WINSTEPS MINISTEP Rasch-model computer programs*. Chicago: http://www.winsteps.com.

Linacre, J. M. & Wright, B. D. (1989). Mantel-Haenszel DIF and PROX are equivalent. *Rasch Measurement Transactions, 3*, (2), 52–53.

Livingston, S. A., & Zieky, M. J. (1982). *Passing scores: A manual for setting standards of performance on educational and occupational tests* [Whitepaper]. Educational Testing Service. https://www.ets.org/Media/Research/pdf/passing_scores.pdf.

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

National Research Council. (2001). *Knowing what students know: The science and design of educational assessment*. The National Academies Press. https://doi.org/10.17226/10019

Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Danmarks Paedagogiske Institut: Copenhagen.

Rudner, L. M. (2000). Computing the expected proportions of misclassified examinees. *Practical Assessment, Research, and Evaluation, 7*(14). https://doi.org/10.7275/an9m-2035

Rudner, L. M. (2005). Expected classification accuracy. *Practical Assessment, Research, and Evaluation*, *10*(1), 13.

Ryan, R. M., & Deci, E. L. (2018). *Self-determination theory: Basic psychological needs in motivation, development, and wellness*. The Guilford Press. https://doi.org/10.1521/978.14625/28806

Sireci, S.G., Thissen, D., & Wainer, H. (1991). On the reliability of testlet-based tests. *Journal of Educational Measurement, 28*, 237–247.

Webb, N. L. (1997). *Criteria for alignment of expectations and assessments in mathematics and science education* (Council of Chief State School Officers and National Institute for Science Education Research Monograph No. 6). Madison: University of Wisconsin, Wisconsin Center for Education Research.

Webb, N. L. (1999). *Alignment study in language arts, mathematics, science, and social studies of state standards and assessments for four states: A study of the state collaborative on assessment & student standards (SCASS), technical issues in large-scale assessment (TILSA).* Council of Chief State School Officers.

Webb, N. L. (2007). Issues related to judging the alignment of curriculum standards and assessments. *Applied Measurement in Education, 20*(1), 7–25.

Willett, J. (1989). Questions and answers in the measurement of change. In E. Z. Rothkopf (Ed.), *Review of research in education* (Vol. 15, pp. 345–422). AERA.

Wilson, M. R. (2005). *Constructing measures: An item response modeling approach.* Psychology Press.

Worldwide Web Consortium. (2022). *Web content accessibility guidelines (WCAG)*. W3C: Web accessibility initiative (WAI). https://www.w3.org/WAI/standards-guidelines/wcag

Wyse, A. E., & Hao, S. (2012). An evaluation of item response theory classification accuracy and consistency indices. *Applied Psychological Measurement*, *36*(7), 602-624. https://journals.sagepub.com/doi/pdf/10.1177/0146621612451522

Zwick, R., Thayer, D. T., & Lewis, C. (1999). An empirical Bayes approach to Mantel-Haenszel DIF analysis. *Journal of Educational Measurement, 36*(1), 1–28.