



2019 No. 062

# An Impact Evaluation of Mathematics and Reading *i-Ready Instruction* for Elementary Grades

## Final Report

**Prepared for:** Curriculum Associates  
153 Rangeway Road  
North Billerica, MA 01862

**Authors:** Rebecca Norman Dvorak  
Bruce Randel (Century Analytics)  
Matthew Swain

**Date:** September 16, 2019

# An Impact Evaluation of Mathematics and Reading *i-Ready Instruction* for Elementary Grades

## Table of Contents

Introduction .....	1
Defining i-Ready Instruction .....	1
Research Questions .....	1
Methodology .....	2
Design Decisions .....	2
Cluster-Level Design .....	2
Outcome Measure .....	2
Required Number of Schools.....	3
School and Student Selection Criteria.....	3
Initial School and Student Selection .....	4
i-Ready Instruction Schools .....	4
Comparison Schools .....	6
Matching .....	7
School-Level Matching .....	7
Student-Level Matching.....	9
Impact Analysis.....	12
Impact Analysis Results.....	13
Mathematics Results .....	13
Reading Results .....	14
Mathematics and Reading Student-Achievement Gains .....	14
Impact Summary and Discussion.....	19
Quality Control Procedures .....	20
References .....	21
Appendix A. <i>i-Ready Instruction</i> Theory of Action .....	A-1
Appendix B. i-Ready® Diagnostic Reliability Information from Curriculum Associates’ Technical Report.....	B-1
Appendix C. Impact HLM Coefficients.....	C-1

## Table of Contents (Continued)

### List of Tables

Table 1. School Characteristics of i-Ready Instruction Schools at Key Points in the Filtering Process.....	6
Table 2. School Characteristics of Comparison Schools at Key Points in the Filtering Process..	7
Table 3. School Demographic Variables and Effect Size Differences Between the i-Ready Instruction (Treatment) and Comparison Group. ....	8
Table 4. Urbanicity Differences Between the i-Ready Instruction (Treatment) and Comparison Group.....	9
Table 5. Mean Difference in Baseline Student Achievement in Mathematics and Reading for Students in the i-Ready Instruction (treatment) and Comparison Matched Schools, prior to Student-Level Matching. ....	10
Table 6. Mathematics Baseline Equivalence Statistics for i-Ready Instruction (Treatment) and Comparison Groups by Grade .....	11
Table 7. Reading Baseline Equivalence Statistics for i-Ready Instruction (Treatment) and Comparison Groups by Grade .....	12
Table 8. Impact Analysis Results for i-Ready Diagnostic Schools Compared to Comparison Schools for Mathematics Student Achievement at grades K–5. ....	15
Table 9. Impact Analysis Results for i-Ready Diagnostic Schools Compared to Comparison Schools for Reading Student Achievement at grades K–5. ....	16
Table C.1. HLM Results for Kindergarten Mathematics.....	C-1
Table C.2. HLM Results for First Grade Mathematics. ....	C-2
Table C.3. HLM Results for Second Grade Mathematics. ....	C-2
Table C.4. HLM Results for Third Grade Mathematics. ....	C-3
Table C.5. HLM Results for Fourth Grade Mathematics. ....	C-3
Table C.6. HLM Results for Fifth Grade Mathematics. ....	C-4
Table C.7. HLM Results for Kindergarten Reading.....	C-5
Table C.8. HLM Results for First Grade Reading. ....	C-6
Table C.9. HLM Results for Second Grade Reading. ....	C-6
Table C.10. HLM Results for Third Grade Reading. ....	C-7
Table C.11. HLM Results for Fourth Grade Reading. ....	C-7
Table C.12. HLM Results for Fifth Grade Reading. ....	C-8

# An Impact Evaluation of Mathematics and Reading *i-Ready Instruction* for Elementary Grades

## Introduction

Founded in 1969, Curriculum Associates provides a variety of educational products and services with the goal of improving education for students and teachers. Two Curriculum Associates products include *i-Ready<sup>®</sup> Diagnostic* (available for K–12) and *i-Ready<sup>®</sup> Instruction* (available for K–8). The *i-Ready Diagnostic* assessments (a) are online, computer-adaptive assessments that pinpoint student needs at the sub-skill level and (b) help monitor the extent to which students are on track to achieve end-of-year targets. *i-Ready Instruction* provides online, individualized instruction for students and can be used with the *i-Ready Diagnostic*.

The Human Resources Research Organization (HumRRO) conducted an evaluation to examine the impact of *i-Ready Instruction* on mathematics and reading achievement for students in elementary grades K–5. This study was designed to meet the required rigor of the What Works Clearinghouse (WWC) 4.0 standards to achieve a rating of *Meets WWC Group Design Standards with Reservations* (WWC, 2017a), and to meet guidelines for a Level 2 (or *Moderate*) rating for the Every Student Succeeds Act (ESSA) guidance for evidence-based research (U.S. Department of Education, 2016). This was achieved by using quasi-experimental design (QED), establishing baseline equivalence between the treatment and comparison groups, using an outcome measure acceptable by WWC, including baseline achievement as a covariate, and sampling design that mitigates the effects of any confounding factors.

## Defining *i-Ready Instruction*

*i-Ready Instruction* is an online personalized instruction program aligned to college- and career-ready standards that includes engaging multimedia instruction and progress monitoring into online lessons. Lessons are intended to provide a consistent best-practice lesson structure and build students' conceptual understanding. *i-Ready Instruction* is intended to be used in conjunction with *i-Ready Diagnostic* which monitors student progress and identifies student performance in mathematics. This diagnostic information helps target student-specific intervention, which is provided through *i-Ready Instruction*.

Curriculum Associates has developed a Theory of Action (TOA) noting the key implementation components of *i-Ready Instruction*, the intended intermediate outcomes, and the intended long-term outcomes. The key implementation components highlight actions by students, teachers, and leaders recommended to obtain the long-term outcome of improved student learning in reading and mathematics. Among others, the key components include support at the school and district leadership level, monitoring of student progress by teachers, and student use of *i-Ready Instruction* to work through a personalized, scaffolded instruction path. The *i-Ready Instruction* TOA is provided in Appendix A.

Curriculum Associates provides guidance to districts and schools on how to implement *i-Ready Instruction* to best benefit student learning (Curriculum Associates, 2019). Guidance indicates students receive greater gains with at least 30 minutes of *i-Ready Instruction* use for each subject area. In addition, Curriculum Associates recommends use for at least 18 weeks between a fall *i-Ready Diagnostic* administration and a spring administration (Curriculum Associates, 2018).

## Research Questions

The primary purpose of this evaluation was to estimate the impact of using *i-Ready Instruction* on student mathematics and reading achievement. Our research was focused on two primary research questions, addressed separately for each grade from kindergarten to grade 5.

- What was the impact of *i-Ready Instruction* use in mathematics on student mathematics achievement as compared to schools and students who do not use *i-Ready Instruction* in mathematics?
- What was the impact of *i-Ready Instruction* use in reading on student reading achievement as compared to schools and students who do not use *i-Ready Instruction* in reading?

Our hypotheses were that student achievement for mathematics and reading would be higher for students in schools that implement *i-Ready Instruction* with fidelity according to the criteria described in the TOA and user guidance (Curriculum Associates, 2019). These hypotheses were based on the belief that students benefit from the *i-Ready Instruction* targeted to their specific needs in mathematics and reading.

## Methodology

This section describes initial design decisions, school and student selection and matching, analysis, and results.

### Design Decisions

#### Cluster-Level Design

Though it is most common for Curriculum Associates to sell a product, or set of products, at the district-level, usage data and on-the-ground experiences support the notion that the decision to use *i-Ready Instruction*, and to what degree, happens at the school-level. Therefore, we determined the unit of assignment for this study to be schools. The unit of observation in this study was the student and our analyses focused on the impact of *i-Ready Instruction* use on student achievement in mathematics and reading for students and schools using *i-Ready Instruction* with fidelity.

#### Outcome Measure

The mathematics and reading *i-Ready Diagnostic* assessments were designed to be aligned to today's college- and career-ready standards and to provide results that inform student placement decisions, offer explicit instructional advice, and prescribe resources for targeted instruction and intervention. The *i-Ready Diagnostic* is currently used by more than 6.5 million students across the United States – and users sometimes incorporate other *i-Ready* products (i.e., *Instruction*, *Teacher Resources*), though this is not a requirement.

To provide evidence the *i-Ready Diagnostic* measures skills consistent with student expectations, Curriculum Associates has conducted multiple linking studies to examine the correlation of *i-Ready Diagnostic* scores with scores from national and state summative tests for mathematics and reading at grades 3 – 8. Linking studies using 2016 data examined the correlation between *i-Ready Diagnostic* and the Smarter Balanced summative assessments, the Partnership for Assessment of Readiness for College and Careers (PARCC), and multiple state

testing programs (North Carolina, New York, Tennessee, Ohio, Mississippi, Michigan, Indiana, Florida, and Georgia). These studies show strong correlations between *i-Ready Diagnostic* scores and scores on these national and state tests. The average correlations across grades between the *i-Ready Diagnostic* for mathematics and the national and state mathematics assessments ranged from 0.82 (North Carolina End-of-Grade assessments) to 0.88 (Smarter Balanced and Michigan M-STEP). For reading, the correlations ranged from 0.78 (Tennessee TNReady) and 0.85 (Smarter Balanced). These studies provide evidence that the *i-Ready Diagnostic* content is highly consistent with what students across the United States are expected to learn (Curriculum Associates, 2019). Curriculum Associates has also recently completed linking studies for Colorado, Kentucky, and Missouri. In addition, Curriculum Associates has commissioned Odell Education and others to complete alignment studies to demonstrate the degree of alignment between the content on *i-Ready Diagnostic* and current sets of state standards. Specifically, they have conducted alignment studies for the Common Core State Standards (CCSS), and for the Louisiana, Indiana, Ohio, Michigan, Florida, and South Carolina state standards.

Curriculum Associates released *i-Ready Diagnostic* in the summer of 2011. Since then, *i-Ready* has been reviewed and approved at the state level as an assessment, instructional resource, or intervention in Arizona, California, Colorado, Connecticut, Delaware, Florida, Georgia, Idaho, Indiana, Massachusetts, Mississippi, Nevada, New Mexico, New York, North Carolina, Ohio, Oklahoma, Oregon, Tennessee, Utah, and Virginia.

*i-Ready Diagnostic* for mathematics and reading measures achievement aligned to common mathematics and reading content and skills with demonstrated test score reliability (see Appendix B). Marginal reliability ranges from .91 to .97 and test-retest reliabilities range from .70 to .86 for mathematics and reading in kindergarten through grade 5. Therefore, this assessment meets the WWC 4.0 standards for an acceptable outcome measure (WWC, 2017a). The *i-Ready Diagnostic* was used as the baseline and outcome measure for all students participating in this study (i.e., *i-Ready Instruction* students and comparison group students).

The *i-Ready Diagnostic* is intended to be administered in a standardized manner across schools (Curriculum Associates, 2019b). Specifically, at grades 6–8 teachers are to schedule the first (fall) Diagnostic 2 – 3 weeks into the school year in two 45- to 50-minute sessions. They are to test technology to ensure proper function and have pencils and paper available as scratch paper. Test administrators are to provide instructions to their students and motivate them to do their best. As students are testing, teachers are to monitor students.

### **Required Number of Schools**

We conducted power analyses using Optimal Design software (Spybrook, et al., 2011) to identify the total number of schools required at each grade level for sufficient statistical power to reject the null hypothesis that there is no difference in student mathematics or reading achievement between the treatment and comparison group. Statistical power is influenced by various factors. We used data from previous studies HumRRO conducted using *i-Ready Diagnostic* as an outcome to estimate conservative and optimistic parameters for use in the power analysis. These parameters were: (a) 0.90 for the relationship between the baseline and outcome variable, (b) 40 and 60 for the number of students per school, and (c) 0.10 and 0.30 for the intraclass correlation coefficient (ICC). Optimal Design estimated that approximately 40 schools would be needed to achieve statistical power of 0.80. A 0.80 power level provides an 80% chance of detecting a statistically significant difference if one exists with 95% confidence.

## School and Student Selection Criteria

This section provides an overview of the process of selecting schools and students eligible for inclusion in the study. To make eligibility determinations, we used a combination of publicly available data, and Curriculum Associates *i-Ready Diagnostic* and *i-Ready Instruction* usage data. Data provided by MDR<sup>1</sup> summarized recent publicly available school-level enrollment data, free and reduced price lunch participation data, and urbanicity data compiled from the Department of Education Common Core of Data and other sources. We used data available through the Department of Education Civil Rights Data Collection Website<sup>2</sup> to obtain the most recent publicly available school-level information on percent white, percent limited English proficiency (LEP), and percent students with disabilities (SWD).

HumRRO started with Curriculum Associates *i-Ready* 2017-18 datasets for mathematics and reading. Because the *i-Ready Diagnostic* was selected as our student baseline and outcome measure for the treatment and comparison group, we included only students with both fall 2017 and spring 2018 *i-Ready Diagnostic* scores for mathematics and reading. Because the *i-Ready Diagnostic* was selected as our student-baseline and outcome measure for the treatment and comparison group, a student would need to have both fall 2017 and spring 2018 *i-Ready Diagnostic* scores for mathematics and reading for inclusion in our study.

For the purpose of generalizability and interpretability, we identified elementary schools with a similar set of characteristics. Particularly, we filtered our file to include only schools meeting the following criteria: (a) public institutions, (b) not charter or magnet schools, and (c) K – 5 institutions. We allowed schools with and without pre-K programs into the study; however, all included schools had grades K – 5, and grade 5 was the highest grade offered. Restricting our sample in this manner helped reduce potential confounds due to institution-type.

We next sought to identify schools qualified for inclusion in either the treatment or comparison groups. We separated out those schools with at least some *i-Ready Instruction* use (potential treatment) and those including only students with *i-Ready Diagnostic* use (potential comparison). Additional filtering was required for both groups to ensure we included schools meeting minimum fidelity requirements. These additional steps are described separately for each group.

### *i-Ready Instruction* Schools

One of our primary goals for this study was to include schools that implemented *i-Ready Instruction* for mathematics and reading with fidelity. Specifically, we wanted to ensure that students in our schools were using *i-Ready Instruction* in a manner consistent with the logic model and the guidance provided by Curriculum Associates to its users. HumRRO, Century Analytics, and Curriculum Associates worked collaboratively to establish the definition of implementing *i-Ready Instruction* with fidelity at the student-level. First, based on current *i-Ready Diagnostic* guidance Curriculum Associates provides to its users, we determined a student would need to use *i-Ready Instruction* for an average of 30 minutes or more per subject, per week of use during the 2017-18 school year. Second, a student would need to use *i-Ready Instruction* for at least 18 weeks between the fall and spring *i-Ready Diagnostic* administrations to be considered using with fidelity. Based on this definition, HumRRO filtered the *i-Ready*

---

<sup>1</sup> <https://mdrededucation.com/our-story/>

<sup>2</sup> <https://www2.ed.gov/about/offices/list/ocr/data.html>

*Instruction* dataset to include only schools with students meeting this definition for mathematics and reading.

Curriculum Associates staff indicated most schools were not fully implementing *i-Ready Instruction* until the second school-year of use. They found it generally took some time for teacher buy-in of the new intervention and to find ways to carve out the time recommendation for student use. Therefore, we determined it was appropriate to include schools in their second year of *i-Ready Instruction* use in our study to achieve our goal of examining the impact of use with fidelity. We used 2015-16 and 2016-17 *i-Ready Diagnostic* and *Instruction* usage data with our 2017-18 data to make this determination. Schools were kept in our *i-Ready Instruction* school list if (a) they did not have any *i-Ready Instruction* use in 2015-16, (b) they had at least some *i-Ready Instruction* use in 2016-17, and (c) included students who met *i-Ready Instruction* fidelity criteria for the 2017-18 school year. There were 452 schools that met this requirement.

For an elementary school to be using with fidelity, we determined the majority of students should be meeting the student-level *i-Ready Instruction* usage criteria. Using publicly available enrollment data, we calculated the approximate percentage of students per grade using *i-Ready Instruction* with fidelity. We then calculated the average percentage of students meeting the fidelity criteria across grades. Schools that had an average of 50% of students or more across all grades meeting the fidelity criteria were identified as those meeting all implementation fidelity criteria. This left us with 75 eligible schools.

As a final step, we sought confirmation from Curriculum Associates staff who work directly with the identified schools to ensure that (a) they were, in fact, using *i-Ready Instruction* with fidelity as our data suggested, and (b) that these schools were using *i-Ready Instruction* with their general population of students. The Curriculum Associates staff confirmed that the schools we identified were using *i-Ready Instruction* as our data indicated.

Table 1 presents key school characteristics for *i-Ready Instruction* schools at key points in the filtering process described above. Step 1 includes all public K–5 schools in their second year of *i-Ready Instruction* implementation who have at least one student meeting minimum usage requirements for *i-Ready Instruction*. The final column includes schools after filtering for Step 2 and Step 3 - Step 2 includes only those schools with, on average, 50% students or more eligible for inclusion across grades. Step 3 summarizes data of our final set of schools confirmed by Curriculum Associates to be using *i-Ready Instruction* with a general education population of students. For this particular study, Step 3 did not reduce our sample beyond Step 2. As shown, the student demographic make-up of students changes only slightly between Step 1 and the final set of schools meeting all school- and student-level eligibility criteria for the *i-Ready Instruction* group. The urbanicity make-up of schools meeting full eligibility requirements, however, changed between the initial set of all public elementary schools with any *i-Ready Instruction* use to the final set of schools meeting our eligibility requirements – particularly, there was a smaller percentage of urban schools in the final filtered group compared to the earlier steps in the filtering process.

**Table 1. School Characteristics of *i-Ready Instruction* Schools at Key Points in the Filtering Process.**

Variable	Step 1. School-Level Second Year <i>i-Ready Instruction</i> Users Criteria Applied (n = 452)	Step 2. School-Level 50% Minimum <i>i-Ready Instruction</i> Student Use Criteria Applied, and Step 3. Final Confirmed for Matching*. (n = 75)
Percent FRL	65.6%	65.6%
Percent LEP	15.0%	15.5%
Percent SWD	15.0%	15.6%
Percent White	44.1%	43.7%
Urbanicity		
Rural	16.3%	19.7%
Suburban	40.0%	50.7%
Town	12.1%	14.1%
Urban	31.7%	15.5%

\*Step 2 and Step 3 include the same sample of schools because all identified schools were confirmed by Curriculum Associates as *using i-Ready Instruction* with fidelity.

### Comparison Schools

Filtering comparison schools began by removing schools with past exposure to *i-Ready Instruction*. HumRRO used *i-Ready Instruction* data from the two school years prior to the study year to identify and remove schools with any use of *i-Ready Instruction* in mathematics or reading for 2015-16 and 2016-17. To make sure our comparison sample was a pure comparison, we removed schools with as few as one student who had used *i-Ready Instruction* for mathematics or reading either of those two years, as indicated by the usage data. Following this filter, our comparison dataset included 113 schools having *i-Ready Diagnostic* data for fall and spring mathematics and reading.

Next, following similar guidelines to our *i-Ready Instruction* sample, we identified schools where, at minimum, an average of 50% of students across grades K-5 were administered the *i-Ready Diagnostic* for mathematics and reading in the spring and fall of 2017-18. At this point, we identified 87 schools our comparison dataset.

Finally, consistent with the approach taken for *i-Ready Instruction* schools, we provided our filtered list of comparison schools to the Curriculum Associates staff who work directly with the schools to confirm our schools used *i-Ready Diagnostic* with their general student population. We learned that one school was a special education school and therefore did not meet our definition of a traditional public institution. We removed this school from our eligible comparison dataset. This left 86 schools meeting eligibility requirements for inclusion as comparisons.

Table 2 presents school characteristics for comparison schools at key points in the filtering process described above. Step 1 included all schools identified as public, K – 5 institutions with

*i-Ready Diagnostic* users who have not even one *i-Ready Instruction* student user in 2017-18, 2016-17 or 2015-16. 2 included only those schools with, on average, most students with available *i-Ready Diagnostic* scores for reading and mathematics across grades. The final column, Step 3, are those schools from Step 3 that were confirmed by Curriculum Associates staff. These schools meet all comparison eligibility criteria. As shown, the schools eligible for inclusion in the final comparison group (Step 3) are of similar demographic and urbanicity make-up to those in Step one prior to applying filters.

**Table 2. School Characteristics of Comparison Schools at Key Points in the Filtering Process.**

Variable	Step 1. School-Level Previous <i>i-Ready Instruction</i> Exposure Criteria Applied (n = 113)	Step 2. 50% Minimum <i>i-Ready Diagnostic</i> Student Use Criteria Applied (n = 87)	Step 3. Final Confirmed for Matching (n = 86)
Percent FRL	55.6%	55.0%	55.3%
Percent LEP	12.8%	12.6%	12.7%
Percent SWD	13.5%	13.7%	13.6%
Percent White	62.6%	63.7%	63.7%
Urbanicity			
Rural	15.3%	7.0%	7.0%
Suburban	41.4%	41.9%	41.9%
Town	5.4%	4.7%	4.7%
Urban	37.8%	46.5%	46.5%

## Matching

### School-Level Matching

Once we had identified eligible *i-Ready Instruction* and comparison schools, described above, we matched schools to ensure key school-level demographic characteristics were similar between the *i-Ready Instruction* schools and comparison schools selected for analysis. Schools were matched on the following variables:

- Percent students eligible for free or reduced lunch (FRL)
- Percent students limited English proficient (LEP)
- Percent students with disabilities (SWD)
- Percent students white

We selected these variables as they are known to be related to student achievement and reliable data are available for public schools across the country, including all schools meeting the criteria for our sample.

We used a logistic regression model to compute a propensity score for each school in the *i-Ready Instruction* pool and the comparison pool (Guo & Fraser, 2010). Propensity scores range between 0 and 1 and indicate the probability that each school belonged to the *i-Ready*

*Instruction* group given their values on the matching variables. We used a nearest neighbor matching approach (Stuart, 2010) to match each *i-Ready Instruction* school with an available comparison school with the closest propensity score. We matched without replacement – meaning all schools could only be matched once. Following the first round of matching including all 63 *i-Ready Instruction* schools, baseline equivalence was not achieved. Baseline equivalence (BE) is the similarity in means of the matching variables between the groups as measured by a standardized effect size. An examination of the propensity scores revealed that two of the *i-Ready Instruction* schools had scores that were outside the range of values observed for both the intervention and comparison groups (i.e., the area of common support). Thus, we used trimming to remove these two schools (Stuart, 2010) and conducted the matching procedure again.

Following the above steps, we achieved BE between the two study groups of schools on these characteristics with no effect size larger than 0.25. According to WWC guidance, school demographic variables meet baseline equivalence if they are below 0.50 for mathematics (WWC, 2015) and 0.25 for reading (WWC, 2014). Table 3 provides descriptive statistics for each matching variable for the *i-Ready Instruction* (treatment) and comparison group, as well as the effect size of the difference between the means for the final sample of study schools.

**Table 3. School Demographic Variables and Effect Size Differences Between the *i-Ready Instruction* (Treatment) and Comparison Group.**

Variable	<i>i-Ready Instruction</i> (N = 61)		Comparison (N = 61)		Adj. Mean Diff.	Effect Size
	Mean	SD	Mean	SD		
Percent FRL	64.6%	24.1	61.1%	20.2	3.5%	0.16
Percent LEP	15.0%	15.5	14.4%	12.4	0.6%	0.04
Percent SWD	15.0%	4.9	14.7%	4.1	0.2%	0.06
Percent White	49.4%	28.2	55.2%	23.8	5.8%	0.22

Though not part of our matching criteria, we also examined the final urbanicity make-up of our *i-Ready Instruction* and comparison schools. These data reveal that a similar number of suburban schools were included in our final *i-Ready Instruction* and comparison school samples; however, there were more urban schools in the comparison set and more rural and town schools in the *i-Ready Instruction* set (see Table 4).

**Table 4. Urbanicity Differences Between the *i-Ready Instruction (Treatment)* and Comparison Group.**

Urbanicity	<i>i-Ready Instruction</i> ( <i>n</i> =61)	Comparison ( <i>n</i> = 61)
Rural	21.1%	3.3%
Suburban	52.6%	49.2%
Town	15.8%	1.6%
Urban	10.5%	45.9%

\*Four *i-Ready Instruction* schools were missing urbanicity data.

### **Student-Level Matching**

Following school-level matching, we compared the adjusted mean difference in mathematics and reading achievement at the student level. The adjusted mean difference between students at each grade level was estimated using a mixed model that nested students within schools and estimated the difference between the treatment group and the comparison group at the school level. This model mirrors the model used to estimate impacts, which is described in the next section. For each grade level, we compared the mathematics and reading fall *i-Ready Diagnostic* scores of all students in the matched treatment and comparison schools. Within the treatment schools, only those students who met the inclusion criteria for using *i-Ready Instruction* with fidelity were included in the sample (e.g., students using *i-Ready Instruction* for an average of at least 30 minutes per week for a minimum of 18 weeks). Table 5 provides the mean Fall 2017 mathematics and reading achievement, as measured by the mathematics and reading *i-Ready Diagnostic*, for the full set of pre-matched students at each grade level in matched schools. Because effect size differences were greater than the WWC threshold of 0.25 for baseline equivalence for at least one subject area at each grade level (WWC, 2017b), student-level matching was required for all six grades.

**Table 5. Mean Difference in Baseline Student Achievement in Mathematics and Reading for Students in the i-Ready Instruction (treatment) and Comparison Matched Schools, prior to Student-Level Matching.**

Grade	i-Ready Inst. N	Comp. N	Fall 2017 Mathematics i-Ready Diagnostic		Adj. Mean Diff.	Effect Size	Fall 2017 Reading i-Ready Diagnostic		Adj. Mean Diff.	Effect Size
			i-Ready Inst. Mean (SD)	Comp Mean (SD)			i-Ready Inst. Mean (SD)	Comp Mean (SD)		
K	2949	1267	349.7 (21.3)	343.5 (21.9)	6.20	0.29	354.5 (31.3)	343.9 (29.0)	0.56	0.34
1	4249	3096	379.3 (22.7)	374.8 (22.5)	4.45	0.20	409.6 (40.8)	399.2 (39.1)	10.43	0.26
2	4755	3323	406.8 (25.1)	400.0 (23.9)	6.80	0.28	468.1 (48.6)	454.1 (48.3)	13.92	0.29
3	4653	3637	432.4 (25.3)	422.4 (26.8)	10.01	0.39	507.1 (49.8)	491.9 (50.8)	15.22	0.30
4	4537	3977	456.1 (27.8)	444.6 (28.0)	11.54	0.41	536.5 (49.2)	525.2 (50.4)	11.30	0.23
5	3854	3741	472.4 (28.4)	462.4 (30.0)	10.00	0.34	558.9 (50.5)	548.1 (51.2)	10.83	0.21

Similar to school-level matching methodology, we used propensity score matching to obtain a matched sample of *i-Ready Instruction* and Comparison students separately at each grade K–5. We included only students from our 61 *i-Ready Instruction* and 61 comparison baseline equivalent schools. Because at each grade-level our comparison group of students was smaller than our *i-Ready Instruction* group, we computed propensity scores based on the probability each student would be in the comparison group. Next, we matched each comparison student to an *i-Ready Instruction* student using a nearest neighbor matching approach separately at each grade level. This means that for each comparison student we matched to the *i-Ready Instruction* student with the closest propensity score. We matched without replacement; all students could only be matched once.

This approach resulted in baseline equivalence at grades K–3. For grades 4 and 5, we were required to trim 15 and 60 students from the comparison group in grades 4 and 5, respectively, to obtain students matched on baseline mathematics and reading achievement. Trimming was conducted by removing comparison students with propensity scores outside the area of common support.

Tables 6 and 7 provide student-level descriptive statistics on mathematics and reading achievement, respectively, for the study sample used in the impact analyses. The adjusted mean difference between students at each grade level was estimated using a mixed model that nested students within schools and estimated the difference between the treatment group and the comparison group at the school level.

**Table 6. Mathematics Baseline Equivalence Statistics for *i-Ready Instruction* (Treatment) and Comparison Groups by Grade**

Grade	Group	Schools	Students	i-Ready Mean	i-Ready SD	Adj Mean Diff (SE)	Effect Size
K	<i>i-Ready Instruction</i>	60	1267	344.18	21.45	0.40 (2.06)	0.019
	Comparison	30	1267	343.54	21.85		
1	<i>i-Ready Instruction</i>	61	3096	375.33	22.32	-0.07 (1.42)	-0.003
	Comparison	59	3096	374.82	22.50		
2	<i>i-Ready Instruction</i>	61	3323	400.30	23.67	0.94 (1.49)	0.039
	Comparison	60	3323	400.02	23.89		
3	<i>i-Ready Instruction</i>	61	3637	422.84	20.65	2.12 (1.51)	0.089
	Comparison	61	3637	422.39	26.78		
4	<i>i-Ready Instruction</i>	61	3962	449.55	24.43	6.29 (1.78)	0.242
	Comparison	61	3962	444.92	27.49		
5	<i>i-Ready Instruction</i>	61	3681	469.62	26.97	6.60 (1.94)	0.239
	Comparison	61	3681	463.76	28.14		

Note. ICC = intraclass correlation, SD = standard deviation of i-Ready scores, Adj Mean Diff = adjusted mean difference between *i-Ready Instruction* and Comparison groups, and SE = standard error of the adjusted mean difference.

**Table 7. Reading Baseline Equivalence Statistics for *i-Ready Instruction (Treatment)* and Comparison Groups by Grade**

Grade	Group	Schools	Students	i-Ready Mean	i-Ready SD	Adj Mean Diff (SE)	Effect Size
K	<i>i-Ready Instruction</i>	60	1267	343.49	29.06	-0.82 (2.63)	-0.028
	Comparison	30	1267	343.89	28.99		
1	<i>i-Ready Instruction</i>	61	3096	400.79	37.67	0.47 (2.36)	0.012
	Comparison	59	3096	399.15	39.05		
2	<i>i-Ready Instruction</i>	61	3323	455.12	47.11	2.26 (3.02)	0.047
	Comparison	60	3323	454.13	48.29		
3	<i>i-Ready Instruction</i>	61	3637	492.77	46.40	4.24 (3.13)	0.087
	Comparison	61	3637	491.88	50.78		
4	<i>i-Ready Instruction</i>	61	3962	530.53	48.60	7.64 (3.25)	0.155
	Comparison	61	3962	525.44	50.18		
5	<i>i-Ready Instruction</i>	61	3681	555.97	50.35	7.56 (3.58)	0.151
	Comparison	61	3681	549.64	49.78		

Note. ICC = intraclass correlation, SD = standard deviation of i-Ready scores, Adj Mean Diff = adjusted mean difference between *i-Ready Instruction* and Comparison groups, and SE = standard error of the adjusted mean difference.

### Impact Analysis

Following the selection of baseline equivalent students, we used hierarchical linear modeling (HLM) to address our two research questions to estimate the impact of *i-Ready Instruction* use on student achievement in mathematics and reading. We implemented a two-level model to account for the clustered nature of the data with students nested within schools. Because effect size differences between the treatment and comparison on student achievement at baseline fell between 0.05 and 0.25 standard deviations, we included baseline mathematics and reading achievement in the model as covariates. We ran separate models at each grade level and for each content area (e.g., mathematics and reading) for a total of 12 analyses.

We specified Level 1 of the model as:

$$Y_{ij} = \beta_{0j} + \beta_{1j}(\text{PRE\_MATH})_{ij} + \beta_{2j}(\text{PRE\_READ})_{ij} + e_{ij}$$

where  $Y_{ij}$  is the spring i-Ready mathematics or reading diagnostic score for student  $i$  in school  $j$ .  $\beta_{0j}$  is the adjusted mean outcome for students in school  $j$ .  $\beta_{1j}$  is the adjusted difference in outcome due to the student's baseline score in mathematics (school mean centered).  $\beta_{2j}$  is the adjusted difference in outcome due to the student's baseline score in reading (school mean centered).  $e_{ij}$  is the random error in the achievement outcome associated with student  $i$  in school  $j$  not accounted for in the model.

We specified Level 2 of the model as:

$$\beta_{0j} = \gamma_{00} + \gamma_{01}(\text{GROUP})_j + \gamma_{02}(\text{PRE\_MATH}_{.j} - \text{PRE\_MATH}_{..})_j + \gamma_{03}(\text{PRE\_READ}_{.j} - \text{PRE\_READ}_{..})_j + \sum \gamma_{\kappa}(\text{SCHOOL})_j + u_{0j}$$

where  $\gamma_{00}$  is the adjusted comparison group grand mean of the outcome,  $\gamma_{01}$  is the adjusted mean difference in the outcome between school study groups, and GROUP is an indicator variable coded as 1 for schools in the *i-Ready Instruction* group and 0 for schools in the comparison group.  $\gamma_{02}$  is the regression slope of the school-level baseline achievement in mathematics (grand mean centered), and  $\gamma_{03}$  is the regression slope of the school-level baseline achievement in reading (grand mean centered).  $\gamma_{02}$  and  $\gamma_{03}$  are included to explain additional between-school variance not explained in level 1 of the model.  $\gamma_{\kappa}$  is a vector of school-level demographic dummy variables added to increase statistical precision.  $u_{0j}$  is the random error in the achievement outcome associated with school  $j$ .

The student-level covariates used in each analysis were:

- *i-Ready Diagnostic* mathematics baseline performance
- *i-Ready Diagnostic* reading baseline performance

The school-level covariates included:

- Group membership (*i-Ready Instruction* or comparison)
- Average *i-Ready Diagnostic* baseline score in mathematics
- Average *i-Ready Diagnostic* baseline score in reading
- Demographic variables
  - Percent white students
  - Percent Limited English Proficient (LEP) students
  - Percent of students eligible for free and reduced-price lunch (FRL)
  - Percent of students with disabilities (SWD)

The above variables were selected for inclusion in the model because they are exogenous and could reasonably be expected to be related to student achievement. To indicate the size of impacts on outcome variables, effect sizes were computed for all comparisons using Hedge's  $g$  with an adjustment for small sample sizes (WWC, 2017a).

### **Impact Analysis Results**

This section describes the results of the HLM analysis. Full information on the HLM model results, including student- and school-level covariate parameters, can be found in Appendix C.

#### **Mathematics Results**

Students in the *i-Ready Instruction* schools were found to perform statistically significantly better ( $\alpha = .05$ ) than the students in the comparison schools on mathematics achievement as measured by student mathematics *i-Ready Diagnostic* scores for all six grades examined (Table 8).

The effect sizes, as measured by Hedge's  $g$ , for the six positive impacts ranged from 0.10 (grade 5) to 0.35 (kindergarten). The effect size for kindergarten and grade 1 were above 0.25

which are considered large for education interventions (Lipsey et al., 2012) and substantively important (WWC, 2018b). The effect sizes for grades 2 and 4 are moderate whereas those for grades 3 and 5 are small.

The intra-class correlations (ICCs) are also presented in Table 7. The ICCs measure the proportion of the variance that is between schools—that is, how much of the variance in mathematics *i-Ready Diagnostic* scores can be explained by school-level differences. The ICCs range from 0.12 (grades K and 3) to 0.18 (grade 4). This suggests the majority of variance in mathematics achievement is due to factors other than school-level differences.

### **Reading Results**

Students in the *i-Ready Instruction* schools were found to perform significantly better (using  $\alpha = .05$ ) than the comparison schools on reading achievement as measured by student reading *i-Ready Diagnostic* scores for grades K–2 (Table 9). There were no significant differences found at grades 3–5.

The effect sizes, as measured by Hedge's *g*, ranged from less than 0.01 (grade 3) to 0.34 (kindergarten). The effect size for kindergarten was above 0.25 which is considered large for education interventions (Lipsey et al., 2012) and substantively important (WWC, 2018b). The effect sizes for grade 1 was moderate (0.16), while those for grades 2 – 5 were small. For grades 2 – 5, only at grade 2 was there a statistically significant different.

The intra-class correlations (ICCs) are also presented in Table 8. The ICCs range from 0.10 (grades 3) to 0.13 (grades K and 5). This suggests the majority of variance in reading achievement is due to factors other than school-level differences.

### **Mathematics and Reading Student-Achievement Gains**

The adjusted mean differences presented in Tables 8 and 9 indicate students in the *i-Ready Instruction* (treatment) schools earned higher scores on the spring mathematics and reading *i-Ready Diagnostic* as compared to students in the comparison schools. Figures 1 and 2 graphically present the changes in student performance between the fall of 2017 and spring of 2018 for mathematics and reading, respectively. Though schools in the *i-Ready Instruction* groups had higher spring 2018 scores, schools in both groups saw gains in mathematics and reading achievement at all grade levels.

**Table 8. Impact Analysis Results for i-Ready Diagnostic Schools Compared to Comparison Schools for Mathematics Student Achievement at grades K–5.**

Grade	Group	ICC	Schools	Students	i-Ready Mean	i-Ready SD	Adj Mean Diff (SE)	p-value	Effect Size
K	<i>i-Ready Instruction</i>	0.12	60	1267	384.19	23.41	8.32 (1.35)	<0.001	0.35
	Comparison		30	1267	375.87	23.66			
1	<i>i-Ready Instruction</i>	0.14	61	3096	411.91	24.67	7.13 (1.05)	<0.001	0.29
	Comparison		59	3096	404.79	25.18			
2	<i>i-Ready Instruction</i>	0.14	61	3323	432.36	26.35	4.64 (1.05)	<0.001	0.18
	Comparison		60	3323	427.72	26.56			
3	<i>i-Ready Instruction</i>	0.12	61	3637	454.86	24.63	3.80 (1.08)	<0.001	0.14
	Comparison		61	3637	451.05	29.69			
4	<i>i-Ready Instruction</i>	0.18	61	3962	474.84	28.42	5.92 (1.04)	<0.001	0.20
	Comparison		61	3962	468.93	30.91			
5	<i>i-Ready Instruction</i>	0.15	61	3681	486.49	30.00	3.03 (0.99)	0.002	0.10
	Comparison		61	3681	483.46	30.27			

Notes: ICC = intraclass correlation, SD = standard deviation of i-Ready scores, Adj Mean Diff = adjusted mean difference between *i-Ready Instruction* and *Comparison* groups, SE = standard error of the adjusted mean difference, and Effect Size = Hedge's *g*.

**Table 9. Impact Analysis Results for i-Ready Diagnostic Schools Compared to Comparison Schools for Reading Student Achievement at grades K–5.**

Grade	Group	ICC	Schools	Students	<i>i-Ready</i> Mean	<i>i-Ready</i> SD	Adj Mean Diff (SE)	<i>p</i> -value	Effect Size
K	<i>i-Ready Instruction</i>	0.13	60	1267	406.80	35.95	11.74 (1.78)	<0.001	0.34
	Comparison		30	1267	395.06	33.89			
1	<i>i-Ready Instruction</i>	0.12	61	3096	458.83	44.12	7.17 (1.64)	<0.001	0.16
	Comparison		59	3096	451.66	45.80			
2	<i>i-Ready Instruction</i>	0.12	61	3323	499.59	45.94	2.93 (1.45)	0.043	0.06
	Comparison		60	3323	496.66	48.41			
3	<i>i-Ready Instruction</i>	0.10	61	3637	525.93	44.24	0.17 (1.14)	0.883	<0.01
	Comparison		61	3637	525.76	49.53			
4	<i>i-Ready Instruction</i>	0.12	61	3962	549.84	47.22	1.71 (1.29)	0.186	0.04
	Comparison		61	3962	548.13	48.87			
5	<i>i-Ready Instruction</i>	0.13	61	3681	572.95	49.55	2.47 (1.31)	0.060	0.05
	Comparison		61	3681	570.48	48.70			

Notes: ICC = intraclass correlation, SD = standard deviation of i-Ready scores, Adj Mean Diff = adjusted mean difference between *i-Ready Instruction* and *Comparison* groups, SE = standard error of the adjusted mean difference, and Effect Size = Hedge's *g*.

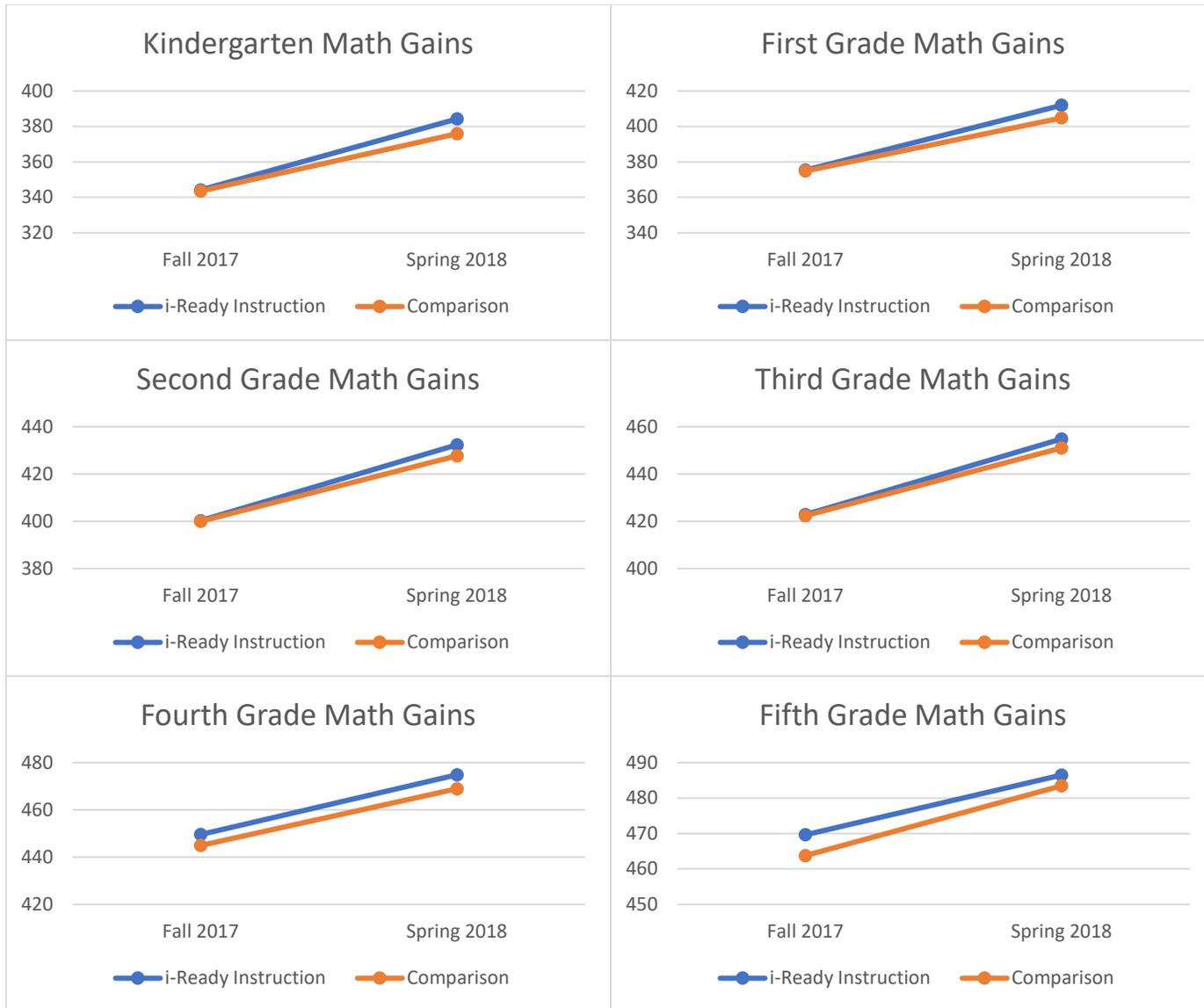


Figure 1. Changes in Mathematics Achievement Between Fall 2017 and Spring 2018 by Grade.

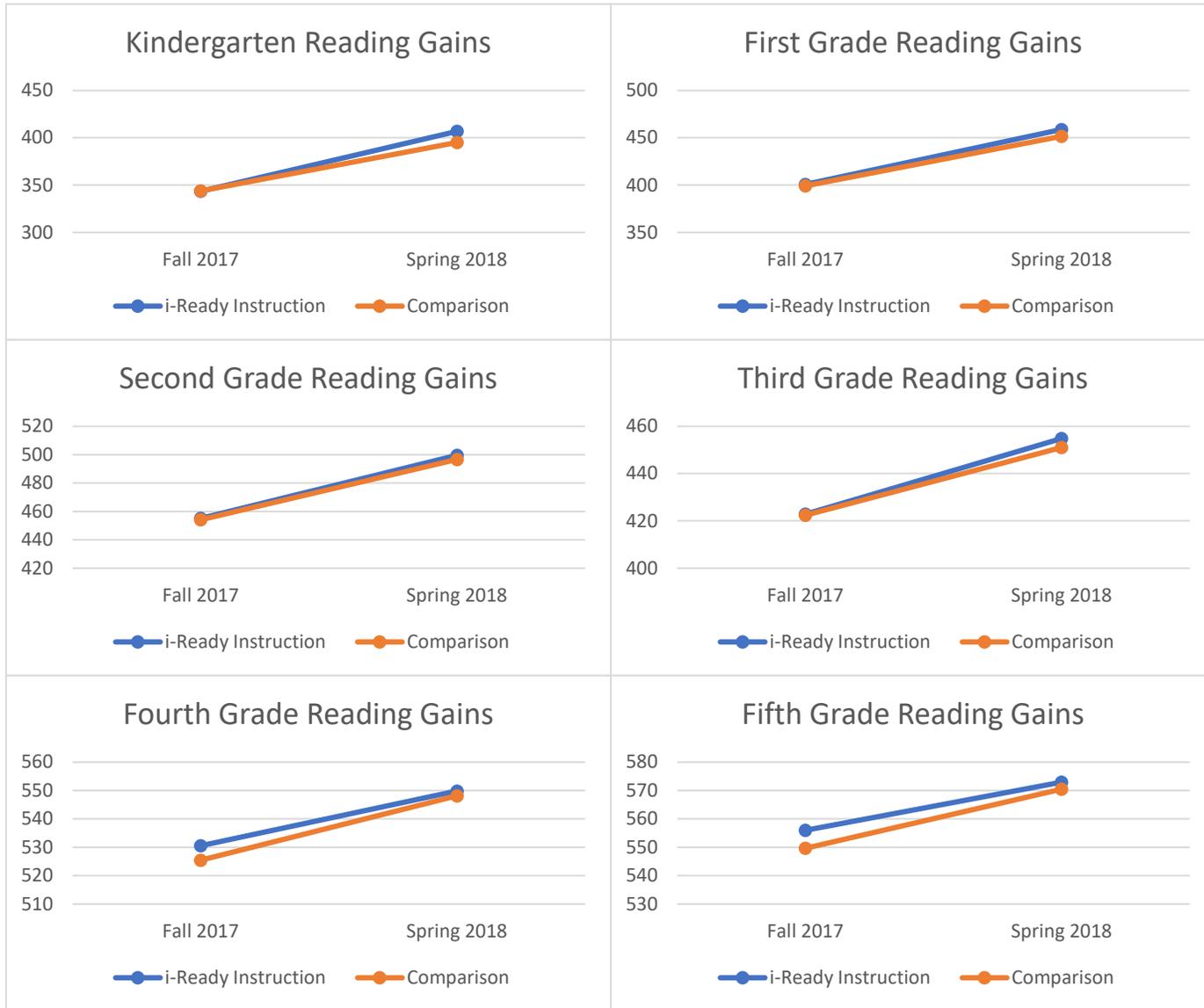


Figure 2. Changes in Reading Achievement Between Fall 2017 and Spring 2018 by Grade.

## Impact Summary and Discussion

Our findings suggest that implementation of *i-Ready Instruction* at the school and student level with fidelity resulted in higher student-level achievement in mathematics for all grades, and for reading at grades K–2, as measured by the *i-Ready Diagnostic*, compared to a matched group of schools and students. The mean mathematics achievement for the *i-Ready Instruction* group was statistically significantly higher in all six of the grades examined, and for grades K–2 for reading.

Moreover, the effect sizes showed additional support that students in *i-Ready Instruction* schools benefitted from their school's adoption and implementation of *i-Ready Instruction*, particularly in the early grades and for mathematics. For kindergarten mathematics and reading, and grade 1 mathematics, the effect sizes were 0.25 or higher. A standardized effect size of this magnitude is considered noteworthy in educational research (Lipsey et al., 2012; WWC, 2017b) where it is often difficult to observe large effects. In addition, moderate effect sizes (larger than 0.15) were found for mathematics at grades 2 and 4, and for reading at grade 1.

All schools attempt to provide a valuable education to their students by implementing quality curriculum and classroom assessments. Therefore, all students should expect to see gains in student achievement. This was evidenced when we examined gains in *i-Ready Diagnostic* scores between Fall of 2017 and Spring of 2018 for the *i-Ready Instruction* and comparison groups for mathematics and reading. It is important to note the effects of the *i-Ready Instruction* intervention were beyond the mean performance of schools that elected not to use it, more so for mathematics than reading.

There are various possible reasons why there was no statistically significant impact found at grades 3 - 5 for reading. One possibility is sampling error which is the chance that the sample does not show an effect when there is one in the population. A second possibility may be that it is more difficult to target reading instruction to the needs of kids compared to mathematics. A third possibility is that schools may have more solid reading curriculum and interventions in place for our *i-Ready Instruction* and comparison schools than mathematics. If other good things are happening in school for reading, it may be more difficult to detect the impact of *i-Ready Instruction*.

The study was conducted as a rigorous quasi-experimental design (QED) to meet the standards described in the WWC 4.0 standards to achieve a rating of *Meets WWC Group Design Standards with Reservations*. In addition, because we found statistically significant positive results for nine of the 12 grade and subject areas examined, this study meets the guidelines set forth by ESSA for a *Level 2 (or Moderate)* rating for evidence-based research (U.S. Department of Education, 2016).

This study was a QED in that schools in both groups were not participants in a research study but actual customers and everyday users of educational products, and we relied on implementation of *i-Ready Instruction* carried out in real-world conditions. Implementation of *i-Ready Instruction*, therefore, likely varied between schools. We may have found different results had the study been conducted under more controlled circumstances. Impacts are typically greater for studies that aim for ideal or close to ideal implementation and less for studies that examine real-world implementation. However, despite this limitation, we were able to find solid impacts for many of the grades and content areas examined.

Finally, our treatment group was compared to a comparison group using the *i-Ready Diagnostic*. It is possible that use of *i-Ready Diagnostic* itself increases student achievement; however, the

design of this study did not allow for an estimation of that impact. Further, use of the *i-Ready Diagnostic* only schools as a comparison group may have attenuated the effects of the treatment had that group been compared to a “business-as-usual” comparison group. Future studies might examine the impact of *i-Ready Instruction* use to a set of a comparison schools and students not implementing any Curriculum Associates products. This would require an external achievement measure, potentially state assessments, for use as an achievement baseline measure and outcome variable.

### **Quality Control Procedures**

We employed various quality control checks throughout the data cleaning, analysis, and reporting process. HumRRO, Curriculum Associates, and Century Analytics worked together to identify a rigorous methodology based on proper implementation of *i-Ready Instruction*, the WWC 4.0 standards, and ESSA Level 2 guidelines.

Rules for identifying treatment and comparison groups were determined upfront through collaboration between the three groups. Curriculum Associates provided information on the various components of *i-Ready Instruction*, and the frequency for which it should be used for solid implementation. They provided *i-Ready Diagnostic* and *Instruction* data to allow HumRRO and Century Analytics to empirically examine the extent to which these recommendations were followed by *i-Ready Instruction* schools. These discussions led to a treatment and comparison school criteria that all partners were confident in.

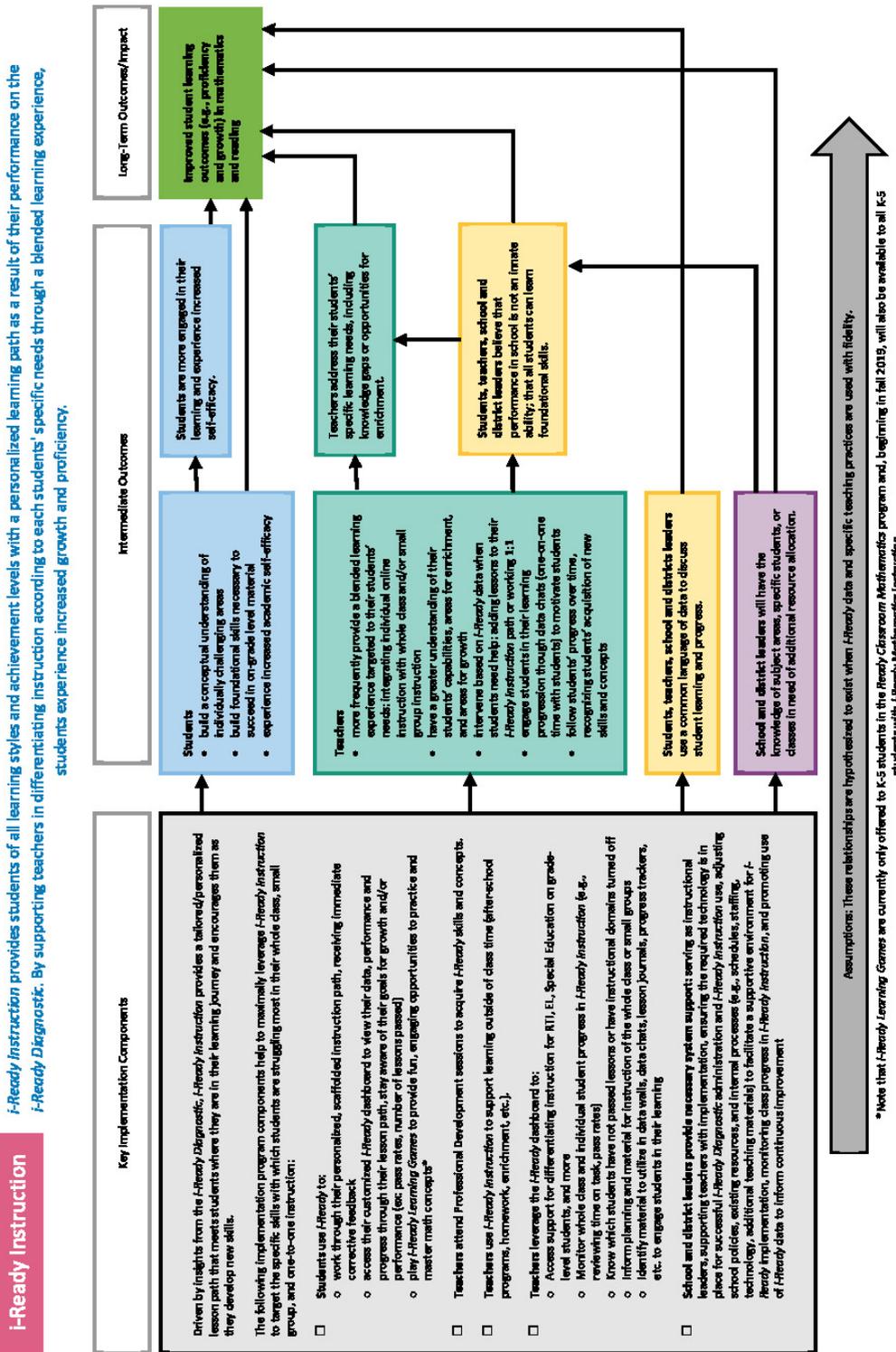
Data analysis work was completed collaboratively by HumRRO and Century Analytics. Century Analytics and HumRRO independently conducted matching and HLM analyses for each grade. The researchers reviewed results against each other and worked out any discrepancies. All data reported in this study were verified by both researchers.

## References

- Curriculum Associates. (2018). *i-Ready Spring 2019: Leadership Success Guide*. Billerica, MA: Curriculum Associates, LLC. Retrieved from file:///C:/Users/rdvorak/Downloads/iready-pd-success-guide-for-leaders-2019-2.pdf
- Curriculum Associates. (2019, June 20). *Getting the MOst from Online Instruction*. Retrieved from i-Ready Central: <http://i-readycentral.com/articles/getting-the-most-from-online-instruction/>
- Curriculum Associates. (2019, June 27). *The Research Behind our Programs*. Retrieved from <https://www.curriculumassociates.com/research-and-efficacy>
- Curriculum Associates. (2019a, June 27). *The Research Behind our Programs*. Retrieved from <https://www.curriculumassociates.com/research-and-efficacy>
- Curriculum Associates. (2019b, July 22). *Get Good Data*. Retrieved from i-Ready Central: <http://i-readycentral.com/articles/get-good-data/>
- Guo, S. and Fraser, M.W. (2010). *Propensity Score Analysis*. Thousand Oaks: SAGE.
- Lipsey, M.W., Puzio, K., Yun, C., Hebert, M.A., Steinka-Fry, K., Cole, M.W., Roberts, M., Anthony, K.S., Busick, M.D. (2012). *Translating the Statistical Representation of the Effects of Education Interventions into More Readily Interpretable Forms*. (NCSER 2013-3000). Washington, DC: National Center for Special Education Research, Institute of Education Sciences, U.S. Department of Education. This report is available on the IES website at <http://ies.ed.gov/ncser/>.
- Spybrook, S., Bloom, H., Congdon, R., Hill, C., Liu, X., Martinez, A., & Raudenbush, S. (2011) *Optimal Design Plus Empirical Evidence* (Version 3.01): W. T. Grant Foundation.
- Stuart, E. A. (2010). Matching methods for causal inference: A review and a look forward. *Statistical Science*, 25, 1–21.
- U.S. Department of Education, Office of Elementary and Secondary Education. (2016, September). *Non-Regulatory Guidance: Using Evidence to Strengthen Education Investments*. Retrieved from <https://www2.ed.gov/policy/elsec/leg/essa/guidanceuseseseinvestment.pdf>.
- What Works Clearinghouse, Institute of Education Sciences, U.S. Department of Education (2014). *Review Protocol for Beginning Reading Interventions* (Version 3.0). Retrieved from <http://whatworks.ed.gov>.
- What Works Clearinghouse, Institute of Education Sciences, U.S. Department of Education (2015). *Review Protocol for Primary Mathematics* (Version 3.1). Retrieved from <http://whatworks.ed.gov>.
- What Works Clearinghouse, Institute of Education Sciences, U.S. Department of Education. (2017a, October). *What Works Clearinghouse: Procedures Handbook* (Version 4.0). Retrieved from <http://whatworks.ed.gov>.

What Works Clearinghouse, Institute of Education Sciences, U.S. Department of Education.  
(2017b, October). *What Works Clearinghouse: Standards Handbook (Version 4.0)*.  
Retrieved from <http://whatworks.ed.gov>.

## Appendix A. *i-Ready Instruction Theory of Action*



## Appendix B.

# i-Ready® Diagnostic Reliability Information from Curriculum Associates’ Technical Report

i-Ready Assessments Technical Manual

March 2018

### Chapter 4. Reliability

Test reliability generally refers to the precision with which measurements are made (Haertel, 2006). All psychometric models rely on the notion of an underlying “true score” for each student which is measured imperfectly (i.e., with error) by any test. Analyses of test reliability endeavor to quantify the relative magnitude of true score and error variance that exist in observed test scores. The greater the proportion of true score variance relative to error variance in the observed variance of test scores, the more reliable the test. This proportional relationship exposes the indirect relationship between test reliability and measurement error: more reliable tests tend to give rise to scores with less measurement error. Because one cannot directly observe either the true score or error, a variety of methods are employed to predict measurement error and analyze test reliability. This chapter presents the reliability estimates for the Diagnostic and Growth Monitoring.

#### 4.1. i-Ready Diagnostic

Three ways of characterizing test reliability for the i-Ready Diagnostic are included. First, because the i-Ready Diagnostic is an IRT-based CAT assessment, a student’s standard error of measurement (SEM) may be calculated directly from students’ item responses and the item difficulty parameters. In addition, the SEM varies with students’ scores, so the conditional standard error of measurement (CSEM) is also presented. More reliable tests will give rise to lower SEMs.

Second, because the i-Ready Diagnostic adapts to students’ item responses and delivers items that are targeted to their exhibited proficiency, a slightly different approach to estimating reliability is called for than for a fixed-form assessment. Classical test theory statistics like coefficient  $\alpha$  assume uniform error variance across the scale. Under the IRT model on which the i-Ready Diagnostic is based, the standard error of measurement has a well-known and variable relationship with students’ scale scores. As such, a more appropriate method for estimating reliability for a CAT assessment like the i-Ready Diagnostic is to calculate marginal reliability (Sireci, Thissen, & Wainer, 1991). Marginal reliability is more closely conceptually related to the definition of reliability based on the standard error of estimation given by Lord and Novick’s (1968) Expression 3.8.4 than to internal consistency measures like coefficient  $\alpha$  (Sireci, Thissen, & Wainer, 1991). In short, measurement error is calculated for each student and then the marginal error variance is calculated across all scale score levels, thus appropriately accounting for the fact that scale scores toward the middle of the scale tend to be more precise than at either extreme. More reliable tests will have greater marginal reliability estimates.

Third, test-retest reliability analyses are presented, in which students test twice and the strength of the linear relationship between their scores is yet another lens through which to view reliability. More reliable test scores are strongly, linearly related, because they contain relatively little error variance (when compared with true score variance).

##### 4.1.1. Standard Error of Measurement (SEM)

The Standard Error of Measurement (SEM) is a measure of the degree of precision of students’ i-Ready Diagnostic scores. SEMs are affected by factors such as how well the data fit the underlying model, student response consistency, student location on the ability continuum, match of items to student ability, and test length. In the context of i-Ready assessments, a high SEM could be caused by students performing erratically or having extreme response vectors (e.g., getting all items correct or incorrect). Although there are no specific targets for observed standard errors, lower values of standard errors are preferable to higher values because they suggest reliable student ability measures. Given the adaptive nature of i-Ready and the wide difficulty range in the item bank, standard errors are expected to approach the theoretical minimum for the test of the given length. The theoretical minimum would be reached if each item difficulty value matched the student’s ability estimate perfectly. Theoretical minimums are restricted by the number of items served in the assessment—the more items that are served up, the lower the SEM could potentially be.

The theoretical minimum SEM for the scale score is given by the following equation:

$$\text{Min}(SEM) = \frac{b}{\sqrt{n \times 0.25}} \tag{5}$$

where  $b$  is the scaling constant which is the slope for scale transformation from  $\theta$  to the i-Ready scale (more detail about scale scores is provided in section 7.1.1),  $n$  is the number of items, 0.25 is the expected maximum information (hence minimum error) under the Rasch model (when the probability of a correct and of an incorrect response are both equal to 0.5). For the overall score, a typical assessment consists of 72 items. The number of items within a domain varies between 12 and 36 depending on the test flow (see Appendices D and E for more detail about the test flow). The theoretical minimum SEM for overall scores is 8.9 and 6.0 scale

score points for Reading and Mathematics, respectively. The theoretical minimum SEMs for domain scores range from 17.8 to 21.8 for Reading (12 to 18 items per domain) and range from 8.5 to 13.6 for Mathematics (14 to 36 items per domain).

Table 4.1 shows the mean and standard deviation of the overall score SEM across all Diagnostic assessments taken during August and September of the 2016–2017 school year<sup>7</sup>. The mean SEMs for overall scores are low in both the Reading (9.3–10.9) and Mathematics assessments (6.3–6.5), with many approaching the theoretical minimum SEM.

**Table 4.1. August and September of 2016–2017 School Year: Diagnostic Mean SEMs for Overall Scores and Sample Size by Grade and Subject**

		Reading											
Statistic	K	1	2	3	4	5	6	7	8	9	10	11	12
Mean	9.3	9.3	10.4	10.0	10.1	10.3	10.5	10.6	10.7	10.7	10.8	10.9	10.8
N*	184.3	287.6	323.3	366.3	346.0	345.2	251.3	225.9	223.6	33.9	22.3	14.6	7.4
		Mathematics											
Statistic	K	1	2	3	4	5	6	7	8	9	10	11	12
Mean	6.5	6.4	6.4	6.4	6.4	6.4	6.4	6.4	6.4	6.3	6.3	6.3	6.4
N*	191.2	298.5	334.2	376.1	366.0	366.1	276.3	254.2	238.8	39.5	25.4	16.4	8.5

\*: Sample size in thousands of students.

Table 4.2 shows the mean SEMs by domain in the Diagnostic tests. Each domain is assessed by 12–36 items (see Appendices B & C for details about the number of items per domain). The observed mean SEMs are also close to the respective minimal value given the length of the domain.

**Table 4.2. August and September of 2016–2017 School Year: Diagnostic Mean SEMs for Domain Scores by Grade**

Reading Domain	K	1	2	3	4	5	6	7	8	9	10	11	12
Comprehension: Informational Text	24.3	24.1	24.1	19.2	19.4	19.5	19.6	19.7	19.8	19.9	20.0	20.2	20.3
Comprehension: Literature	24.2	24.2	24.5	19.4	19.5	19.6	19.7	19.8	19.9	20.1	20.2	20.4	20.4
High-Frequency Words	24.4	26.9	36.9	28.3	29.0	29.1	28.8	28.5	28.5	33.4	34.1	34.5	30.4
Phonological Awareness	24.1	24.6	24.4	24.8	24.7	24.4	24.9	24.8	24.3	25.3	24.7	25.1	23.8
Phonics	23.6	23.6	23.5	24.5	24.6	24.8	24.8	24.9	25.0	25.0	25.2	25.1	25.3
Vocabulary	25.6	24.0	23.9	18.8	18.8	18.7	18.7	18.8	18.8	18.9	19.0	19.1	19.4
Mathematics Domain	K	1	2	3	4	5	6	7	8	9	10	11	12
Algebra and Algebraic Thinking	12.9	12.6	12.6	12.5	12.6	12.6	12.6	12.6	12.6	10.3	10.2	10.1	10.6
Geometry	14.8	14.9	14.7	14.5	14.5	14.5	14.5	14.5	14.5	11.0	10.9	10.5	11.0
Measurement and Data	14.8	14.7	14.5	14.6	14.6	14.6	14.6	14.6	14.6	14.8	14.8	14.8	14.9
Number and Operations	11.9	11.9	11.9	12.0	11.9	11.9	11.9	11.9	11.9	12.1	12.1	12.1	12.2

Table 4.3 shows the mean overall score SEMs across Diagnostic assessments taken by three special groups of students for August and September of the 2016–2017 school year: Economically Disadvantaged, English Learner (EL), and Special Education students. The mean SEMs for these special groups are low and very similar to the mean SEMs for the entire i-Ready population presented in Table 4.1.

<sup>7</sup> For SEM and CSEM based on the 2015–2016 data based on the previous scale, please refer to Appendix M.

**Table 4.3. August and September of 2016–2017 School Year: Diagnostic Mean SEMs for Overall Scores and Sample Size by Grade and Subject for Special Groups**

		Reading												
Group	N	K	1	2	3	4	5	6	7	8	9	10	11	12
Economically Disadvantaged	258.7	9.3	9.3	10.2	9.9	10.0	10.2	10.4	10.5	10.6	10.7	10.8	10.8	10.7
EL	237.9	9.3	9.3	10.1	9.7	9.7	9.9	10.0	10.2	10.3	10.6	10.6	10.7	10.6
Special Ed.	158.4	9.3	9.4	10.0	9.7	9.8	10.0	10.0	10.2	10.3	10.4	10.5	10.6	10.6
		Mathematics												
Group	N	K	1	2	3	4	5	6	7	8	9	10	11	12
Economically Disadvantaged	333.7	6.5	6.4	6.4	6.4	6.4	6.4	6.4	6.4	6.4	6.3	6.3	6.4	6.4
EL	255.2	6.5	6.4	6.4	6.4	6.4	6.4	6.4	6.5	6.5	6.4	6.4	6.4	6.4
Special Ed.	172.7	6.5	6.5	6.5	6.4	6.5	6.4	6.5	6.5	6.5	6.5	6.5	6.5	6.6

4.1.2. Conditional Standard Error of Measurement

In addition to the mean SEMs for both overall and domain scores, Figures 9 and 10 present the observed mean conditional SEMs by overall score across the full spectrum of the i-Ready scale based on Diagnostic assessments administered in August and September of the 2016–2017 school year. The middle 98% range of the observed scores is indicated for both subjects. Except at the very tail ends of the scale, more than 98% of the score ranges have conditional SEMs around 10 (logit value of 0.26) for Reading, and lower than 7.5 (logit value of 0.30) for Mathematics. The results shown in these figures demonstrate that the measurement error of the i-Ready Diagnostic is consistent and at a desirable (i.e., low) level across a wide score range. Considering these values relative to the theoretical SEM minima of 8.9 and 6.0 scale score points for Reading and Mathematics, respectively, reveals that the i-Ready Diagnostic scale scores exhibit very low measurement error and therefore a high degree of reliability.

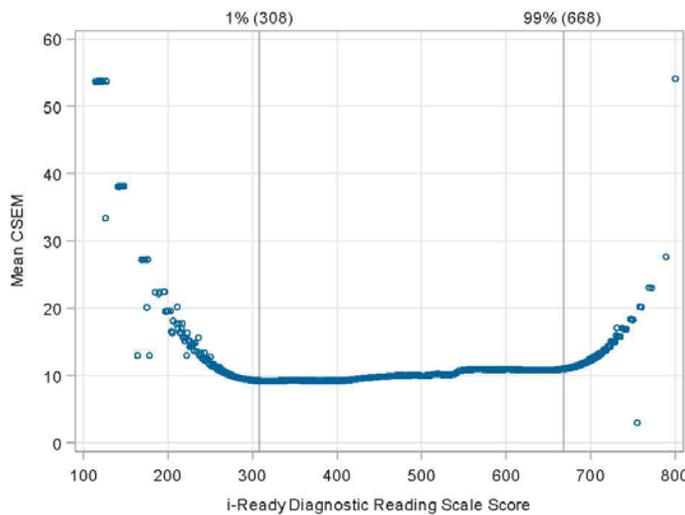


Figure 9. Conditional SEM of i-Ready Diagnostic Assessments for Reading

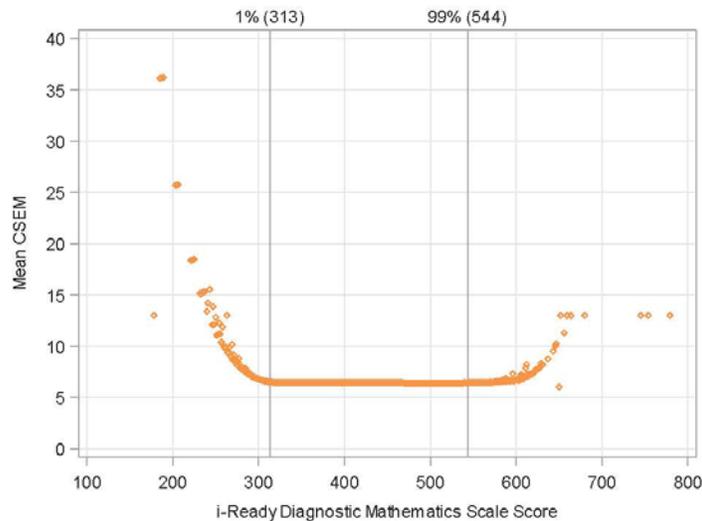


Figure 10. Conditional SEM of i-Ready Diagnostic Assessments for Mathematics

#### 4.1.3. Marginal Reliability Estimates

The IRT analogue to classical reliability estimation is called marginal reliability and operates on the variance of the theta scores and the mean of the expected error variance (Samejima, 1977; Sireci, Thissen, & Wainer, 1991). This marginal reliability uses the classical definition of reliability as proportion of variance in the total observed score due to true score. The true score variance is computed as the observed score variance minus the error variance (see equation below).

$$\rho_{\theta} = \frac{\sigma_{\theta}^2 - \sigma_E^2}{\sigma_{\theta}^2} \quad (6)$$

where  $\rho_{\theta}$  is the marginal reliability estimate,  $\sigma_{\theta}^2$  is the observed error variance of the ability estimate, and  $\sigma_E^2$  is the observed mean conditional error variance. Like a classical reliability coefficient, the marginal reliability estimate increases as the standard error of measurement decreases; it approaches 1 when the standard error of measurement approaches 0.

Other relevant measures provided by WINSTEPS (Linacre, 2012; Linacre, 2002) in Rasch analysis are separation indices and reliability estimates. Separation indices reflect the ratio of person (or item) standard deviation to the standard deviation of error and are closely related to reliability estimates (Wright, 1996). Values in excess of 2.0 indicate that greater than 80% of the variance in scores is not due to error, but rather to person or item differences. For the more common reliability measures, person reliability (or marginal reliability) is equivalent to the commonly recognized test score reliability in classical test theory settings such as the KR-20 internal consistency reliability coefficient. Further, Rasch analysis provides item reliability or the ratio of true item variance to observed item variance, which has no direct counterpart in classical test analysis. The higher item reliability values indicate greater consistency of item ordering.

Based on the spring 2014 calibration, the estimated reliability for the Reading test was 0.97 with a separation index of 5.43 and the estimated reliability of the Mathematics test was 0.96 with a separation index of 5.22. In addition, data from August and September 2016 were used to estimate marginal reliability for each domain score by grade (Table 4.5) and for overall scores by grade (Table 4.4). Table 4.5 shows that—apart from Grade 2 High-Frequency Words—reliability estimates by domain and grade vary between 0.63 and 0.96. The marginal reliability estimates for some domains were lower than others (e.g., High-Frequency Words and Phonological Awareness), likely due to the shorter test length as well as students' grade restriction related to the content for these

domains. Since the overall score is based on all domains and hence a greater number of items, Table 4.4 shows that the overall scores tend to be much more reliable than any individual component domain, with marginal reliability estimates of between 0.91 and 0.99.

**Table 4.4. August and September of 2016–2017 School Year: Marginal Reliability Estimates for Overall Scores by Grade**

Statistic	Reading												
	K	1	2	3	4	5	6	7	8	9	10	11	12
Marginal Reliability	0.91	0.95	0.96	0.97	0.97	0.97	0.97	0.97	0.97	0.98	0.98	0.98	0.98
N*	184.3	287.6	323.3	366.3	346.0	345.2	251.3	225.9	223.6	33.9	22.3	14.6	7.4
Statistic	Mathematics												
	K	1	2	3	4	5	6	7	8	9	10	11	12
Marginal Reliability	0.92	0.94	0.94	0.95	0.96	0.96	0.96	0.97	0.97	0.98	0.98	0.98	0.99
N*	191.2	298.5	334.1	376.0	365.7	365.9	276.2	254.0	238.6	39.5	25.4	16.4	8.5

\*: Sample size in thousands of students.

**Table 4.5. August and September of 2016–2017 School Year: Marginal Reliability Estimates for Domain Scores by Grade**

Reading Domain	K	1	2	3	4	5	6	7	8	9	10	11	12
Comprehension: Informational Text	0.70	0.79	0.85	0.90	0.91	0.91	0.91	0.92	0.92	0.93	0.94	0.94	0.95
Comprehension: Literature	0.68	0.77	0.84	0.89	0.90	0.90	0.91	0.91	0.91	0.92	0.93	0.93	0.95
High-Frequency Words	0.64	0.76	0.57	0.75	0.76	0.76	0.76	0.76	0.75	0.78	0.74	0.81	0.80
Phonological Awareness	0.66	0.76	0.73	0.78	0.79	0.77	0.83	0.81	0.82	0.86	0.81	0.83	0.80
Phonics	0.71	0.80	0.84	0.79	0.80	0.79	0.79	0.79	0.80	0.81	0.82	0.83	0.87
Vocabulary	0.63	0.80	0.82	0.88	0.88	0.89	0.89	0.90	0.90	0.91	0.92	0.93	0.95
Mathematics Domain	K	1	2	3	4	5	6	7	8	9	10	11	12
Algebra and Algebraic Thinking	0.75	0.81	0.82	0.84	0.85	0.86	0.87	0.88	0.89	0.92	0.93	0.94	0.96
Geometry	0.78	0.79	0.81	0.81	0.83	0.85	0.86	0.87	0.88	0.92	0.93	0.94	0.95
Measurement and Data	0.74	0.79	0.81	0.83	0.84	0.86	0.87	0.87	0.88	0.86	0.87	0.87	0.91
Number and Operations	0.76	0.80	0.81	0.82	0.85	0.87	0.89	0.90	0.91	0.84	0.84	0.85	0.89

#### 4.1.4. Test-retest Reliability Estimates

Evidence of test-retest reliability was assessed based on a subsample of students who, during the 2016–2017 school year, took the i-Ready Diagnostic once in the fall and once in the winter testing windows with the recommended 12–18 weeks between tests and rushed in neither test. The mean testing interval was 106 days (15.2 weeks), and over that period students would have received classroom instruction and hence have been anticipated to have grown. The fact that the test administrations were more than a negligible amount of time apart suggests that not only was there variability in the test scores due to error variance, but also that students' true scores may have changed and therefore introduced additional variance that is not attributable to the precision of the i-Ready Diagnostic test itself. Table 4.6 presents the correlations between the two overall Diagnostic scores. For all grades and both subjects, test-retest correlations are above 0.70, which, while considered strong, tends to be lower than the marginal reliability estimates, primarily because of the likelihood that students' content knowledge did grow over the 12–18 weeks that passed between their two test occasions.

**Table 4.6. Test-Retest Correlations by Grade and Subject**

		Reading											
Statistic	K	1	2	3	4	5	6	7	8	9	10	11	12
Test-Retest Reliability	0.70	0.82	0.85	0.85	0.86	0.86	0.86	0.85	0.85	0.86	0.86	0.85	0.88
N*	118.1	159.1	173.3	199.5	193.5	190.3	137.1	119.9	114.4	12.5	8.2	4.9	2.0
		Mathematics											
Statistic	K	1	2	3	4	5	6	7	8	9	10	11	12
Test-Retest Reliability	0.71	0.77	0.81	0.82	0.85	0.86	0.87	0.87	0.87	0.85	0.85	0.86	0.89
N*	113.4	161.9	184.4	210.8	210.2	206.9	153.9	134.1	123.3	13.0	7.0	3.7	1.8

\*: Sample size in thousands of students.

#### 4.2. Standard Error of Measurement for Growth Monitoring

Table 4.7 shows the mean SEMs for Growth Monitoring (GM) assessments taken during August and September of the 2016–2017 school year<sup>8</sup>. Because the Growth Monitoring tests are much shorter (19–21 items), the theoretical minimum SEM is between 16.5 and 17.3 points for Reading and 11.1 and 11.7 points for Mathematics. The observed SEMs are close to the lowest obtainable value.

**Table 4.7. August and September of 2016–2017 School Year: Growth Monitoring SEM and Sample Size by Grade**

		Reading							
Statistic	K	1	2	3	4	5	6	7	8
Mean	17.7	17.7	17.7	17.8	17.8	17.8	17.9	17.9	17.9
N*	36.5	63.0	70.0	78.5	69.0	65.2	29.2	20.3	17.4
		Mathematics							
Statistic	K	1	2	3	4	5	6	7	8
Mean	11.8	11.8	11.8	11.8	11.8	11.8	11.8	11.8	11.9
N*	26.3	46.9	54.4	65.3	62.6	60.1	35.3	26.8	22.5

\*: Sample size in thousands of students.

<sup>8</sup> For SEM and CSEM based on the 2015–2016 data based on the previous scale, please refer to Appendix M.

## Appendix C. Impact HLM Coefficients

### Mathematics

**Table C.1. HLM Results for Kindergarten Mathematics.**

Covariates	Coef.	SE	z	p-value	95% Conf. Interval	
<b>Student-Level Covariates</b>						
Treatment Group Membership	8.32	1.35	6.16	<0.001	5.67	10.96
Fall 2017 Reading i-Ready School Mean Centered	-0.04	0.10	-0.40	0.686	-0.24	0.16
Fall 2017 Mathematics i-Ready School Mean Centered	0.64	0.13	4.94	<0.001	0.39	0.90
<b>School-Level Covariates</b>						
Percent LEP	-0.17	7.51	-0.02	0.982	-14.89	14.55
Percent SWD	1.66	16.65	0.10	0.920	-30.98	34.31
Percent White	0.06	0.04	1.43	0.153	-0.02	0.14
Percent FRL	-1.36	3.77	-0.36	0.719	-8.75	6.04
Fall Reading i-Ready School-Level Grand Mean Centered	0.20	0.02	11.84	<0.001	0.16	0.23
Fall Mathematics i-Ready School-Level Grand Mean Centered	0.53	0.02	23.33	<0.001	0.48	0.57
Intercept	372.70	4.56	81.72	<0.001	363.77	381.64

*Note.* LEP = limited English proficient; SWD = students with disabilities; FRL = free or reduced lunch; Coef. = coefficient; SE = standard error of the coefficient; z = standardized score.

**Table C.2. HLM Results for First Grade Mathematics.**

Covariates	Coef.	SE	z	p-value	95% Conf. Interval	
<b>Student-Level Covariates</b>						
Treatment Group Membership	7.13	1.05	6.77	<0.001	5.06	9.19
Fall 2017 Reading i-Ready School Mean Centered	0.13	0.09	1.34	0.179	-0.06	0.31
Fall 2017 Mathematics i-Ready School Mean Centered	0.45	0.15	2.94	0.003	0.15	0.76
<b>School-Level Covariates</b>						
Percent LEP	11.68	6.34	1.84	0.065	-0.75	24.11
Percent SWD	27.64	13.51	2.05	0.041	1.16	54.12
Percent White	0.10	0.04	2.72	0.006	0.03	0.18
Percent FRL	-9.35	3.02	-3.10	0.002	-15.26	-3.43
Fall Reading i-Ready School-Level Grand Mean Centered	0.15	0.01	19.72	<0.001	0.14	0.17
Fall Mathematics i-Ready School-Level Grand Mean Centered	0.61	0.01	46.03	<0.001	0.59	0.64
Intercept	399.14	3.66	109.14	<0.001	391.97	406.31

**Table C.3. HLM Results for Second Grade Mathematics.**

Covariates	Coef.	SE	z	p-value	95% Conf. Interval	
<b>Student-Level Covariates</b>						
Treatment Group Membership	4.64	1.05	4.44	<0.001	2.59	6.69
Fall 2017 Reading i-Ready School Mean Centered	0.12	0.07	1.74	0.083	-0.02	0.25
Fall 2017 Mathematics i-Ready School Mean Centered	0.61	0.14	4.29	<0.001	0.33	0.88
<b>School-Level Covariates</b>						
Percent LEP	-13.10	6.24	-2.10	0.036	-25.33	-0.87
Percent SWD	25.77	13.35	1.93	0.054	-0.40	51.94
Percent White	-0.02	0.04	-0.66	0.507	-0.10	0.05
Percent FRL	-5.56	3.38	-1.65	0.100	-12.19	1.06
Fall Reading i-Ready School-Level Grand Mean Centered	0.12	0.01	19.55	<0.001	0.11	0.13
Fall Mathematics i-Ready School-Level Grand Mean Centered	0.67	0.01	55.73	<0.001	0.65	0.70
Intercept	431.07	3.79	113.71	<0.001	423.64	438.50

*Note.* LEP = limited English proficient; SWD = students with disabilities; FRL = free or reduced lunch; Coef. = coefficient; SE = standard error of the coefficient; z = standardized score.

**Table C.4. HLM Results for Third Grade Mathematics.**

Covariates	Coef.	SE	z	p-value	95% Conf. Interval	
<b>Student-Level Covariates</b>						
Treatment Group Membership	3.80	1.08	3.52	<0.001	1.68	5.92
Fall 2017 Reading i-Ready School Mean Centered	0.15	0.07	2.11	0.035	0.01	0.28
Fall 2017 Mathematics i-Ready School Mean Centered	0.72	0.14	5.09	<0.001	0.44	1.00
<b>School-Level Covariates</b>						
Percent LEP	-5.48	6.21	-0.88	0.378	-17.65	6.69
Percent SWD	4.24	13.14	0.32	0.747	-21.53	30.00
Percent White	-0.05	0.04	-1.30	0.193	-0.12	0.02
Percent FRL	1.44	3.31	0.43	0.664	-5.05	7.94
Fall Reading i-Ready School-Level Grand Mean Centered	0.11	0.01	19.15	<0.001	0.10	0.12
Fall Mathematics i-Ready School-Level Grand Mean Centered	0.73	0.01	63.30	<0.001	0.71	0.76
Intercept	453.57	3.46	131.16	<0.001	446.79	460.35

**Table C.5. HLM Results for Fourth Grade Mathematics.**

Covariates	Coef.	SE	z	p-value	95% Conf. Interval	
<b>Student-Level Covariates</b>						
Treatment Group Membership	5.92	1.04	5.69	<0.001	3.88	7.96
Fall 2017 Reading i-Ready School Mean Centered	0.13	0.08	1.75	0.080	-0.02	0.28
Fall 2017 Mathematics i-Ready School Mean Centered	0.86	0.12	6.94	<0.001	0.61	1.10
<b>School-Level Covariates</b>						
Percent LEP	2.24	6.07	0.37	0.712	-9.66	14.13
Percent SWD	-1.76	11.83	-0.15	0.882	-24.95	21.44
Percent White	0.02	0.03	0.70	0.485	-0.04	0.09
Percent FRL	1.25	2.99	0.42	0.676	-4.61	7.12
Fall Reading i-Ready School-Level Grand Mean Centered	0.10	0.01	17.49	<0.001	0.09	0.11
Fall Mathematics i-Ready School-Level Grand Mean Centered	0.78	0.01	72.42	<0.001	0.76	0.81
Intercept	469.40	3.29	142.68	<0.001	462.95	475.85

Note. LEP = limited English proficient; SWD = students with disabilities; FRL = free or reduced lunch; Coef. = coefficient; SE = standard error of the coefficient; z = standardized score.

**Table C.6. HLM Results for Fifth Grade Mathematics.**

Covariates	Coef.	SE	z	p-value	95% Conf. Interval	
<b>Student-Level Covariates</b>						
Treatment Group Membership	3.03	0.99	3.06	0.002	1.09	4.97
Fall 2017 Reading i-Ready School Mean Centered	0.05	0.05	0.91	0.361	-0.05	0.15
Fall 2017 Mathematics i-Ready School Mean Centered	0.85	0.09	9.26	<0.001	0.67	1.03
<b>School-Level Covariates</b>						
Percent LEP	-9.19	5.37	-1.71	0.087	-19.71	1.33
Percent SWD	0.95	11.18	0.08	0.932	-20.97	22.87
Percent White	0.01	0.03	0.29	0.770	-0.05	0.07
Percent FRL	-0.46	2.78	-0.16	0.870	-5.91	5.00
Fall Reading i-Ready School-Level Grand Mean Centered	0.10	0.01	17.27	<0.001	0.09	0.11
Fall Mathematics i-Ready School-Level Grand Mean Centered	0.80	0.01	79.00	<0.001	0.78	0.82
Intercept	487.45	2.93	166.10	<0.001	481.69	493.20

*Note.* LEP = limited English proficient; SWD = students with disabilities; FRL = free or reduced lunch; Coef. = coefficient; SE = standard error of the coefficient; z = standardized score.

## Reading

**Table C.7. HLM Results for Kindergarten Reading.**

Covariates	Coef.	SE	z	p-value	95% Conf. Interval	
<b>Student-Level Covariates</b>						
Treatment Group Membership	11.74	1.78	6.59	<0.001	8.25	15.24
Fall 2017 Reading i-Ready School Mean Centered	0.42	0.14	3.02	0.003	0.15	0.69
Fall 2017 Mathematics i-Ready School Mean Centered	0.61	0.18	3.48	<0.001	0.27	0.95
<b>School-Level Covariates</b>						
Percent LEP	-10.83	10.03	-1.08	0.280	-30.48	8.82
Percent SWD	3.59	22.09	0.16	0.871	-39.71	46.88
Percent White	0.00	0.05	-0.04	0.967	-0.11	0.11
Percent FRL	-0.98	5.02	-0.19	0.846	-10.81	8.85
Fall Reading i-Ready School-Level Grand Mean Centered	0.47	0.02	19.38	<0.001	0.42	0.52
Fall Mathematics i-Ready School-Level Grand Mean Centered	0.57	0.03	17.16	<0.001	0.50	0.63
Intercept	395.93	6.07	65.24	<0.001	384.04	407.83

*Note.* LEP = limited English proficient; SWD = students with disabilities; FRL = free or reduced lunch; Coef. = coefficient; SE = standard error of the coefficient; z = standardized score.

**Table C.8. HLM Results for First Grade Reading.**

Covariates	Coef.	SE	z	p-value	95% Conf. Interval	
<b>Student-Level Covariates</b>						
Treatment Group Membership	7.17	1.64	4.39	<0.001	3.97	10.38
Fall 2017 Reading i-Ready School Mean Centered	0.70	0.15	4.81	<0.001	0.41	0.98
Fall 2017 Mathematics i-Ready School Mean Centered	0.18	0.24	0.74	0.457	-0.29	0.65
<b>School-Level Covariates</b>						
Percent LEP	10.29	9.86	1.04	0.297	-9.03	29.61
Percent SWD	25.49	20.96	1.22	0.224	-15.59	66.57
Percent White	0.09	0.06	1.52	0.130	-0.03	0.20
Percent FRL	-16.93	4.70	-3.60	<0.001	-26.13	-7.72
Fall Reading i-Ready School-Level Grand Mean Centered	0.69	0.01	50.67	<0.001	0.66	0.71
Fall Mathematics i-Ready School-Level Grand Mean Centered	0.45	0.02	19.43	<0.001	0.41	0.50
Intercept	452.06	5.68	79.61	<0.001	440.93	463.19

**Table C.9. HLM Results for Second Grade Reading.**

Covariates	Coef.	SE	z	p-value	95% Conf. Interval	
<b>Student-Level Covariates</b>						
Treatment Group Membership	2.93	1.45	2.02	0.043	0.09	5.78
Fall 2017 Reading i-Ready School Mean Centered	0.68	0.10	7.19	<0.001	0.50	0.87
Fall 2017 Mathematics i-Ready School Mean Centered	0.11	0.20	0.58	0.561	-0.27	0.50
<b>School-Level Covariates</b>						
Percent LEP	-18.44	8.69	-2.12	0.034	-35.46	-1.41
Percent SWD	22.40	18.53	1.21	0.227	-13.91	58.72
Percent White	-0.04	0.05	-0.70	0.483	-0.13	0.06
Percent FRL	-13.78	4.71	-2.93	0.003	-23.01	-4.56
Fall Reading i-Ready School-Level Grand Mean Centered	0.65	0.01	65.93	<0.001	0.63	0.67
Fall Mathematics i-Ready School-Level Grand Mean Centered	0.39	0.02	19.69	<0.001	0.35	0.43
Intercept	506.81	5.27	96.09	<0.001	496.47	517.15

*Note.* LEP = limited English proficient; SWD = students with disabilities; FRL = free or reduced lunch; Coef. = coefficient; SE = standard error of the coefficient; z = standardized score.

**Table C.10. HLM Results for Third Grade Reading.**

Covariates	Coef.	SE	z	p-value	95% Conf. Interval	
<b>Student-Level Covariates</b>						
Treatment Group Membership	0.17	1.14	0.15	0.883	-2.07	2.40
Fall 2017 Reading i-Ready School Mean Centered	0.80	0.08	10.53	<0.001	0.65	0.94
Fall 2017 Mathematics i-Ready School Mean Centered	0.02	0.15	0.16	0.873	-0.27	0.32
<b>School-Level Covariates</b>						
Percent LEP	-19.18	6.59	-2.91	0.004	-32.09	-6.26
Percent SWD	-23.22	13.80	-1.68	0.092	-50.26	3.82
Percent White	-0.08	0.04	-1.93	0.054	-0.15	0.00
Percent FRL	-5.13	3.52	-1.46	0.144	-12.03	1.76
Fall Reading i-Ready School-Level Grand Mean Centered	0.67	0.01	74.12	<0.001	0.65	0.69
Fall Mathematics i-Ready School-Level Grand Mean Centered	0.36	0.02	19.48	<0.001	0.32	0.39
Intercept	539.56	3.66	147.50	<0.001	532.39	546.73

**Table C.11. HLM Results for Fourth Grade Reading.**

Covariates	Coef.	SE	z	p-value	95% Conf. Interval	
<b>Student-Level Covariates</b>						
Treatment Group Membership	3.80	1.08	3.52	<0.001	1.68	5.92
Fall 2017 Reading i-Ready School Mean Centered	0.15	0.07	2.11	0.035	0.01	0.28
Fall 2017 Mathematics i-Ready School Mean Centered	0.72	0.14	5.09	<0.001	0.44	1.00
<b>School-Level Covariates</b>						
Percent LEP	-5.48	6.21	-0.88	0.378	-17.65	6.69
Percent SWD	4.24	13.14	0.32	0.747	-21.53	30.00
Percent White	-0.05	0.04	-1.30	0.193	-0.12	0.02
Percent FRL	1.44	3.31	0.43	0.664	-5.05	7.94
Fall Reading i-Ready School-Level Grand Mean Centered	0.11	0.01	19.15	<0.001	0.10	0.12
Fall Mathematics i-Ready School-Level Grand Mean Centered	0.73	0.01	63.30	<0.001	0.71	0.76
Intercept	453.57	3.46	131.16	<0.001	446.79	460.35

*Note.* LEP = limited English proficient; SWD = students with disabilities; FRL = free or reduced lunch; Coef. = coefficient; SE = standard error of the coefficient; z = standardized score.

**Table C.12. HLM Results for Fifth Grade Reading.**

Covariates	Coef.	SE	z	p-value	95% Conf. Interval	
<b>Student-Level Covariates</b>						
Treatment Group Membership	5.92	1.04	5.69	<0.001	3.88	7.96
Fall 2017 Reading i-Ready School Mean Centered	0.13	0.08	1.75	0.080	-0.02	0.28
Fall 2017 Mathematics i-Ready School Mean Centered	0.86	0.12	6.94	<0.001	0.61	1.10
<b>School-Level Covariates</b>						
Percent LEP	2.24	6.07	0.37	0.712	-9.66	14.13
Percent SWD	-1.76	11.83	-0.15	0.882	-24.95	21.44
Percent White	0.02	0.03	0.70	0.485	-0.04	0.09
Percent FRL	1.25	2.99	0.42	0.676	-4.61	7.12
Fall Reading i-Ready School-Level Grand Mean Centered	0.10	0.01	17.49	<0.001	0.09	0.11
Fall Mathematics i-Ready School-Level Grand Mean Centered	0.78	0.01	72.42	<0.001	0.76	0.81
Intercept	469.40	3.29	142.68	<0.001	462.95	475.85

*Note.* LEP = limited English proficient; SWD = students with disabilities; FRL = free or reduced lunch; Coef. = coefficient; SE = standard error of the coefficient; z = standardized score.