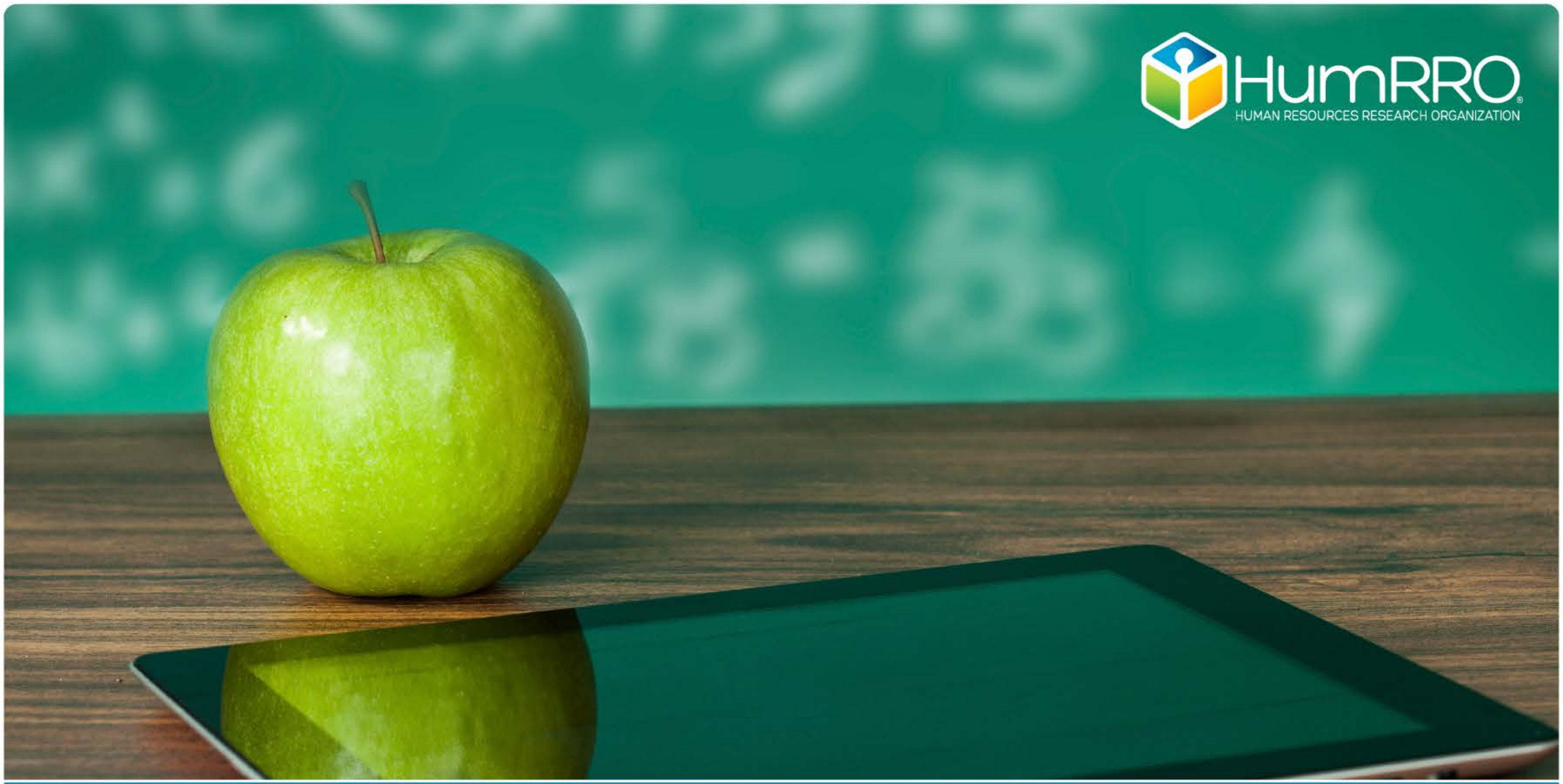


2019 No. 109

Impact Evaluation of Mathematics *i-Ready Instruction* for Middle School Grades using 2018 – 19 Data

Final Report

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

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

Date: January 15, 2020

Headquarters: 66 Canal Center Plaza, Suite 700, Alexandria, VA 22314 | Phone: 703.549.3611 | **humrro.org**

Impact Evaluation of Mathematics *i-Ready Instruction* for Middle School Grades using 2018 – 19 Data

Table of Contents

List of Tables

Table of Contents (Continued)

List of Figures

Impact Evaluation of Mathematics *i-Ready Instruction* for Middle School Grades using 2018 – 19 Data

Abstract

Curriculum Associates' *i-Ready® Instruction* is a supplemental, online personalized instruction program available for reading and mathematics[1]. The Human Resources Research Organization (HumRRO), in collaboration with Century Analytics, implemented a quasi-experimental design (QED) using 2018–19 *i-Ready Diagnostic and Instruction* data to evaluate the impact of Curriculum Associates' mathematics *i-Ready Instruction* on student mathematics achievement at grades 6–8. We hypothesized student achievement, as measured by the *i-Ready® Diagnostic,* would be higher for students using *i-Ready Instruction* for mathematics over a comparison group of students who did not use this instruction. We conducted matching to identify a set of comparison students demographically similar to our *i-Ready Instruction* treatment students for each grade level. First, we stratified our sample by gender, English learner status, disability status, and economic disadvantage status. Next, we used propensity score matching to identify analytic samples of *i-Ready Instruction* and comparison students matched on baseline mathematics student achievement. Students who received the *i-Ready Instruction* and students in the comparison group were administered the mathematics *i-Ready Diagnostic* assessments. To evaluate impact, hierarchical-linear modeling (HLM) was conducted separately for each analytic sample with students at level 1 and school at level 2. Results suggest students using *i-Ready Instruction* with fidelity performed statistically significantly better on mathematics performance than students in grades 6–8 who did not use this instruction. The effect sizes fall at the upper end of the range for which recent research by Kraft (2019) has found is typical of education within the interventions. These findings provide support that, when used with fidelity, student use of *i-Ready Instruction* for mathematics is tied to higher student mathematics achievement.

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® Diagnostic* (available for K–12) and *i-Ready® 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. The *i-Ready Diagnostic* assessments are independent measures often used by educators as classroom benchmark assessments. They can be used with or without *i-Ready Instruction.* We provide additional information on the validity and reliability of the *i-Ready Diagnostic* as a measure of student achievement in our methodology discussion below. *i-Ready Instruction* is a supplemental program that provides online, individualized instruction adjusted to student needs.

The Human Resources Research Organization (HumRRO) is an independent research organization that specializes in program evaluation and quantitative methodology. Century Analytics is a small business with various education research expertise including quasi-experimental design and What Works Clearinghouse (WWC) standards.

[1] https://www.curriculumassociates.com/products/i-ready

HumRRO and Century Analytics conducted an evaluation to examine the impact of *i-Ready Instruction* on mathematics achievement for students in middle school grades 6–8 using 2018–19 data. This was one in a series of evaluations examining the impact of Curriculum Associates' interventions on student achievement. This study was designed to meet the required rigor of the 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). To accomplish this, we used a quasi-experimental design (QED), established baseline equivalence between the treatment and comparison groups, included baseline achievement as a covariate, and used a sampling design that mitigates the effects of any confounding factors.

There were key differences between this study and past studies. Specifically, previous studies considered school as the unit of *i-Ready Instruction* assignment, whereas this study considered students as the unit of assignment. This change in unit of assignment acknowledges the inherent flexibility of *i-Ready Instruction* implementation. For example, some schools may implement at the school-level, the grade-level, or the classroom-level, while other schools may implement *i-Ready Instruction* at the individual student-level so they can target specific groups of students. In addition, our past studies included only schools using *i-Ready Diagnostic* and *Instruction*, or *i-Ready Diagnostic* only for the comparison group, with general education students. Thus, those schools using *i-Ready Diagnostic* (with or without *Instruction*) with select subsets of students were removed from our sample. Because our data support various types of implementation occurring across schools, and we understand it is Curriculum Associates intent that these different implementations are valid uses, this study includes students from schools that are implementing *i-Ready Diagnostic* with or without *Instruction* in a variety of ways.

Defining i-Ready Instruction

The impact of *i-Ready Instruction* on student achievement was the focus of this evaluation. *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 can be provided through *i-Ready Instruction.*

Curriculum Associates developed a Theory of Action (TOA) that features the key implementation components of *i-Ready Instruction*, the intended intermediate outcomes, and the intended long-term outcomes. The key implementation components highlight actions recommended by students, teachers, and leaders 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 levels, 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 achieve greater gains when using *i-Ready Instruction* for an average of at least 30 minutes per week, per subject area. In addition, Curriculum Associates recommends use for 12 to 18 calendar weeks between two administrations of the *i-Ready Diagnostic* (Curriculum Associates, 2018).

Research Questions

The purpose of this study was to determine the impact of *i-Ready Instruction* on student achievement in mathematics. We examined the following key research question separately for each grade 6–8 of our study:

> Do students who use *i-Ready Instruction* for mathematics have higher mathematics achievement as measured by the *i-Ready Diagnostic* than students who use *i-Ready Diagnostic* only?

We hypothesized that student achievement for mathematics would be higher for students who used *i-Ready Instruction* with fidelity, based on the criteria described in the TOA and user guidance (Curriculum Associates, 2019). Our hypothesis was based on the belief that students benefit from the *i-Ready Instruction* targeted to their specific needs in mathematics.

Methodology

In this section, we describe the methodology for conducting our impact analysis. We begin with initial design decisions. We then discuss the student selection and matching process, as well as our analytic model and examination of baseline equivalence. Finally, we discuss our impact analysis results.

Initial Design Decisions

Cluster-Level Design

We used the student as the unit of assignment for this study to acknowledge the flexibility intended by *i-Ready Instruction* and to include students from schools with various implementation types. Matching was conducted at the student-level and, thus, the analytic model examined the outcome at the student level. However, we also considered potential influence of school-level factors and thus decided to include a two-level analytic model with school characteristics at level 2 and students at level 1.

Baseline and Outcome Measure

We selected the *i-Ready Diagnostic* as both the baseline and outcome measure for all students participating in this study (i.e., *i-Ready Instruction* students and comparison group students). *i-Ready Diagnostic* for mathematics measures achievement aligned to common mathematics content and skills with demonstrated test score reliability. Marginal reliabilities are 0.98 and test-retest reliabilities are 0.87 for mathematics in grades 6–8. Therefore, this assessment meets the WWC 4.0 standards for an acceptable baseline and outcome measure (WWC, 2017a).

The *i-Ready Diagnostic* assessments align to college- and career-ready standards so that results can inform student placement decisions, offer explicit instructional advice, and prescribe resources for targeted instruction and intervention. The assessments are used by some schools and districts in conjunction with *i-Ready Instruction* and by others as a stand-alone diagnostic assessment without the use of *i-Ready Instruction*. The *i-Ready Diagnostic* assessments for mathematics and reading are currently used by more than 6.5 million students across the United

States. Thus, the use of *i-Ready Diagnostic* as the outcome measure allowed us to include a large sample of students from across the United States. The *i-Ready Diagnostic* is intended to be administered in a standardized manner across schools (Curriculum Associates, 2019b). Specifically, teachers are to schedule the first (fall) Diagnostic 2–3 weeks into the school year in two 45–50-minute sessions. Teachers also are encouraged to test technology to ensure proper function and have pencils and paper available as scratch paper. Test administrators provide instructions to their students and motivate them to do their best. Teachers monitor students as they complete the assessments.

Multiple studies have been conducted to support the reliability and validity of the mathematics *i-Ready Diagnostic* as well as its consistency with education standards used across the United States. Since being released in summer 2011, *i-Ready Diagnostic* 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.

Curriculum Associates has conducted multiple linking studies examining *i-Ready Diagnostic* scores for mathematics at grades 3–8 that provide evidence the *i-Ready Diagnostic* measures skills consistent with student expectations and can be used as a student mathematics achievement measure. For example, a study 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 state testing programs in Florida, Georgia, Indiana, Michigan, Mississippi, New York, North Carolina, Ohio, and Tennessee. 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 ranged from 0.82 (North Carolina End-of-Grade assessments) and 0.88 (Smarter Balanced and Michigan M-STEP). These studies also 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 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 Florida, Indiana, Louisiana, Michigan, Ohio, and South Carolina state standards.

Required Number of Students

We conducted power analyses using Optimal Design software (Spybrook et al., 2011) to identify the total number of students required at each grade level to reject the null hypothesis that there is no difference in student mathematics 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). Results of the power analyses indicated sample sizes of a minimum of 400 students would be sufficient to reach our desired statistical power of 0.80. This level of statistical power provides an 80% chance of detecting a statistically significant difference with 95% confidence, if one exists. Our student samples across all grades far exceeded the minimum.

Analytic Model

Our model for the impact analyses incorporated student- and school-variables. The baseline difference model used to estimate baseline equivalence for our matched sample was based on the impact model. As previously discussed, we chose a two-level model with level 1 as the student and level 2 as school.

Impact Model

We used HLM to estimate the impact of *i-Ready Instruction* on student achievement. We included the following student-level covariates in each analysis:

- Group membership (0 = comparison; 1 = treatment)
- *i-Ready Diagnostic* mathematics baseline performance (grand mean centered)
- Blocking variables (i.e., dummy codes) to account for strata used in matching (described in the matching section of this report)

Although we considered the student to be our unit of assignment, with the understanding that many schools intentionally do not use *i-Ready Instruction* with all students, we also wanted to capture and control for potential school-level factors. We were especially interested in identifying variables that would provide unique information from the student-level variables. We used the following school-level covariates in each analysis:

- Traditional school indicator (0 = 6–8 structure; 1 = other)
- Location (town, suburban, rural, city)
- Charter/magnet school indicator (0 = not charter or magnet; 1 = charter or magnet)
- Percent white students
- Percent of students eligible for free and reduced price lunch (FRL)

Our Level 1 model described the relationship between student outcomes, student-level characteristics, the baseline covariate, and the strata used for matching. This model level also included the treatment indicator. We specified level 1 of the model as follows:

$$Y_{ij} = \beta_{0j} + \beta_{1j}(GROUP_{ij}) + \beta_{2j}(PRE_{ij} - PRE_{..}) + \Sigma\beta_q(STRATA_{ij}) + e_{ij}$$

Where Y_{ij} is the outcome for student *i* in school *j*. β_{0j} is the adjusted mean outcome for comparison students in school *j*. β_{1j} is the adjusted mean difference in outcome due to the student's group membership (i.e., the treatment effect), and GROUP is an indicator variable coded 1 for students in the *i-Ready Instruction* group and 0 for students in the comparison group. β_{2j} is the adjusted difference in outcome due to the student's baseline achievement score (grand mean centered). β_q is a vector of blocking variables to account for the strata used in matching. 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 follows:

$$\beta_{0j} = \gamma_{00} + \gamma_{01}(STRUCTURE_j) + \gamma_{02}(CHARTER_j) + \gamma_{03}(PERWHITE_j) + \gamma_{04}(PERFRL_j) + \Sigma\gamma_k(LOCATION_j) + u_{0j}$$

$\beta 1j = \gamma 10$
$\beta 2j = \gamma 20$
$\Sigma\beta p = \gamma p0$
$\Sigma\beta q = \gamma q0$

Where $\gamma 00$ is the grand mean. $\gamma 01$ is added to control for school grade-level structure where STRUCTURE is coded as 0 for schools with a typical grade level structure (6–8 for middle school) and 1 for schools with an atypical grade structure. $\gamma 02$ is the additive effect for charter or magnet schools. $\gamma 03$ and $\gamma 04$ are added to control for school characteristics of percent white and percent FRL, respectively. $\Sigma\gamma k$ is a vector of three dummy variables to control for school location. $u0j$ is the random error in the achievement outcome associated with school *j*. The regression slopes for the treatment, student baseline achievement, student demographics and strata are fixed across schools.

Baseline Difference Model

We used the model below to estimate the baseline difference between students in the treatment group and the comparison group. This model follows the same structure as the impact analysis model but excludes covariates.

We specified level 1 of the model as:

$$Yij = \beta 0j + \beta 1j(GROUPij) + \Sigma\beta q(STRATAij) + eij$$

Where Yij is the baseline for student *i* in school *j*. $\beta 0j$ is the adjusted mean outcome for comparison students in school *j*. $\beta 1j$ is the adjusted difference in outcome due to the student's study group membership (i.e., the baseline difference), and GROUP is an indicator variable coded 1 for students in the *i-Ready Instruction* group and 0 for students in the comparison group. βq is a vector of blocking variables to account for the strata used in matching. eij 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:

$\beta 0j = \gamma 00 + u0j$
$\beta 1j = \gamma 10$
$\Sigma\beta q = \gamma q0$

Identifying a Student Sample

Defining Eligibility

For each grade level, we started with a student-level *i-Ready* usage file of mathematics *i-Ready Diagnostic* and *i-Ready Instruction* use in 2018–19 for students who had at a minimum fall and spring *i-Ready Diagnostic* scores. We next filtered to include only public-school students, which included traditional public schools and public charter and magnet schools. This ensured we were including only students in a relatively traditional school environment with expectations to follow state adopted college and career ready standards.

We also filtered our sample based on availability of student-level demographic variables that were identified for inclusion in matching and the impact analysis model. Only students with available demographic data for (a) gender, (b) English learner (EL) status, (c) special education status, and (d) economic disadvantage status were included. We conducted data checks prior to removing schools that indicated students with available demographic data were not different on academic achievement, as measured by the *i-Ready Diagnostic*, than those who did not have demographic data. These checks provided assurance that data were missing at random. However, we also note that users of the *i-Ready* products tend to be of higher percentage minority and low income schools compared to all United States schools; thus, though we were confident our student sample used for matching was academically representative of the public school students using *i-Ready Diagnostic* or *i-Ready Diagnostic and Instruction*, we do not expect they are representative of all students in the United States.

In addition, for a student to be eligible for the treatment group, they must have used *i-Ready Instruction* for mathematics a minimum of 18 distinct weeks for an average of at least 30 minutes per week (Curriculum Associates, 2018). This was consistent with guidance on the minimum *i-Ready Instruction* usage at the student level for attaining intended goals of improved student mathematics achievement. These students also needed to have attended a school that began using *i-Ready Instruction* to some extent prior to the 2018–19 school year. This requirement is based on the understanding that *i-Ready Instruction* implementation requires a start-up time to learn the technology and adjustments to scheduling before *i-Ready Instruction* is fully up and running. To be eligible for the comparison group, students must not have used any *i-Ready Instruction* for mathematics in 2018–19. We removed students not meeting the treatment or comparison eligibility requirements from the datafile used in matching.

For all middle school grades, between 20 to 25% of schools had students assigned to the treatment group and students assigned to the comparison group. Though we expected some overlap, this proved problematic for achieving baseline equivalence. The was likely because these schools were not assigning students randomly to receive either *i-Ready Diagnostic and Instruction* or *i-Ready Diagnostic* only. Rather, the data suggested that lower achieving students were being assigned to receive *i-Ready Diagnostic and Instruction*, and the higher performing students were using *i-Ready Diagnostic* only. By making such assignments at the school level and including schools as a level in the predetermined baseline difference model, comparisons within schools resulted in two groups dissimilar on baseline achievement. For the purpose of this study, we eliminated all schools that included students in both groups.

Matching

We conducted matching at the student level using a multi-step process. Matching was conducted separately by grade (6–8). Thus, we conducted each matching step three separate times to identify three analytic samples (i.e., three grades).

First, we stratified our sample by gender, EL status, special education status, and economic disadvantage status. This assured that students were only matched to students with identical demographic characteristics on these four variables. The variables were selected because they are known to be related to student achievement (Hanover Research, 2014; van Langen, Bosker, & Dekkers, 2006) and were available through the *i-Ready* usage datafiles. This stratification resulted in 16 strata at each grade. Each stratum contained treatment and comparison students. In some strata, the treatment group was larger than the comparison group, or vice versa. Within each stratum, we used logistic regression to compute a propensity score for each student (Guo & Fraser, 2010). The propensity scores predicted the chance a student belonged to the group (treatment or comparison) with the smallest number of students, indicated by a value ranging between 0 and 1, based on the fall *i-Ready Diagnostic* scores. We used the propensity scores to match each student from the smallest group (treatment or comparison) to a student from the largest group. We matched using the nearest neighbor method without replacement (Stuart, 2010). Once matching was conducted for all strata within a grade, we combined the data from all strata into one analytic sample.

Following specification of our analytic and baseline difference models, we removed an average of 9.1% of students across the three analytic samples who had incomplete data on the school-level variables included in the impact model. This resulted in unequal numbers of students in comparison and treatment groups. Figure 1 summarizes the demographic makeup of the final set of students in each analytic sample. The counts of students included in each group can be found in Table 1 on page 11. As shown, the stratification process used in matching ensured the *i-Ready Instruction* and comparison groups were highly similar on the key demographic variables, despite the need to remove a small percentage of the sample to account for missing school-level variables.

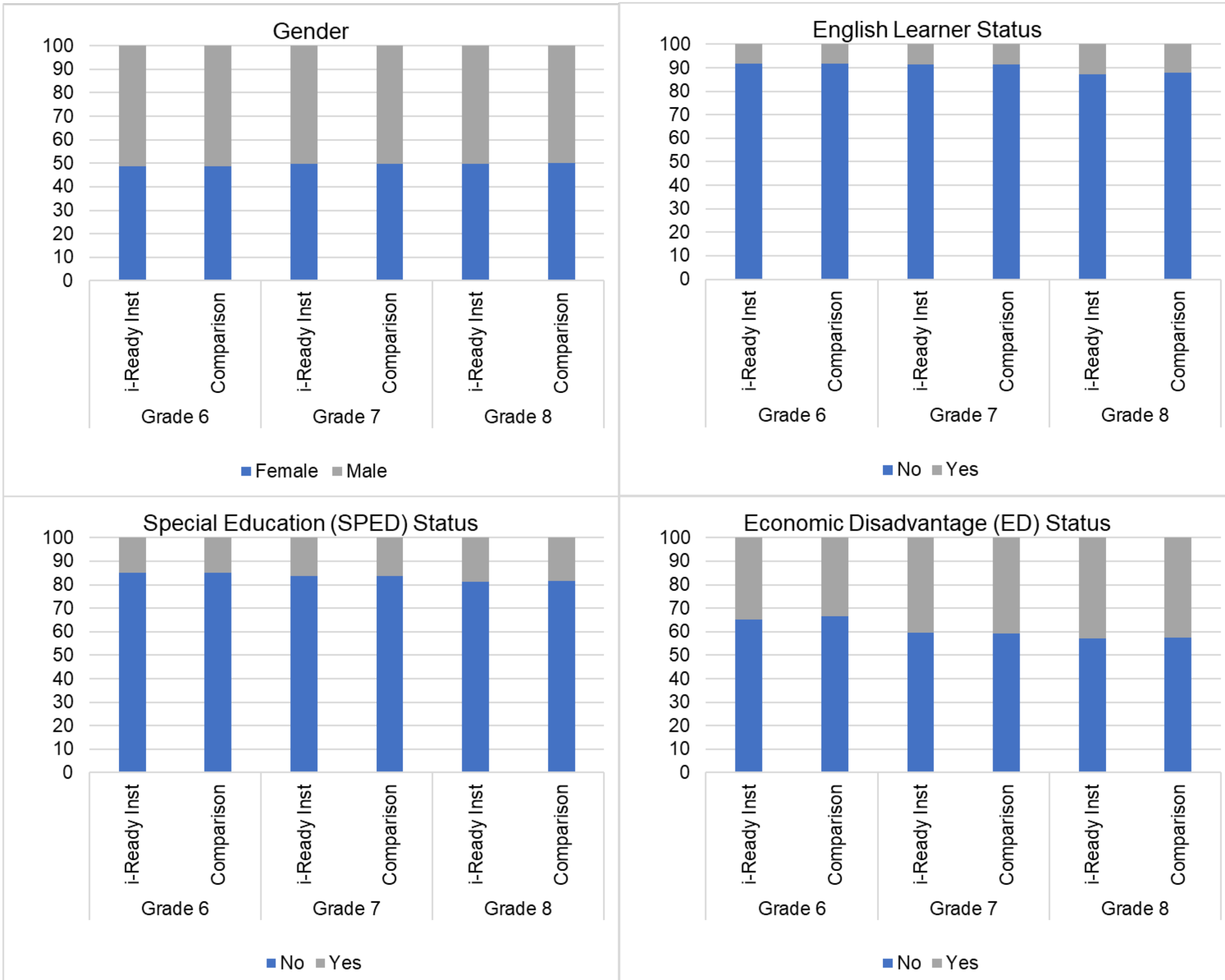


Figure 1. Demographic makeup of final matched i-Ready Instruction and comparison samples

Although our sampling focused on the student-level, to gain additional understanding of where our student-sample was from, we examined the distribution of students across urbanicity categories, as defined through school-level variables of the National Center for Education Statistics (NCES) publicly available database. Figure 2 shows that schools in the *i-Ready Instruction* and comparison groups share a relatively similar urbanicity distribution.

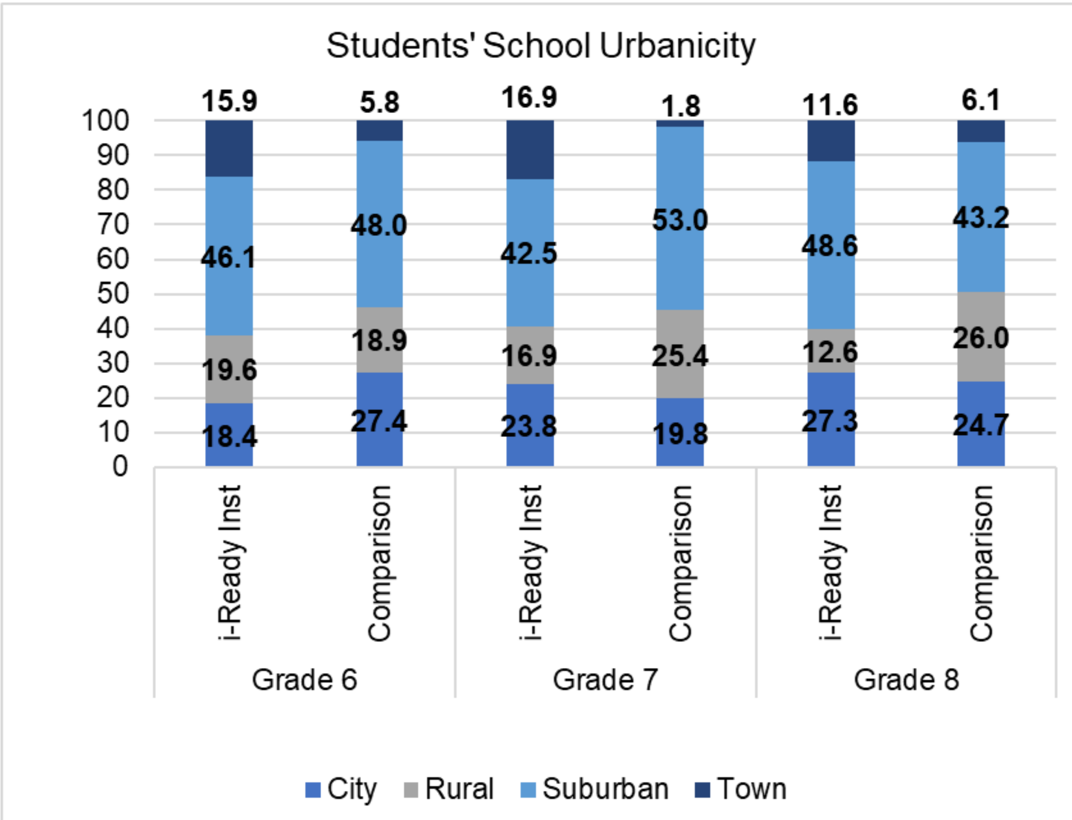


Figure 2. Students' school urbanicity for final matched i-Ready Instruction and comparison samples

Baseline Equivalence

Once our analytic samples were identified, we used our baseline difference model to estimate the adjusted mean differences between our *i-Ready Instruction* and comparison groups of students at each grade level. We converted the estimated baseline difference between students in the two groups to an effect size to evaluate baseline equivalence for each of the three analytic samples. For all three samples, Hedges' *g* was much smaller than the WWC required threshold of 0.25 (see Table 1), so we determined the groups were baseline equivalent (WWC, 2017b).

Table 1. Mathematics Baseline Equivalence Statistics for i-Ready Instruction (Treatment) and Comparison Groups by Grade

Grade	Group	Students	*i-Ready* Mean	*i-Ready* SD	Adj Mean Diff (SE)	Effect Size
6	*i-Ready Instruction*	10,688	477.02	29.91	-1.14	-0.04
	Comparison	10,335	478.16	30.43	(1.39)	
7	*i-Ready Instruction*	8,474	485.37	31.79	-0.77	-0.02
	Comparison	8,035	486.14	32.01	(1.99)	
8	*i-Ready Instruction*	4,917	484.01	34.67	-7.18	-0.21
	Comparison	4,937	491.19	32.95	(2.11)	

Notes: SD = standard deviation of *i-Ready* scores, Adj Mean Diff = adjusted mean difference between *i-Ready Instruction* and comparison groups, and Effect Size = Hedge's *g.*

Impact Analysis Results

After confirming our matched samples were baseline equivalent at each grade, we estimated the impact of *i-Ready Instruction* on student achievement using the analytic model described above, with spring 2019 *i-Ready Diagnostic* scores as the outcome. Analyses were conducted separately for each grade. This section describes the results of the analysis. Full information on the model results, including student- and school-level covariate parameters, are presented in Appendix B.

In addition to estimating the impact of *i-Ready Instruction*, we also examined three model assumptions associated with two-level HLM—residual normality, independence, and homoscedasticity—using the MIXED_DX macro in SAS (Bell, Smiley, Ene, & Blue, 2014). No major violations were found. Additional details regarding the assumption checks are available in Appendix C.

Table 2 contains the impact model results by grade for mathematics spring *i-Ready Diagnostic* scores. For all grade levels, the adjusted mean differences were positive, indicating the *i-Ready Instruction* group earned higher scores than the matched comparison group. All mean differences were statistically significant (α = .05) with Hedge's *g* effect sizes ranging from 0.15 to 0.17. These effect sizes are promising for an education intervention. Traditional guidance has suggested these effect sizes are modest (Lipsey et al. 2012); however, recent research by Kraft (2019) notes traditional guidelines, including those reported by Lipsey, are often too rigid for the realities of education interventions. He specifies effect size ranges of 0.03–0.17 as typical of education interventions and that these often represent a meaningful effect. He suggests effect sizes should be considered in conjunction with all aspects of an intervention, including the magnitude of the treatment contrast and costs.

Table 2 also provides the intra-class correlations (ICCs) by grade. The ICCs measure the proportion of the variance 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 0.31 (grade 6) to 0.33 (grades 7 and 8). This suggests the majority of variance is due to factors other than school-level differences; however, we prefer ICCs to be below 0.20. The elevated ICCs may be impacted by the variation in implementation methods and our decision to model implementation at the student-level. This finding will assist in future efforts for identifying the most appropriate unit of assignment to account for these variations.

Table 2. Impact Analysis Results for i-Ready Instruction (Treatment) and Comparison Groups for Mathematics Student Achievement by Grade

Grade	Group	ICC	Students	*i-Ready* Mean	*i-Ready* SD	Adj Mean Diff (SE)	*p*-value	Effect Size
6	*i-Ready Instruction*	0.31	10,688	498.11	33.83	5.09	<0.001	0.15
	Comparison		10,335	493.02	34.21	(0.81)		
7	*i-Ready Instruction*	0.33	8,474	501.81	36.27	6.06	<0.001	0.17
	Comparison		8,035	495.75	36.91	(0.95)		
8	*i-Ready Instruction*	0.33	4,917	506.49	39.49	6.21	<0.001	0.16
	Comparison		4,937	500.29	36.79	(1.21)		

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.*

Summary and Discussion

At all grades, impact analyses suggest that middle school students who use *i-Ready Instruction* with fidelity have higher achievement in mathematics when compared to students who did not use *i-Ready Instruction*. At each grade, students in the *i-Ready Instruction* group had a statistically significantly higher mathematics *i-Ready Diagnostic* score than did students in a matched comparison group.

The effect sizes provided additional evidence *i-Ready Instruction* is beneficial for improving student mathematics. Recent research (Kraft, 2019) suggests education interventions typically attain effects ranging from 0.03 to 0.17. Our effect sizes for all grades fell at the upper end of this range. Kraft (2019) notes one should consider various factors when interpreting effect sizes, including a program's cost relative to its benefits and the size of the treatment contrast. For example, we note that *i-Ready Instruction* is a supplemental intervention that requires only 12 to 18 weeks of 30 minutes or more per week during a school year to be considered implemented with fidelity at the student level. In addition, because *i-Ready Instruction* is not a full curriculum and there are likely many similarities between what else students are exposed to whether in the *i-Ready Instruction* group or comparison group, we believe the contrast between our treatment and comparison group is likely minimal. Similarly, it is possible some students in our comparison group were exposed to interventions like *i-Ready Instruction*. Thus, given the required effort for using *i-Ready Instruction* with fidelity is relatively low, and the contrast between the *i-Ready Instruction* and comparison group small compared to a more involved intervention or curricular program, we feel confident that our effect sizes are meaningful.

Kraft (2019) also points out that the U.S. education system is decentralized, and implementation procedures are ultimately controlled by local schools and/or teachers. As a QED, this study did not attempt to control for curriculum, supplemental resources, or classroom structure. Students in both groups were not participants in a research study but rather they were actual customers and everyday users, and *i-Ready Instruction* was carried out in real-world conditions. We may have found even larger effect sizes 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. Thus, the fact we were able to find significant findings for all grade levels despite the lack of controls is promising.

We conducted this study differently from a past study using 2017–18 data by considering the unit of assignment to be the student instead of the school. Additionally, we used 2018–19 data to take advantage of the most recent available information. Despite these key differences, our results were consistent – both studies found the treatment performed better than the comparison. This replication provides confidence that students using *i-Ready Instruction* in conjunction with the *i-Ready Diagnostic* show greater mathematics achievement compared to a comparison group using *i-Ready Diagnostic* only.

Our study was conducted as a rigorous QED to meet the current standards described by the WWC (WWC, 2017) to achieve a rating of *Meets WWC Group Design Standards with Reservations*. In addition, because we found statistically significant positive effects for all grades, 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).

Limitations and Implications for Future Studies

This study provides strong evidence supporting mathematics *i-Ready Instruction* use for students. Through our long-standing relationship with Curriculum Associates and multiple impact evaluations, including the current study, we have developed recommendations for the foci of future studies that may provide additional evidence to support the impact of *i-Ready Instruction*.

First our ICCs were above 0.20 for all grades, suggesting school differences may be important for matching and estimating treatment effects. However, the data also revealed large variations in how many students at a given school or grade within a school used *i-Ready Instruction* with fidelity. Future studies may look to explore the grade or classroom as the unit of assignment. We also found that schools choosing to implement *i-Ready Instruction* to select students and *i-Ready Diagnostic* only to others were generally selecting low performing students for the *i-Ready Diagnostic and Instruction* group and other students for the *i-Ready Diagnostic* only group such that baseline equivalence could not be achieved for students within school. We recommend Curriculum Associates collect information directly from schools to understand their intended implementation so this information can be incorporated into sample selection and analytic models.

Second, we note our study was a QED with the typical limitations, including a lack of information on implementation decisions made at each school and within each classroom. We recommend randomized control trials (RCTs) in the future even if only a small sample of schools and students is included. We also suggest including only one district to allow greater control on implementation.

Finally, our treatment group was compared to a matched 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 and students as a comparison group may have attenuated the effects of *i-Ready Instruction* use had this treatment group been compared to a "business-as-usual" comparison group. Future studies might examine the impact of *i-Ready Instruction* using a set of comparison schools and students not implementing any Curriculum Associates products. This would require an external achievement measure, potentially a state assessment, as the baseline and outcome measure.

Quality Control Procedures

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

Rules for identifying treatment and comparison groups were determined through collaboration between the three study partners. Curriculum Associates provided information on the various components of *i-Ready Instruction* and the frequency for which it should be used for implementation with fidelity. They also 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 treatment and comparison group criteria in which all partners were confident.

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 results reported in this study were verified by researchers from both organizations.

References

Bell, B. A., Smiley, W., Ene, M., & Blue, G. L. (2014). *An Intermediate Primer to Estimating Linear Multilevel Models using SAS PROC MIXED.* SAS Global Forum 2014 Proceedings. Retrieved from https://support.sas.com/resources/papers/proceedings14/1869-2014.pdf

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. (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.

Hanover Research. (2014). *Improving Student Achievement and Closing the Achievement Gap.* Washington, DC: Hanover Research. Retrieved from https://www.rcoe.us/educational-services/files/2015/12/10c-Hanover_Improving_Student_Achievement_and_Closing_the_Achievement_Gap__12-2014.pdf

Kraft, Matthew. (2019). Interpreting Effect Sizes of Education Interventions. (EdWorkingPaper: 19-10). Retrieved from Annenberg Institute at Brown University: https://doi.org/10.26300/8pjp-2z74

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/guidanceuseseinvestment.pdf.

van Langen, A., Bosker, R., & Dekkers, H. (2006). Exploring Cross-National Differences in Gender Gaps in Education. *Educational Research and Evaluation, 12*(2), 155-177.

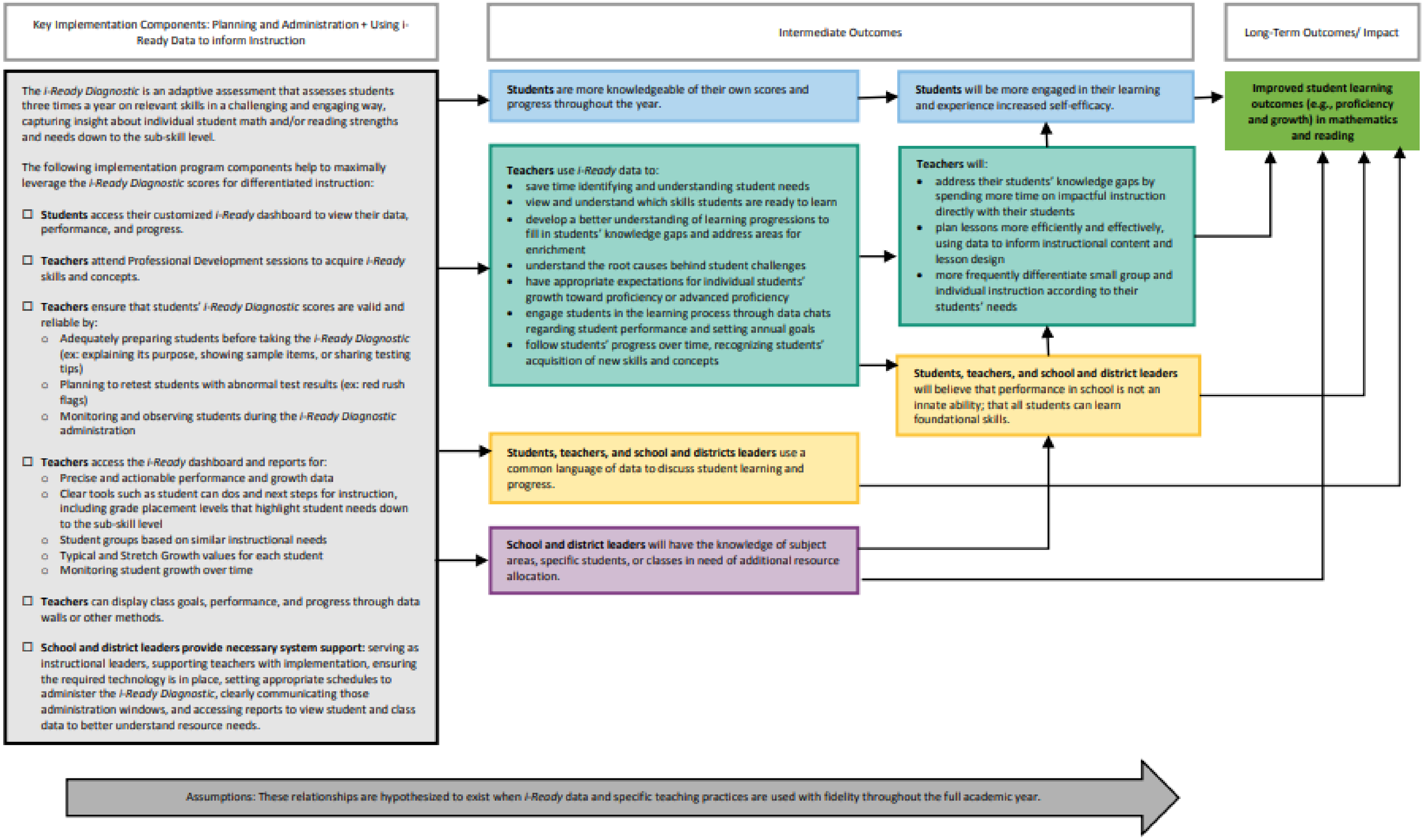
What Works Clearinghouse, Institute of Education Sciences, U.S. Department of Education. (2017a, October). *What Works Clearinghouse: Procedures Handbook (Version 4.0).* Retrieved from http://whatworks.ed.gov.

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

i-Ready Diagnostic

The *i-Ready Diagnostic* is an adaptive assessment that assesses students on relevant skills in a challenging and engaging way, capturing insight about student learning down to the subskill level. Teachers are provided with precise, actionable data and instructional recommendations to more seamlessly differentiate classroom instruction according to their students' needs, saving teachers valuable time. This allows teachers to deliver more impactful instruction to increase student growth and proficiency.



Appendix B. Impact HLM Coefficients

Table B.1. HLM Results for Sixth Grade Mathematics

Covariates	Coef.	SE	z	*p*-value	95% Conf. Interval	
Student-Level Covariates						
Treatment Group Membership	5.09	0.81	6.29	< .001	3.51	6.68
Fall 2018 Mathematics *i-Ready* Grand Mean Centered	0.91	0.00	195.62	< .001	0.90	0.92
Student-Level Stratum						
Female, ELL = 0, SpEd = 0, EcDis = 0	4.85	1.63	2.99	0.003	1.67	8.04
Female, ELL = 1, SpEd = 0, EcDis = 0	5.19	1.88	2.76	0.006	1.50	8.88
Female, ELL = 0, SpEd = 1, EcDis = 0	-1.53	1.73	-0.89	0.375	-4.92	1.85
Female, ELL = 0, SpEd = 0, EcDis = 1	2.80	1.61	1.73	0.083	-0.37	5.96
Female, ELL = 1, SpEd = 1, EcDis = 0	-1.92	2.52	-0.76	0.446	-6.87	3.03
Female, ELL = 0, SpEd = 1, EcDis = 1	-3.01	1.79	-1.68	0.093	-6.53	0.50
Female, ELL = 1, SpEd = 0, EcDis = 1	2.57	1.82	1.41	0.158	-0.99	6.13
Female, ELL = 1, SpEd = 1, EcDis = 1	-2.21	2.87	-0.77	0.442	-7.84	3.42
Male, ELL = 0, SpEd = 0, EcDis = 0	4.25	1.63	2.61	0.009	1.06	7.44
Male, ELL = 1, SpEd = 0, EcDis = 0	4.14	1.83	2.26	0.024	0.55	7.74
Male, ELL = 0, SpEd = 1, EcDis = 0	-1.32	1.67	-0.79	0.430	-4.59	1.96
Male, ELL = 0, SpEd = 0, EcDis = 1	2.06	1.62	1.27	0.204	-1.11	5.22
Male, ELL = 1, SpEd = 1, EcDis = 0	-3.54	2.20	-1.61	0.108	-7.85	0.78
Male, ELL = 0, SpEd = 1, EcDis = 1	-1.53	1.71	-0.89	0.372	-4.87	1.82
Male, ELL = 1, SpEd = 0, EcDis = 1	1.13	1.78	0.63	0.526	-2.36	4.61
School-Level Covariates						
Charter or Magnet Designation	-0.48	1.02	-0.47	0.635	-2.49	1.52
Traditional Middle School	4.79	0.76	6.34	< .001	3.31	6.27
Percent non-white students	-0.04	0.02	-2.31	0.021	-0.07	-0.01
Percent FRL students	-0.03	0.02	-1.40	0.160	-0.06	0.01
Locale – Suburban	-0.77	1.20	-0.64	0.522	-3.13	1.59
Locale – Rural	-0.53	0.90	-0.59	0.552	-2.29	1.22
Locale – City	-1.27	1.38	-0.92	0.359	-3.97	1.44
Intercept						
Intercept	490.49	2.06	237.73	< .001	486.45	494.54

Note. Stratum 16 and Locale – Town were used as reference groups in the model.

Table B.2. HLM Results for Seventh Grade Mathematics

Covariates	Coef.	SE	z	*p*-value	95% Conf. Interval	
Student-Level Covariates						
Treatment Group Membership	6.06	0.95	6.41	< .001	4.21	7.91
Fall 2018 Mathematics *i-Ready* Grand Mean Centered	0.93	0.01	170.21	< .001	0.92	0.94
Student-Level Stratum						
Female, ELL = 0, SpEd = 0, EcDis = 0	7.16	2.42	2.96	0.003	2.42	11.90
Female, ELL = 1, SpEd = 0, EcDis = 0	7.62	2.65	2.88	0.004	2.44	12.81
Female, ELL = 0, SpEd = 1, EcDis = 0	3.85	2.51	1.54	0.125	-1.06	8.76
Female, ELL = 0, SpEd = 0, EcDis = 1	5.52	2.40	2.30	0.022	0.81	10.23
Female, ELL = 1, SpEd = 1, EcDis = 0	7.14	3.31	2.16	0.031	0.66	13.62
Female, ELL = 0, SpEd = 1, EcDis = 1	1.73	2.59	0.67	0.505	-3.35	6.81
Female, ELL = 1, SpEd = 0, EcDis = 1	3.83	2.61	1.47	0.143	-1.29	8.94
Female, ELL = 1, SpEd = 1, EcDis = 1	0.62	3.28	0.19	0.850	-5.82	7.06
Male, ELL = 0, SpEd = 0, EcDis = 0	6.41	2.42	2.65	0.008	1.67	11.16
Male, ELL = 1, SpEd = 0, EcDis = 0	7.07	2.62	2.70	0.007	1.94	12.20
Male, ELL = 0, SpEd = 1, EcDis = 0	2.87	2.46	1.17	0.243	-1.95	7.70
Male, ELL = 0, SpEd = 0, EcDis = 1	4.00	2.40	1.66	0.096	-0.71	8.71
Male, ELL = 1, SpEd = 1, EcDis = 0	2.98	3.33	0.89	0.371	-3.55	9.51
Male, ELL = 0, SpEd = 1, EcDis = 1	0.78	2.51	0.31	0.757	-4.15	5.70
Male, ELL = 1, SpEd = 0, EcDis = 1	2.85	2.59	1.10	0.270	-2.22	7.92
School-Level Covariates						
Charter or Magnet Designation	0.72	1.14	0.63	0.528	-1.52	2.97
Traditional Middle School	2.27	0.86	2.64	0.008	0.59	3.96
Percent non-white students	0.00	0.02	-0.22	0.829	-0.04	0.04
Percent FRL students	-0.05	0.02	-2.16	0.031	-0.09	0.00
Locale – Suburban	0.19	1.46	0.13	0.894	-2.67	3.06
Locale – Rural	-0.97	1.05	-0.93	0.354	-3.03	1.08
Locale – City	-1.66	1.73	-0.96	0.337	-5.04	1.73
Intercept						
Intercept	493.02	2.84	173.78	< .001	487.46	498.58
Note. Stratum 16 and Locale – Town were used as reference groups in the model.						

Table B.3. HLM Results for Eighth Grade Mathematics

Covariates	Coef.	SE	z	*p*-value	95% Conf. Interval	
Student-Level Covariates						
Treatment Group Membership	6.21	1.21	5.14	< .001	3.84	8.57
Fall 2018 Mathematics *i-Ready* Grand Mean Centered	0.89	0.01	123.30	< .001	0.87	0.90
Student-Level Stratum						
Female, ELL = 0, SpEd = 0, EcDis = 0	8.58	2.32	3.70	< .001	4.04	13.11
Female, ELL = 1, SpEd = 0, EcDis = 0	7.23	2.67	2.71	0.007	2.00	12.47
Female, ELL = 0, SpEd = 1, EcDis = 0	1.04	2.49	0.42	0.676	-3.84	5.93
Female, ELL = 0, SpEd = 0, EcDis = 1	6.32	2.28	2.78	0.005	1.86	10.78
Female, ELL = 1, SpEd = 1, EcDis = 0	4.75	4.27	1.11	0.267	-3.63	13.13
Female, ELL = 0, SpEd = 1, EcDis = 1	2.07	2.48	0.84	0.404	-2.79	6.94
Female, ELL = 1, SpEd = 0, EcDis = 1	5.92	2.46	2.41	0.016	1.11	10.73
Female, ELL = 1, SpEd = 1, EcDis = 1	-0.14	4.16	-0.03	0.973	-8.29	8.01
Male, ELL = 0, SpEd = 0, EcDis = 0	7.35	2.32	3.17	0.002	2.80	11.89
Male, ELL = 1, SpEd = 0, EcDis = 0	7.07	2.60	2.72	0.007	1.97	12.17
Male, ELL = 0, SpEd = 1, EcDis = 0	1.83	2.39	0.76	0.446	-2.87	6.52
Male, ELL = 0, SpEd = 0, EcDis = 1	4.51	2.28	1.98	0.048	0.04	8.98
Male, ELL = 1, SpEd = 1, EcDis = 0	2.85	3.74	0.76	0.445	-4.48	10.18
Male, ELL = 0, SpEd = 1, EcDis = 1	-1.00	2.41	-0.41	0.679	-5.72	3.73
Male, ELL = 1, SpEd = 0, EcDis = 1	3.58	2.46	1.45	0.146	-1.25	8.41
School-Level Covariates						
Charter or Magnet Designation	0.31	1.53	0.20	0.842	-2.70	3.31
Traditional Middle School	1.21	1.14	1.07	0.286	-1.02	3.45
Percent non-white students	0.02	0.03	0.84	0.399	-0.03	0.08
Percent FRL students	-0.06	0.03	-1.89	0.059	-0.12	0.00
Locale – Suburban	0.22	1.93	0.12	0.908	-3.55	4.00
Locale – Rural	-2.70	1.38	-1.95	0.051	-5.41	0.01
Locale – City	0.35	2.25	0.15	0.878	-4.06	4.75
Intercept						
Intercept	495.12	3.06	161.71	< .001	489.12	501.12
Note. Stratum 16 and Locale – Town were used as reference groups in the model.						

Appendix C. Model Assumption Checks

We examined three model assumptions associated with two-level HLM – residual normality, independence, and homoscedasticity – using the MIXED_DX macro in SAS (Bell, Smiley, Ene, & Blue, 2014) based on the analytic model for all three grade levels of this study. The MIXED_DX macro provides visual output including box-and-whisker plots, histograms, scatter plots, and summary tables to examine residual normality, linearity, homoscedasticity, and influential outliers. The macro provides this information for level 1 and level 2 residuals.

We reviewed plots and summary tables at level 1 and level 2 for each grade level. These checks provided assurance that our analytic model was appropriate for our data. We examined histograms, box and whisker plots, and scatter plots to check residual normality. These plots supported that our residuals were generally normally distributed – particularly, the histograms of level-2 residuals produced highly symmetrical bell shape with little skewness or kurtosis. The level-1 residuals had some skewness but were close enough to normal to allow confidence. There was no evidence when examining level 1 residuals of clearly non-normal distributions such as a bi-modal distribution. Violation of assumptions of normality of level 1 residuals can adversely affect estimation of random effect coefficients and variance-covariance components, but typically will not adversely affect estimation of standard errors and, therefore, inferences regarding statistical significance. Given the primary purpose of the models was estimating treatment effects, the slight lack of normality of the level 1 residuals likely did not have implications for the findings presented in this report.

Scatter plots of predicted values against residuals at level 1 and level 2 clearly illustrated random distributions and provided support for that assumptions regarding independence and homoscedasticity were not violated.