2020 No. 048

Impact Evaluation of Mathematics *i-Ready* for Striving Learners Using 2018–19 Data

Final Report

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

Authors:
Bruce Randel (Century Analytics)
Matthew Swain
Rebecca Norman Dvorak
Elisabeth Spratto
Jordan Yee Prendez

Date: December 10, 2020

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

Impact Evaluation of Mathematics *i-Ready* for Striving Learners Using 2018–19 Data

Table of Contents

Key Findings 1

Abstract 2

Introduction 3
- Defining i-Ready Implementation 3

Research Questions 4

Methodology 4
- Design 5
- Analytic Samples 7
- Analytic Models 7

Results 8
- Benchmark Impact Analysis 8
- Exploratory Analyses 19
- Summary and Discussion 20
- Limitations and Implications for Future Studies 21
- Quality Control Procedures 22

References 23

List of Appendices

Appendix A: Selection of Analytic Samples A-1

Appendix B: Analytic Model Descriptions B-1

Appendix C: Confirmatory Impact HLM Coefficients for Striving Learners C-1

Appendix D: Confirmatory Impact HLM Coefficients for Striving Learners at the Bottom 20th Percentile D-1

Appendix E: Exploratory HLM with Black or African American by Treatment Interaction E-1

Appendix F: Impacts and Baseline to Outcome Gains for Black or African American by Treatment Interactions F-1

Appendix G: Exploratory HLM with Hispanic Origin by Treatment Interaction G-1

Appendix H: Impacts and Baseline to Outcome Gains for Hispanic Origin by Treatment Interactions H-1

Appendix I: Model Assumption Checks I-1

Table of Contents (Continued)

List of Tables

Table 1. Impact Analysis Results for Striving learners of i-Ready (Treatment) Schools Compared to Comparison Schools for Mathematics Student Achievement at Grades 2–5 10

Table 2. Baseline to Outcome Change in Mathematics Diagnostic Performance for the Striving Learners in i-Ready (Treatment) Schools Compared to Comparison Schools at Grades 2–5 11

Table 3. Impact Analysis Results for Striving Learners at the Bottom 20th Percentile of i-Ready (Treatment) Schools Compared to Comparison Schools for Mathematics Student Achievement at Grades 2–5 15

Table 4. Baseline to Outcome Change in Mathematics Diagnostic Performance for Students in i-Ready (Treatment) Schools Compared to Comparison Schools at Grades 2–5 16

Table 5. Summary of Black or African American by Treatment Interactions, by Grade 19

Table 6. Summary of Hispanic Origin by Treatment Interactions by Grade 20

Table A.1. Demographic Characteristics of Matched Mathematics i-Ready (Treatment) and Comparison Schools, by Grade A-1

Table A.2. Mathematics Baseline Equivalence Statistics for i-Ready (Treatment) and Comparison Striving Learner Groups, by Grade A-2

Table A.3. Mathematics Baseline Equivalence Statistics for i-Ready (Treatment) and Comparison Bottom 20th Percentile Striving Learner Groups, by Grade A-2

Table A.4. Mathematics Baseline Equivalence Statistics for Grade 4 Students of Hispanic Origin of the i-Ready (Treatment) and Comparison Groups A-3

Table C.1. Grade 2 Mathematics HLM Results for Striving learners C-1

Table C.2. Grade 3 Mathematics HLM Results for Striving learners C-1

Table C.3. Grade 4 Mathematics HLM Results for Striving learners C-2

Table C.4. Grade 5 Mathematics HLM Results for Striving learners C-2

Table D.1. Grade 2 Mathematics HLM Results for Striving learners at the Bottom 20th Percentile D-1

Table D.2. Grade 3 Mathematics HLM Results for Striving learners at the Bottom 20th Percentile D-1

Table D.3. Grade 4 Mathematics HLM Results for Striving learners at the Bottom 20th Percentile D-2

Table D.4. Grade 5 Mathematics HLM Results for Striving learners at the Bottom 20th Percentile D-2

Table of Contents (Continued)

List of Tables

Table E.1. HLM Results for Grade 2 Mathematics with Black or African American by Treatment Interaction.....E-1
Table E.2. HLM Results for Grade 3 Mathematics with Black or African American by Treatment Interaction.....E-1
Table E.3. HLM Results for Grade 4 Mathematics with Black or African American by Treatment Interaction.....E-2
Table E.4. HLM Results for Grade 5 Mathematics with Black or African American by Treatment Interaction.....E-2

Table F.1. Differences between Treatment and Comparison Black or African American groups, by Grade.....F-1
Table F.2. Baseline to Outcome Change in Mathematics Diagnostic Performance for Black or African American and Non-Black or African American Striving Learners in i-Ready (Treatment) Schools Compared to Comparison Schools, by Grade. F-2

Table G.1. HLM Results for Grade 2 Mathematics with Hispanic Origin by Treatment Interaction.....G-1
Table G.2. HLM Results for Grade 3 Mathematics with Hispanic Origin by Treatment Interaction.....G-1
Table G.3. HLM Results for Grade 4 Mathematics with Hispanic Origin by Treatment Interaction.....G-2
Table G.4. HLM Results for Grade 5 Mathematics with Hispanic Origin by Treatment Interaction.....G-2

Table H.1. Differences between i-Ready (Treatment) and Comparison Hispanic Origin groups, by Grade.....H-1

Table I.2. Baseline to Outcome Change in Mathematics i-Ready Diagnostic Performance for Hispanic Origin and Non-Hispanic Origin Students in Treatment Schools Compared to Comparison Schools, by grade..... I-2

List of Figures

Figure 1. Gains in mathematics Diagnostic achievement between baseline and outcome for the striving learner treatment and comparison groups at grade 2.....12
Figure 2. Gains in mathematics Diagnostic achievement between baseline and outcome for the striving learner treatment and comparison groups at grade 3.....12
Figure 3. Gains in mathematics Diagnostic achievement between baseline and outcome for the striving learner treatment and comparison groups at grade 4.....13
Figure 4. Gains in mathematics Diagnostic achievement between baseline and outcome for the striving learner treatment and comparison groups at grade 5.....13

Table of Contents (Continued)

List of Figures

Figure 5. Gains in mathematics Diagnostic achievement between baseline and outcome for the striving learner treatment and comparison groups at grade 2.....17
Figure 6. Gains in mathematics Diagnostic achievement between baseline and outcome for the striving learner treatment and comparison groups at grade 3.....17
Figure 7. Gains in mathematics Diagnostic achievement between baseline and outcome for the striving learner treatment and comparison groups at grade 4.....18
Figure 8. Gains in mathematics Diagnostic achievement between baseline and outcome for the striving learner treatment and comparison groups at grade 5.....18

Figure F.1. Gains in Mathematics Diagnostic achievement between baseline and outcome for the i-Ready (treatment) and comparison groups for Black or African American and non-Black or African American groups at grade 2. ..... F-3
Figure F.2. Gains in Mathematics Diagnostic achievement between baseline and outcome for the i-Ready (treatment) and comparison groups for Black or African American and non-Black or African American groups at grade 3. ..... F-3
Figure F.3. Gains in Mathematics Diagnostic achievement between baseline and outcome for the i-Ready (treatment) and comparison groups for Black or African American and non-Black or African American groups at grade 4. ..... F-4
Figure F.4. Gains in Mathematics Diagnostic achievement between baseline and outcome for the i-Ready (treatment) and comparison groups for Black or African American and non-Black or African American groups at grade 5. ..... F-4

Figure H.1. Gains in Mathematics Diagnostic achievement between baseline and outcome for the i-Ready (treatment) and comparison groups for Hispanic Origin and non-Hispanic Origin groups at grade 2. .....H-3
Figure H.2. Gains in Mathematics Diagnostic achievement between baseline and outcome for the i-Ready (treatment) and comparison groups for Hispanic Origin and non-Hispanic Origin groups at grade 3. .....H-3
Figure H.3. Gains in Mathematics Diagnostic achievement between baseline and outcome for the i-Ready (treatment) and comparison groups for Hispanic Origin and non-Hispanic Origin students at grade 4.....H-4
Figure H.4. Gains in Mathematics Diagnostic achievement between baseline and outcome for the i-Ready (treatment) and comparison groups for Hispanic Origin and non-Hispanic Origin groups at grade 5. .....H-4

Impact Evaluation of Mathematics *i-Ready* for Striving Learners Using 2018–19 Data

Key Findings

- Striving learners that used *i-Ready* in treatment schools performed better on mathematics achievement than similar students in comparison schools who did not use *i-Ready*.
  - On average, all striving learners showed gains in mathematics achievement between fall of 2018 and spring of 2019; moreover, those who used *i-Ready* showed significantly greater gains in student achievement.

- A subset of striving learners, defined as those who performed at or below the 20th percentile of mathematics achievement at baseline, that used *i-Ready* in treatment schools performed better on mathematics achievement than similar students in comparison schools striving learners in comparison schools who did not use i-*Ready*.
  - On average, students who performed at or below the 20th percentile of mathematics achievement showed gains in mathematics achievement between fall of 2018 and spring of 2019; moreover, those who used *i-Ready* showed significantly greater gains in student achievement.

- Black or African American striving learners that used *i-Ready* in treatment schools performed better on mathematics achievement than Black or African American striving learners who did not use *i-Ready* in comparison schools.
  - Black or African American striving learners experienced similar benefits from *i-Ready* use as non-Black or African American striving learners.
  - On average, all Black or African American striving learners showed gains in mathematics achievement between fall of 2018 and spring of 2019; moreover, those who used *i-Ready* showed significantly greater gains in student achievement.

- Striving learners of Hispanic origin that used *i-Ready* in treatment schools performed better on mathematics achievement than striving learners of Hispanic origin in comparison schools who did not use *i-Ready*.
  - Striving learners of Hispanic origin in grades 2, 3, and 5 experienced similar benefits from *i-Ready* use as striving learners not of Hispanic origin.
  - Striving learners of Hispanic origin in grade 4 showed even greater benefits of *i-Ready* use on mathematics achievement compared to striving learners not of Hispanic origin.
  - On average, all striving learners of Hispanic origin showed gains in mathematics achievement between fall of 2018 and spring of 2019; moreover, those who used *i-Ready* showed significantly greater gains in student achievement. At grade 4, striving learners of Hispanic origin who used *i-Ready* showed even greater gains in mathematics achievement than the students not of Hispanic origin who used *i-Ready*.

Abstract

Curriculum Associates' *i-Ready® Personalized Instruction (i-Ready)* is a supplemental, online personalized instruction program available for mathematics and reading[1]. Prior research has indicated *i-Ready* has a positive impact on student achievement for students overall (e.g., Swain, Randel, & Norman Dvorak, 2020). The present study furthers that work by examining the impacts of *i-Ready* for striving learners specifically, to provide schools and districts with more targeted information on its effectiveness for these struggling students. The Human Resources Research Organization (HumRRO), in collaboration with Century Analytics, implemented a quasi-experimental design (QED) using academic year 2018–19 *i-Ready* data to evaluate the impact of *i-Ready* mathematics Instruction on student mathematics achievement for striving learners in grades 2–5 on a nationally normed cognitive assessment. Two populations of striving learners were examined at each grade – those who tested two or more grade levels below their current grade in mathematics at baseline and a subset of these students who fell at the bottom 20th percentile of mathematics achievement. The percentiles were based on mathematics achievement measured by the *i-Ready® Diagnostic* (Diagnostic) at baseline. It was hypothesized student achievement, as measured by the Diagnostic, would be higher for striving learners using *i-Ready* for mathematics over comparison groups of students who did not use this instruction. Exploratory analyses examined whether the findings were consistent for Black or African American striving learners and those of Hispanic origin. Matching was conducted at each grade level to meet two needs: 1) identify a set of comparison schools demographically similar to our *i-Ready* schools, and 2) identify a set of academically equivalent comparison students within the matched comparison schools. Students who received *i-Ready* and students in the comparison group took the mathematics version of the Diagnostic assessment. To estimate impacts, hierarchical-linear modeling (HLM) was conducted separately for each grade level with students at level 1 and schools at level 2. This process was conducted for the full sample of striving learners and again for the subsample of students at the bottom 20th percentile. Results suggest both the striving learners and students at the bottom 20th percentile using *i-Ready* with fidelity in the treatment schools performed statistically significantly better on mathematics than students in the comparison schools who did not use this instruction. The effect sizes for striving learners and the subset of the bottom 20th percentile students at grades 3 and 5 fell within the range which recent research characterizes as modest for an education intervention (Kraft, 2019). The effect sizes for striving learners and the bottom 20th percentile students at grades 2 and 4 fell above this range. These findings provide support that *i-Ready* for mathematics used with fidelity in schools can lead to higher mathematics achievement for striving learners. Exploratory analyses found that these impacts were consistent for the Black or African American striving learners and for the striving learners of Hispanic origin at grades 2, 3, and 5. A positive Hispanic origin by treatment group interaction was present at grade 4, indicating *i-Ready* had greater impacts on mathematics achievement of striving learners of Hispanic origin as compared to striving learners not of Hispanic origin who used *i-Ready*.

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

Introduction

For more than 50 years, Curriculum Associates has provided educational products and services with the goal of improving education for students and teachers.

They provide various assessment and instructional resources and professional development for reading and mathematics. One available product is the *i-Ready® Diagnostic* (Diagnostic), available for grades K–12. The Diagnostic assessments, typically taken in the fall, winter, and spring of a given academic year, are (a) 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 Diagnostic assessments are independent measures often used by educators as interim classroom benchmark assessments. Another product is *i-Ready® Personalized Instruction* (*i-Ready*), available for grades K–8. *i-Ready* is personalized instruction included with Curriculum Associate's *i-Ready Learning* products. The instruction provided to students is driven by performance on the Diagnostic and provides tailored instruction that meets students' needs and encourages the development of new skills.

i-Ready is intended for students of all ability levels. Previous research provides evidence of its effectiveness in reading and mathematics when considering K–8 students overall (Swain et. al., 2020). However, Curriculum Associates understands many schools are interested in education programs that are proven to be effective with select groups of students, including striving learners and students from traditionally disadvantaged backgrounds. Identifying successful online learning options for struggling students is particularly relevant in the age of virtual learning, as schools develop virtual options that may need to be implemented for the 2020–2021 school year and beyond. The primary purpose of this study was to examine the impact of *i-Ready* on mathematics achievement for striving learners in elementary grades 2–5 using 2018–19 data. Because achievement gaps in mathematics are often prevalent for Black or African American students and students of Hispanic origin (Stanford CEPA, n.d.), a secondary purpose was to examine if *i-Ready* had a differential impact on Black or African American striving learners or striving learners of Hispanic origin.

The research was conducted by the Human Resources Research Organization (HumRRO) and Century Analytics. 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. HumRRO and Century Analytics designed the study to meet the required rigor of the WWC 4.1 standards to achieve a rating of *Meets WWC Group Design Standards with Reservations (*WWC, 2020a), 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.

Defining i-Ready Implementation

The impact of *i-Ready* on student achievement was the focus of this evaluation. *i-Ready* is an online personalized instruction program aligned to college- and career-readiness standards that includes engaging multimedia instruction and progress monitoring of online lessons. Lessons are intended to provide a consistent best practice lesson structure and build students' conceptual understanding. *i-Ready* is intended to be used in conjunction with the Diagnostic

which monitors student progress and identifies student performance in mathematics and reading. This diagnostic information helps target student-specific intervention, which can be provided through *i-Ready.*

Curriculum Associates has identified key implementation components of *i-Ready* that 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* to work through a personalized, scaffolded instruction path.

Curriculum Associates provides guidance to districts and schools on how to implement *i-Ready* to best benefit student learning (Curriculum Associates, 2019). Guidance indicates students achieve greater gains when using *i-Ready* for an average of between 30–49 minutes of lesson time-on-task per week, per subject area. In addition, Curriculum Associates recommends use for at least 12 to 18 calendar weeks between administrations of the Diagnostic (Curriculum Associates, 2018).

Research Questions

The primary purpose of this study was to estimate the impact of Curriculum Associates' *i-Ready* on student mathematics achievement for striving learners in grades 2–5. Striving learners were defined as those who tested two or more grade levels below their current grade at baseline. The following confirmatory research question was addressed:

- What is the impact of *i-Ready* usage on student mathematics achievement for striving learners in schools that implement *i-Ready* compared to striving learners in schools that implement the Diagnostic only?

In addition, a second research question sought to examine the impact of *i-Ready* on student achievement for a subset of grade 2–5 striving learners that fell in the bottom 20th percentile of mathematics achievement. The following second confirmatory research question was addressed:

- What is the impact of *i-Ready* usage on student mathematics achievement for striving learners at the bottom 20th percentile in schools that implement *i-Ready* compared to these striving learners at schools that implement the Diagnostic only?

In addition to the main research questions, we sought to understand whether the main effects were representative of the experiences with *i-Ready* for Black or African American striving learners and striving learners of Hispanic origin. We addressed the following exploratory questions:

- Do Black or African American striving learners experience similar impacts of *i-Ready* use on student mathematics achievement compared to striving learners overall?
- Do striving learners of Hispanic origin experience similar impacts of *i-Ready* use on student mathematics achievement compared to striving learners overall?

Methodology

In this section, we describe the methodology for conducting our impact analyses. We begin with initial design decisions. We then discuss the matching process to achieve baseline equivalence

and the analytic model. In the subsequent section, we discuss our impact and exploratory analysis results.

Design

Eligible Schools and Students

For each grade, we started with a student-level *i-Ready* usage file of students in public schools with Diagnostic and *i-Ready* use in 2018–19 who had at minimum fall and spring Diagnostic scores. By including only public schools, we sought to include only students in a relatively traditional school environment with expectations to follow state adopted college and career ready standards.

For a student within a treatment school to be eligible for inclusion, they must have used *i-Ready* 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* usage at the student-level for attaining intended goals of improved student mathematics achievement. Treatment schools were only included if they began using *i-Ready* to some extent prior to the 2018–19 school year. This requirement is based on the understanding that *i-Ready* implementation, like the implementation of most new programs, requires a start-up time to learn the technology and adjust to the schedule before *i-Ready* is fully implemented. To be eligible for inclusion as a student in a comparison school, students must not have used any *i-Ready* for mathematics in 2018–19. We removed from the datafile students not meeting the treatment or comparison eligibility requirements when matching students to the two groups.

Prior to the onset of this study, we defined a striving learner as one who tested two or more grade levels below their current grade at baseline. Each student is assigned a grade-classification based on their Diagnostic score. Only students assigned a classification of two or more grade levels below their current grade were included in our study. For example, a grade 2 student was included only if they classified at a kindergarten level and a grade 5 student was included only if they classified as levels K–3. We also identified students at the lowest 20th percentile of mathematics achievement at baseline as a subset of striving students to examine. These students were of interest as they may require intensive academic intervention. Only students who met these definitions were included in our study.

Unit of Assignment

HumRRO and Century Analytics completed investigations to identify the unit of assignment—either school-level or student-level—for the sample of striving learners. Because we understand there are differences by grade-level, we conducted these investigations separately by grade. Using Curriculum Associates usage data, we identified the number of schools for which there were students in (a) only the treatment or the comparison group and (b) both the treatment and comparison groups. Across grades, 94.5% of students attended schools with students classified as only treatment or comparison; thus, we decided that school was the appropriate unit of assignment for investigating the impact of *i-Ready* on the achievement of striving learners. Separately by grade, we excluded the small percentage of schools with some students classified as treatment and other students classified as comparison from our school-level assignment study.

Baseline and Outcome Measure

We selected the Diagnostic as both the baseline and outcome measure for all students participating in this study (i.e., *i-Ready* students and comparison group students). The

Diagnostic for mathematics measures achievement aligned to common mathematics content and skills with demonstrated test score reliability. Marginal reliabilities range from 0.94 to 0.96 and test-retest reliabilities range from 0.81 to 0.86 for mathematics grades 2 through grade 5. Therefore, this assessment meets the WWC 4.1 standards for an acceptable baseline and outcome measure (WWC, 2020a).

The 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* and by others as a stand-alone diagnostic assessment without the use of *i-Ready*. The Diagnostic assessments for mathematics and reading are currently used by approximately eight million, or nearly 25%, of K–8 students across the United States. Thus, the use of Diagnostic as the outcome measure allowed us to include a large sample of students from across the United States. The Diagnostic is intended to be administered in a standardized manner across schools (Curriculum Associates, 2019b). Specifically, teachers of students in the studied grades 2–5 are to schedule the first (fall) Diagnostic assessment 2–3 weeks into the school year in two 45- to 50-minute sessions. Curriculum Associates recommends three administrations over the course of the school year, with 12–18 weeks between each Diagnostic administration. 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 Diagnostic as well as its consistency with education standards used across the United States. Since being released in summer 2011, the Diagnostic has been reviewed and approved at the national and state level as an assessment, instructional resource, or intervention in Alabama, Arizona, Arkansas, California, Colorado, Connecticut, Delaware, Florida, Georgia, Idaho, Indiana, Iowa, Louisiana, Massachusetts, Michigan, Mississippi, Nebraska, Nevada, New Mexico, New York, North Carolina, Ohio, Oklahoma, Oregon, Tennessee, Texas, Utah, and Virginia.

Between 2017 and 2019, Curriculum Associates conducted linking studies examining the relationship between the Diagnostic and 19 state accountability tests, the Partnership for Assessment of Readiness for College and Careers (PARCC) test, and the Smarter Balanced Assessment (SBA at grades 3–8. These studies provide evidence the Diagnostic measures skills consistent with student expectations and can be used as a student mathematics achievement measure. These studies show strong correlations between Diagnostic scores and scores on these national and state tests. The average correlations across grades between these tests and the Diagnostic for mathematics ranged from 0.80 (Texas Assessments of Academic Readiness) to 0.89 (Smarter Balanced Assessment). These findings support that the Diagnostic content is highly consistent with what students across the United States are expected to learn (Curriculum Associates, 2020).

Required Number of Schools

We conducted power analyses using PowerUp! (Dong & Maynard, 2013) to identify the minimum detectable effect size (MDES) needed to reject the null hypothesis that there is no difference in mathematics achievement between the treatment and comparison group. Statistical power is influenced by various factors. We used data from previous studies HumRRO conducted using the Diagnostic as an outcome to estimate conservative and optimistic parameters for use in the power analysis. These parameters were: (a) approximately 1,000

schools available for the analyses per grade, (b) an average of six striving learners eligible for inclusion at each school and grade, and (c) 0.10 and 0.30 for the intraclass correlation coefficient (ICC). Results of the power analyses indicated an MDES of between 0.06 and 0.08 with 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. The available schools for our analyses across all grades far exceeded the minimum.

Analytic Samples

We used a multi-step process to identify analytic samples separately at each grade to address the confirmatory research question. First, we conducted school-level matching to identify a set of *i-Ready* (treatment) and a set of comparison schools from which to match students. Next, we conducted student-level matching with students within the selected matched schools to identify a set of *i-Ready* students and comparison students equivalent on mathematics achievement, as measured by the Diagnostic. We computed effect sizes for all school- and student-level matching variables following matching and found baseline equivalence was achieved according to WWC standards (WWC, 2020a). See Appendix A for details of our matching process and final school and student samples.

Analytic Models

We used hierarchical linear modeling (HLM) to estimate impacts of *i-Ready* on student achievement for the samples of striving learners to address our confirmatory research questions. Similarly, we used HLM for our samples of Black or African American and Hispanic origin striving learners to address our exploratory research questions. For each analysis, we chose a two-level model with level 1 as the student and level 2 as the school. The analytic model acted as the basis for our models to estimate baseline differences. This section describes the analytic models used for impact analyses and for baseline equivalence.

Benchmark Impact Model

We developed a benchmark impact model to address our confirmatory research questions. We used hierarchical linear modeling (HLM) to estimate the impact of *i-Ready* on student achievement. For level 2 of our model, we included an indicator variable of group membership and school-level variables that were publicly available and known to be related to achievement. For level 1 we included baseline Diagnostic performance.

The student-level covariate used in each analysis was:

- Diagnostic mathematics baseline performance

The school-level covariates included:

- Group membership (0 = comparison, 1 = *i-Ready*)
- Urbanicity
- Percent of students eligible for free and reduced-price lunch (FRL)
- Percent students of historically marginalized races (HMR)
- Grade-level enrollment

For additional model details and information on sensitivity analyses to examine the robustness of the benchmark impact model, see Appendix B.

Exploratory Models

We analyzed two additional impact models to address the exploratory research questions focused on striving learners classified as Black or African American and Hispanic origin. For these models, an interaction term was added at Level 1. We added a Black or African American by treatment interaction to address our research question, *Do Black or African American striving learners experience similar impacts of i-Ready use on student mathematics achievement compared to striving learners overall?* We added a Hispanic origin by treatment interaction to address our research question, *Do striving learners of Hispanic origin experience similar impacts of i-Ready use on student mathematics achievement compared to striving learners overall?* Level 2 for both models was specified consistent with our baseline model. See Appendix B for additional model details.

Baseline Difference Model

We used a baseline difference model to provide a model-based estimate of the difference between students in the treatment and comparison groups on the baseline (fall Diagnostic) score separately for each grade level. This model is described in Appendix B.

Results

Benchmark Impact Analysis

Striving learners

Table 1 contains the benchmark impact model results for the samples of striving learners by grade for mathematics spring Diagnostic scores. Full results of the HLM model are available at Appendix C, with a discussion of model assumption checks presented in Appendix I. For all grade levels, the adjusted mean differences were positive and statistically significant, indicating the *i-Ready* group earned higher mathematics scores than the comparison group. Hedge's *g* effect sizes ranged from 0.13 to 0.22. Recent research by Kraft (2019) notes traditional guidelines are often too rigid for the realities of education evaluations designed to meet the rigor required by the U.S. Department of Education, including those developed in accordance to WWC standards. He specifies effect size ranges of 0.03 – 0.17 as typical of education interventions and often represent a meaningful effect. All effect sizes fall at the upper end of this range or above it. Based on Kraft's findings, we consider the mathematics effect sizes modest (grades 3 and 5) to strong (grades 2 and 4) for an education intervention.

We computed the improvement index, as defined by the *WWC Procedures Handbook* (WWC, 2020a), as an additional measure of impact. The improvement indices range between 5.17 (grade 5) and 8.71 (grade 2). Improvement indices show the expected change in percentile rank for an average comparison student if they had been in the intervention group. For example, an improvement index of 8.71 is equivalent to a comparison group student improving from the 50th percentile to better than the 58th percentile if they were to have participated in the treatment.

Table 1 also provides the intraclass correlations. The ICCs measure the proportion of the variance that is between schools—that is, how much of the variance in mathematics Diagnostic scores can be explained by school-level differences. The ICCs range from 0.11 (grade 5) to

0.15 (grade 2). This suggests the majority of variance is due to factors other than school-level differences.

Table 1. Impact Analysis Results for Striving Learners of i-Ready (Treatment) Schools Compared to Comparison Schools for Mathematics Student Achievement at Grades 2–5

Grade	Group	Schools	Students	Diagnostic Mean	Diagnostic SD	ICC	Adj Mean Diff (SE)	*p*-value	Effect Size	Improvement Index
2	*i-Ready*	1,356	11,673	408.14	22.55	0.15	4.97 (0.39)	<.0001	0.22	8.71
	Comparison	1,330	11,673	403.17	22.02					
3	*i-Ready*	1,410	14,679	430.35	24.79	0.15	4.08 (0.39)	<.0001	0.17	6.75
	Comparison	1,376	14,679	426.27	24.62					
4	*i-Ready*	1,475	12,757	443.78	27.21	0.14	5.46 (0.39)	<.0001	0.20	7.93
	Comparison	1,425	12,757	438.31	26.48					
5	*i-Ready*	1,500	12,566	451.73	27.69	0.11	3.64 (0.38)	<.0001	0.13	5.17
	Comparison	1,438	12,566	448.09	27.01					

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

We also provide the gains in mathematics achievement on the Diagnostic between baseline and outcome for the treatment and comparison groups as supplemental information to aid in interpreting the impacts presented above. Table 2 presents the mean baseline scores, outcome scores, and the gains between these two periods for our *i-Ready* and comparison striving learner groups at each grade.

Table 2. Baseline to Outcome Change in Mathematics Diagnostic Performance for Striving Learners of i-Ready (Treatment) Schools Compared to Striving Learners of Comparison Schools at Grades 2–5

Grade	Group	Schools	Students	Diagnostic Baseline Mean	Diagnostic Outcome Mean	*Baseline to Outcome Gain*
2	*i-Ready*	1,356	11,673	371.81	408.14	36.33
	Comparison	1,330	11,673	371.75	403.17	31.42
3	*i-Ready*	1,410	14,679	395.59	430.35	34.76
	Comparison	1,376	14,679	395.41	426.27	30.86
4	*i-Ready*	1,475	12,757	412.50	443.78	31.28
	Comparison	1,425	12,757	412.75	438.31	25.56
5	*i-Ready*	1,500	12,566	427.51	451.73	24.22
	Comparison	1,438	12,566	427.77	448.09	20.32

As shown in Figures 1–4 below, for each grade, the two striving learner groups start with very similar baseline means. Both groups show gains in achievement, however the treatment group gains are greater than the comparison group gains. While the gain scores presented in Table 2 and Figures 1–4 provide a reasonable approximation of achievement gains, caution is warranted when interpreting them. Although the gain scores were calculated by subtracting the baseline mean from the outcome mean, the difference between the gain scores of the two study groups does not provide an accurate or model-based estimate of the impact because they do not adjust for covariates. Please refer to Table 1 above for the impact estimates.

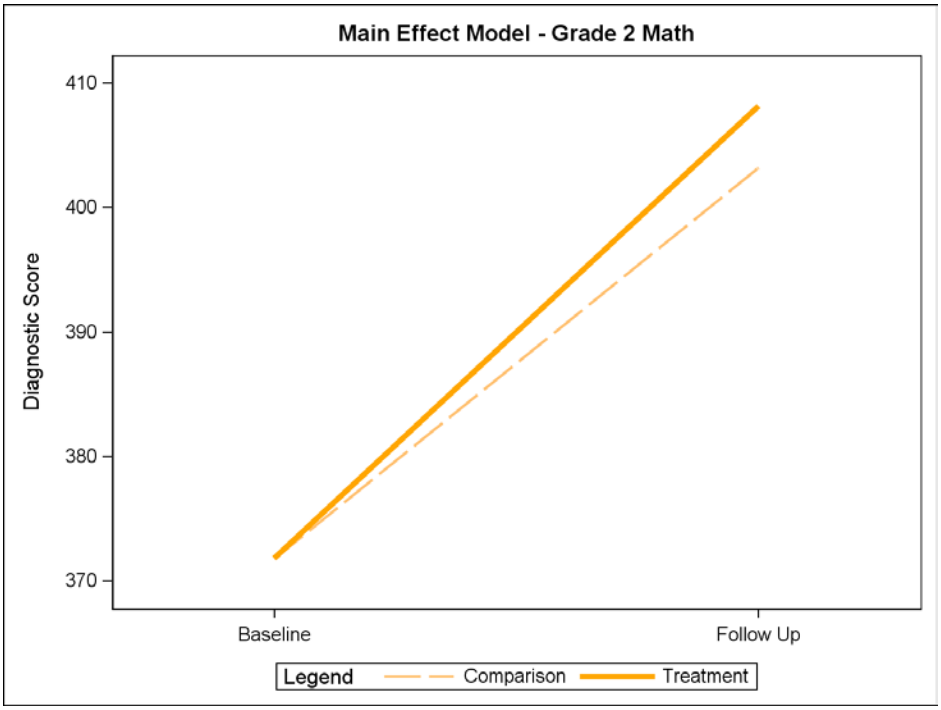

Figure 1. Gains in mathematics Diagnostic achievement between baseline and outcome for the striving learner treatment and comparison groups at grade 2.

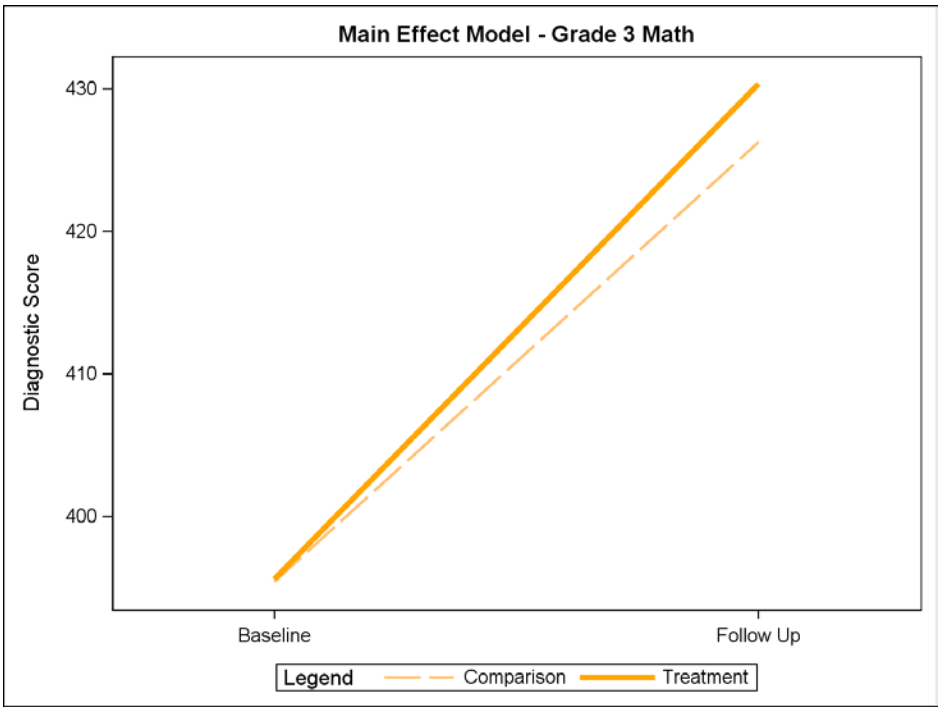

Figure 2. Gains in mathematics Diagnostic achievement between baseline and outcome for the striving learner treatment and comparison groups at grade 3.

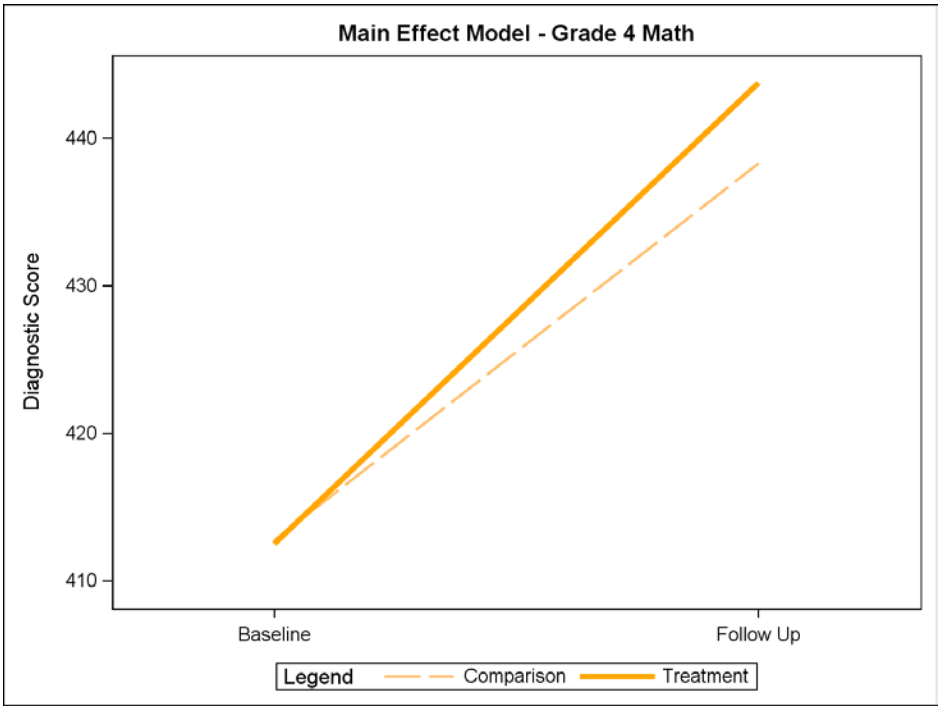

Figure 3. Gains in mathematics Diagnostic achievement between baseline and outcome for the striving learner treatment and comparison groups at grade 4.

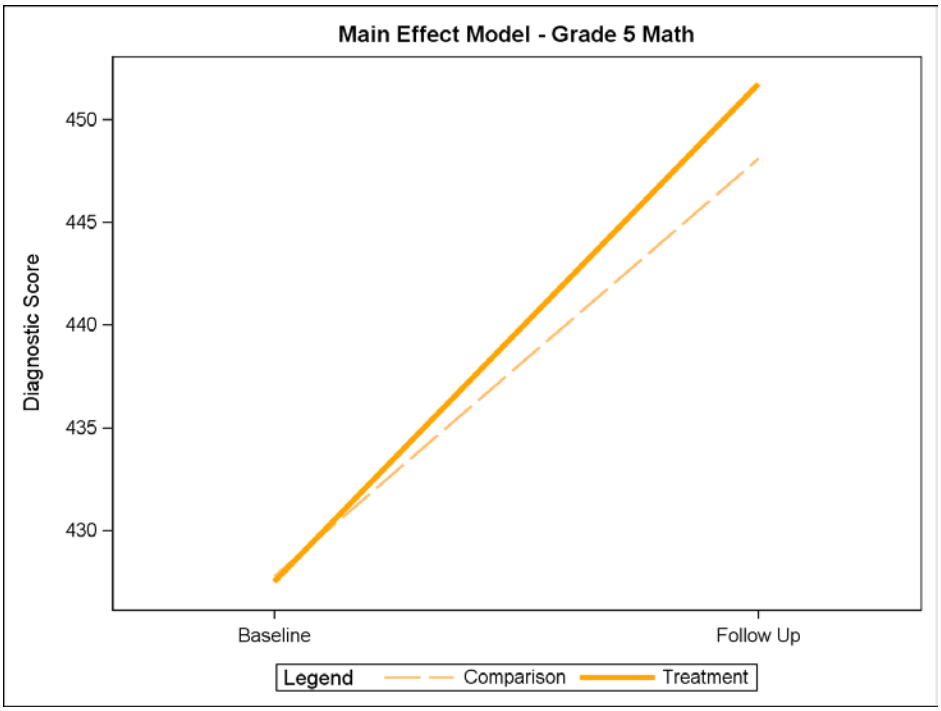

Figure 4. Gains in mathematics Diagnostic achievement between baseline and outcome for the striving learner treatment and comparison groups at grade 5.

Bottom 20th Percentile Striving Learners

Table 3 contains the benchmark impact model results for the samples of striving learners at the bottom 20th percentile by grade for mathematics spring Diagnostic scores. Full results of the HLM model are available at Appendix D. The findings were similar to those for the full group of striving learners – the adjusted mean differences were positive and statistically significant at all grade levels, indicating the *i-Ready* group earned higher mathematics scores than the comparison group. Based on Kraft's (2019) research, we consider the effect sizes for grades 3 and 5 modest, and those for grades 2 and 4 strong. These effect sizes are comparable to those for the striving learners.

The improvement indices for the analyses examining the impact of *i-Ready* on the lowest performing students range between 5.17 (grade 5) and 8.71 (grade 1). Improvement indices show the expected change in percentile rank for an average comparison student in this study if they had been in the intervention group. For example, an improvement index of 8.71 is equivalent to a student at a comparison school improving from the 50th percentile to better than the 58th percentile if they were to have participated in the treatment.

The ICCs for the lowest performing student impact analyses range from 0.12 (grade 5) to 0.15 (grade 3). This suggests the majority of variance is due to factors other than school-level differences.

Table 3. Impact Analysis Results for Striving Learners at the Bottom 20th Percentile of i-Ready (Treatment) Schools Compared to these Striving Learners of Comparison Schools for Mathematics Student Achievement at Grades 2–5

Grade	Group	Schools	Students	Diagnostic Mean	Diagnostic SD	ICC	Adj Mean Diff (SE)	*p*-value	Effect Size	Improvement Index
2	*i-Ready*	1,309	9,114	404.93	22.52	0.14	4.91 (0.42)	<.0001	0.22	8.71
	Comparison	1,283	9,121	400.02	22.12					
3	*i-Ready*	1,314	9,606	423.53	25.27	0.15	4.43 (0.46)	<.0001	0.18	7.14
	Comparison	1,295	9,598	419.10	24.80					
4	*i-Ready*	1,427	10,068	439.43	27.58	0.14	5.64 (0.44)	<.0001	0.21	8.32
	Comparison	1,351	10,031	433.79	26.57					
5	*i-Ready*	1,447	9,956	446.97	27.88	0.12	3.54 (0.42)	<.0001	0.13	5.17
	Comparison	1,372	9,969	443.43	27.03					

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

Table 4 and Figures 5—8 below present the gains in student achievement for the treatment and comparison groups of the bottom 20th percentile striving learners. Both groups show gains in achievement, however the treatment group gains are greater than the comparison group gains. Although the gain scores presented in Table 4 and Figures 5—8 provide a reasonable approximation of achievement gains, caution is warranted when interpreting them. The gain scores were calculated by subtracting the baseline mean from the outcome mean; however, the difference between the gain scores of the two study groups does not provide an accurate or model-based estimate of the impact because they do not adjust for covariates. Please refer to Table 3 for the impact estimates.

Table 4. Baseline to Outcome Change in Mathematics Diagnostic Performance for Students in i-Ready (Treatment) Schools Compared to Comparison Schools at Grades 2–5

Grade	Group	Schools	Students	Diagnostic Baseline Mean	Diagnostic Outcome Mean	*Baseline to Outcome Gain*
2	*i-Ready*	1,309	9,114	368.19	404.93	36.74
	Comparison	1,283	9,121	368.17	400.02	31.85
3	*i-Ready*	1,314	9,606	388.45	423.53	35.08
	Comparison	1,295	9,598	388.26	419.10	30.84
4	*i-Ready*	1,427	10,068	407.69	439.43	31.74
	Comparison	1,351	10,031	407.89	433.79	25.90
5	*i-Ready*	1,447	9,956	422.46	446.97	24.51
	Comparison	1,372	9,969	422.77	443.43	20.66

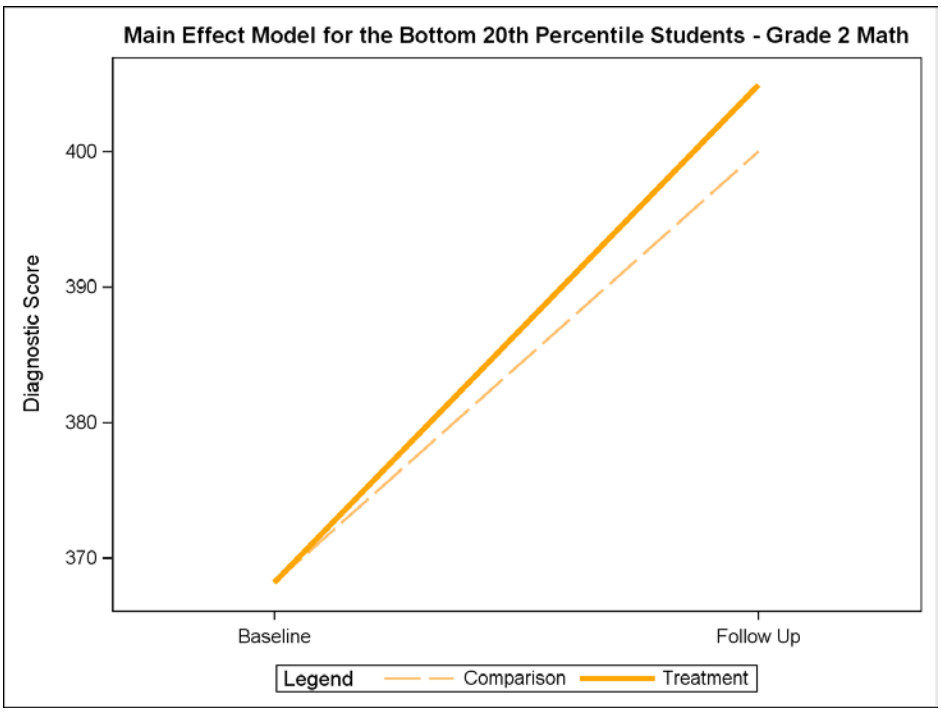

Figure 5. Gains in mathematics Diagnostic achievement between baseline and outcome for the bottom 20th percentile treatment and comparison groups at grade 2.

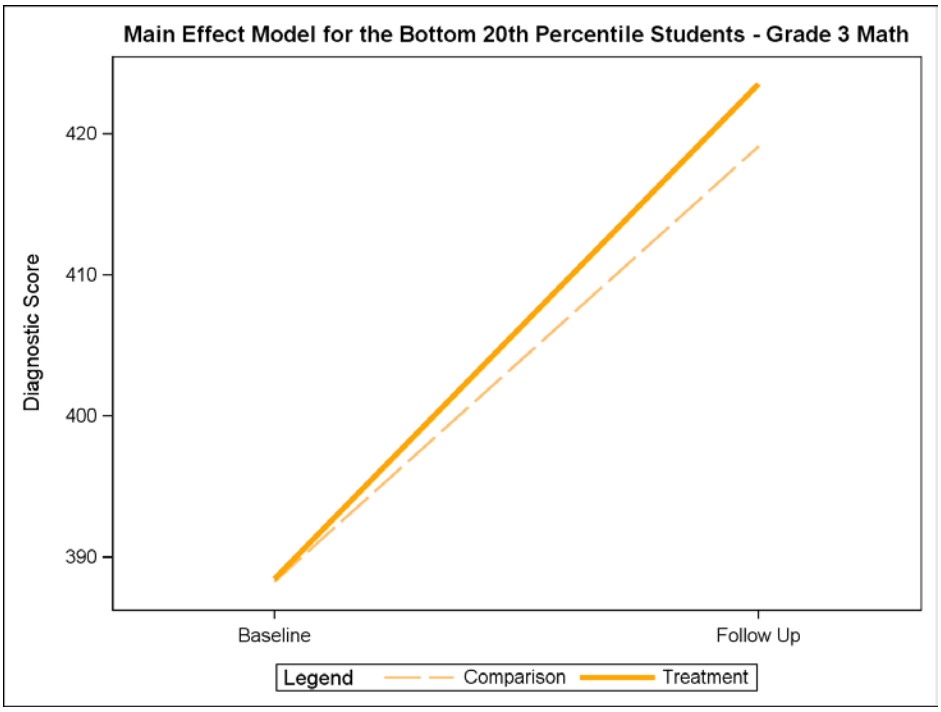

Figure 6. Gains in mathematics Diagnostic achievement between baseline and outcome for the bottom 20th percentile treatment and comparison groups at grade 3.

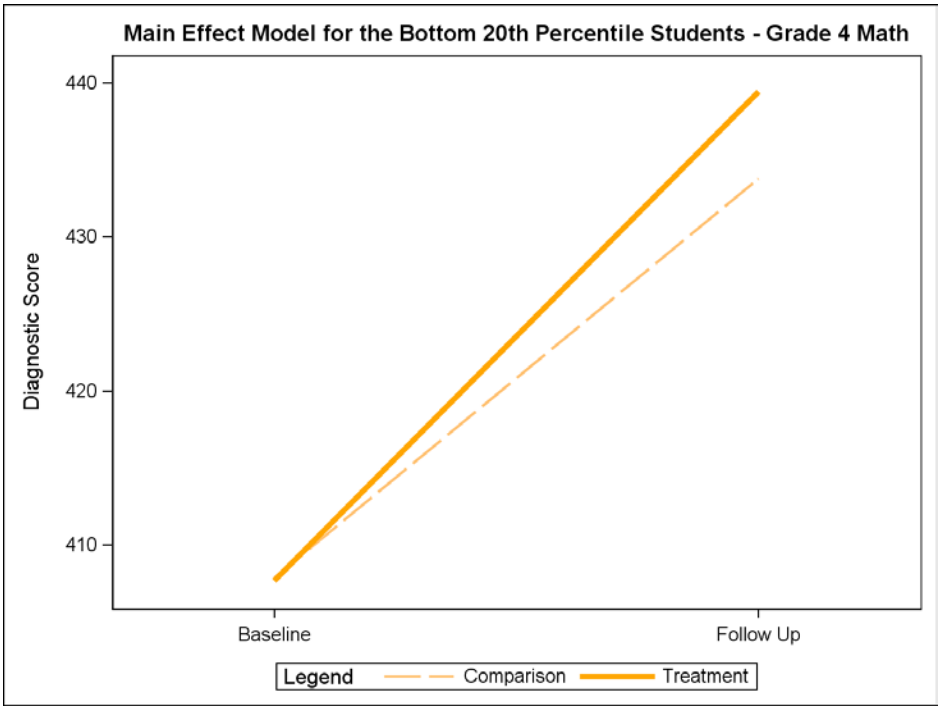

Figure 7. Gains in mathematics Diagnostic achievement between baseline and outcome for the bottom 20th percentile treatment and comparison groups at grade 4.

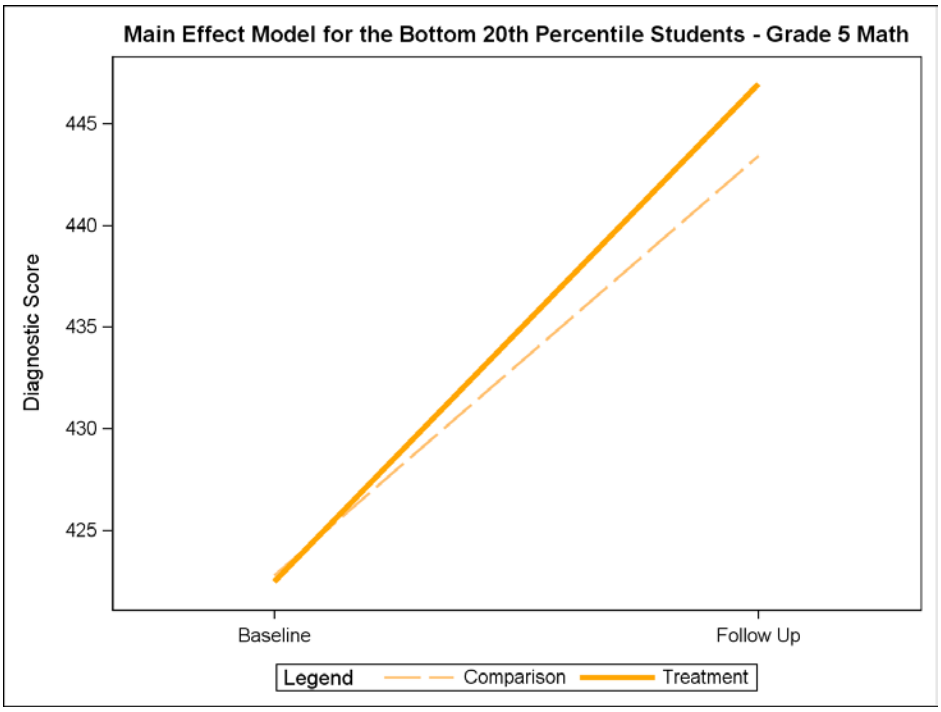

Figure 8. Gains in mathematics Diagnostic achievement between baseline and outcome for the bottom 20th percentile treatment and comparison groups at grade 5.

Exploratory Analyses

This section describes the findings of the analyses we conducted to address our two exploratory research questions. We begin with the findings pertaining to Black or African American students, followed by the findings for students of Hispanic origin.

Black or African American by Treatment Interactions

Table 5 presents the impact model results for striving learners with a Black or African American by treatment interaction added. The table includes the number of students and schools for which there were complete student-level race data for inclusion in the analysis. Full HLM results are presented in Appendix E. For all grade levels, interaction terms were not found to be statistically significant. Thus, we conclude that Black or African American striving learners see similar positive impacts of *i-Ready* on mathematics achievement as striving learners overall. In other words, Black or African American striving learners who used *i-Ready* performed better than Black or African American students in a comparison group. One should refer to the benchmark analysis results to examine the expected impact, including the effect sizes and improvement indices, for Black or African American striving learners. Additional details of the Black or African American by treatment interaction analyses, including score differences, are presented in Appendix F.

Table 5. Summary of Black or African American by Treatment Interactions, by Grade

Grade	Group	Schools	*Students*	*Interaction Coefficient*	*p-value*
2	*i-Ready*	953	8,256	-0.31	0.721
	Comparison	683	5,612		
3	*i-Ready*	990	10,377	0.56	0.475
	Comparison	698	7,329		
4	*i-Ready*	1,016	8,860	0.67	0.442
	Comparison	708	6,384		
5	*i-Ready*	1,034	8,942	0.29	0.733
	Comparison	739	6,725		

Hispanic Origin by Treatment Interactions

Table 6 contains the impact model results for striving learners with a Hispanic origin by treatment interaction added. The table includes the number of students and schools for which there were complete student-level race data for inclusion in the analysis. Full HLM results are presented in Appendix G. The interactions were not statistically significant at grades 2, 3, and 5. For these grade levels we conclude that striving learners of Hispanic origin see similar positive impacts of *i-Ready* on mathematics achievement as striving learners overall. In other words, students of Hispanic origin who used *i-Ready* performed better than students of Hispanic origin in a comparison group. One should refer to the benchmark analysis results to examine the impact one should expect for striving learners of Hispanic origin at these grades. A statistically significant positive interaction was found at grade 4, indicating grade 4 striving learners of Hispanic origin benefitted more from *i-Ready* than similar students not of Hispanic origin. Additional details of the Hispanic origin by treatment interaction analyses, including score differences, are presented in Appendix H.

Table 6. Summary of Hispanic Origin by Treatment Interactions by Grade

Grade	Group	Schools	*Students*	*Interaction Coefficient*	*p-value*
2	*i-Ready*	1,013	9,122	0.99	0.162
	Comparison	748	6,775		
3	*i-Ready*	1,040	11,505	1.17	0.069
	Comparison	750	8,804		
4	*i-Ready*	1,078	9,753	1.62	0.025
	Comparison	772	7,725		
5	*i-Ready*	1,089	9,614	1.39	0.052
	Comparison	775	7,720		

Summary and Discussion

Our study findings suggest *i-Ready* for mathematics in schools, when used by students with fidelity, has a positive impact on student mathematics achievement for striving learners in elementary grades 2–5. This includes striving learners who tested two or more grade levels below their current grade in mathematics achievement at baseline, and a subset of these students who fell at or below the 20th percentile in mathematics achievement. At each grade, striving learners who received *i-Ready* performed statistically significantly better on the mathematics Diagnostic than those in a comparison group.

Effect sizes provide additional support for *i-Ready*'s effectiveness with striving students overall and a subset of those at the bottom 20th percentile. For both groups, effect sizes for grades 3 and 5 were at the upper end of what Kraft (2019) indicates as typical and potentially meaningful in education, and those for grades 2 and 4 exceeded this range. Kraft (2019) suggests effect sizes should be considered in conjunction with all aspects of an intervention, including the magnitude of the treatment contrast and costs. Because *i-Ready* is personalized online learning intended as a supplemental activity to curricula and not an intense intervention, we consider the contrast between treatment and comparison to be relatively low. Thus, we consider the effect sizes for impacts on mathematics highly promising. Moreover, the comparison group implemented Diagnostic, which may have attenuated treatment effects for *i-Ready*.

Further, this study suggests *i-Ready* is equally effective for Black or African American striving learners as it is for all striving learners. The study also suggests *i-Ready* is equally effective for striving learners of Hispanic origin at grades 2, 3, and 5 as it is for all striving learners. Therefore, Black or African American striving learners in the *i-Ready* group performed better than Black or African American striving learners in the comparison group, and striving learners of Hispanic origin in the *i-Ready* group performed better than striving learners of Hispanic origin in the comparison group.

A positive Hispanic origin by treatment interaction suggests students of Hispanic origin at grade 4 benefited more from *i-Ready* than striving learners not of Hispanic origin. Similar to the results of other grades, striving learners of Hispanic origin at grade 4 performed better than striving learners of Hispanic origin in the comparison group. In addition, the positive interaction indicated

striving learners of Hispanic origin in the *i-Ready* group saw greater benefit from *i-Ready* than students in the *i-Ready* group who were not of Hispanic origin. In other words, *i-Ready* provided additional benefit to striving learner treatment group students of Hispanic origin over and above what it provided to other striving learners. Because we only saw this impact in one of the four grades examined, we recommend Curriculum Associates conduct additional studies to examine the impact of *i-Ready* on striving learners of Hispanic origin to determine if this was an anomaly or a meaningful finding.

Kraft (2019) 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 were everyday users, and *i-Ready* 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. The findings from this study, therefore, should be considered quite promising given that statistically significant impacts with modest to strong effect sizes were detected for all grade levels in the context of real-world implementation.

Our study was conducted as a rigorous QED to meet the current standards described by the WWC (WWC, 2020b) 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 the impact on mathematics achievement from *i-Ready* use for striving learners. This study, however, is not without some limitations.

First, 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 and collecting implementation fidelity information from treatment schools as well as collecting information about programs within comparison schools that might be similar in nature to *i-Ready*. We suggest including only one district to allow greater control of implementation and fewer confounds.

Next, our Black or African American and Hispanic origin interactions required use of student-level race and ethnicity demographic data. These data were missing for several students in the *i-Ready* usage datasets provided. We recommend Curriculum Associates continue ongoing efforts to increase the likelihood schools will provide this information for their students.

Finally, our treatment group was compared to a matched comparison group using the Diagnostic. It is possible that use of the Diagnostic itself increases student achievement. However, the design of this study did not allow for an estimation of that impact. Use of the Diagnostic only schools and students as a comparison group may have attenuated the effects of *i-Ready* use had this treatment group been compared to a “business as usual” comparison group. Future studies might examine the impact of *i-Ready* 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* with fidelity, the WWC 4.1 standards, and ESSA Level 2 guidelines.

Eligibility criteria for the treatment and comparison groups were determined through collaboration between the three study partners. Curriculum Associates provided information on the various components of *i-Ready* and the frequency for which it should be used for implementation with fidelity. They also provided *i-Ready* data to allow HumRRO and Century Analytics to empirically examine the extent to which these recommendations were followed by *i-Ready* 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/

Curriculum Associates. (2020, July 20). *The Research Behind our Programs*. Retrieved from https://www.curriculumassociates.com/research-and-efficacy

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

Stanford CEPA, n.d. “Racial and Ethnic Achievement Gaps.” The Educational Opportunity Monitoring Project. Retrieved from: https://cepa.stanford.edu/educational-opportunity-monitoring-project/achievement-gaps/race/

Swain, M., Randel, B., Norman Dvorak, R. (2020). Impact Evaluation of Mathematics *i-Ready Instruction* for Elementary Grades Using 2018-19 Data. (HumRRO Report 2019 No. 106). Retrieved from ERIC: https://files.eric.ed.gov/fulltext/ED604744.pdf

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.

What Works Clearinghouse, Institute of Education Sciences, U.S. Department of Education. (2020a, January). *What Works Clearinghouse: Procedures Handbook (Version 4.1).* Retrieved from http://whatworks.ed.gov

What Works Clearinghouse, Institute of Education Sciences, U.S. Department of Education. (2020b, January). *What Works Clearinghouse: Standards Handbook (Version 4.1).* Retrieved from http://whatworks.ed.gov

Appendix A
Selection of Analytic Samples

This appendix describes the process for identifying our analytic samples used to address confirmatory and exploratory research questions.

Confirmatory Analytic Samples

We used a multi-step process to identify analytic samples separately at each grade to address the confirmatory research question. First, we conducted school-level matching to identify a set of treatment and a set of comparison schools from which to match students. We matched schools on three school-level variables: (a) percent of students of historically marginalized race (i.e., a variable combining students identified as Black or African American, Asian, Pacific Islander, American Indian/Alaskan Native, or two or more races), (b) percent of students eligible for free or reduced-price lunch, and (c) grade-level enrollment. School-level matching was conducted separately for each grade and stratified by four levels of school urbanicity (city, suburb, town, and rural), resulting in four matched sets of schools. Table A.1 presents the average demographic composition of our three matching variables for our final matched mathematics *i-Ready* and comparison schools, and the effect size of the difference between the groups. As shown, the sets of treatment and comparison schools are baseline equivalent on our three matching variables.

Table A.1. Demographic Characteristics of Matched Mathematics i-Ready (Treatment) and Comparison Schools, by Grade

Grade	Variable	*i-Ready* Mean (SD)	Comparison Mean (SD)	Effect Size
2	FRL Percent	51.37 (26.12)	49.62 (28.27)	0.06
	Non-White Percent	48.43 (30.83)	47.07 (29.58)	0.05
	Grade Enrollment	82.80 (38.16)	81.49 (37.98)	0.03
3	FRL Percent	51.89 (26.84)	50.28 (29.42)	0.06
	Non-White Percent	49.28 (30.83)	48.88 (30.24)	0.01
	Grade Enrollment	85.13 (41.10)	84.24 (39.18)	0.02
4	FRL Percent	52.59 (27.03)	50.42 (29.19)	0.08
	Non-White Percent	49.47 (30.97)	48.98 (30.55)	0.02
	Grade Enrollment	88.11 (43.70)	86.31 (42.00)	0.04
5	FRL Percent	52.88 (25.87)	51.02 (28.77)	0.07
	Non-White Percent	49.76 (30.66)	49.05 (30.15)	0.02
	Grade Enrollment	90.45 (48.34)	86.36 (47.95)	0.08

Note. HMR = historically marginalized race; FRL = free or reduced lunch; SD = standard deviation of variables; Effect Size = Cohen's *d*.

Next, at each grade level we reduced our datafile to include only striving learners from the matched schools. To conduct student-level matching, we first stratified to ensure a treatment student was matched to a comparison student with the exact same placement levels on all

mathematics domains reported by the Diagnostic. We then used propensity score matching to match on composite fall Diagnostic scores within each stratum. Once matching was complete by stratum, we combined students across strata to generate one set of matched students for each grade, resulting in four separate analytic samples.

We calculated effect sizes to compare differences in student baseline achievement between our treatment and comparison groups in each grade using our planned baseline difference model (described in the next section). For all grades, Hedges' *g* was smaller than 0.25 after matching, and thus considered baseline equivalent (WWC, 2020a). Table A.2 presents our final matched samples of students.

We then reduced our sample to include only a subset of striving learners in the bottom 20th percentile. We examined the effect size differences between the treatment and comparison groups at each grade and similarly found all groups met baseline equivalence (see Table A.3).

Table A.2. Mathematics Baseline Equivalence Statistics for i-Ready (Treatment) and Comparison Striving Learner Groups, by Grade

Grade	Group	Schools	Students	Diagnostic Mean	Diagnostic SD	Adj Mean Diff (SE)	Effect Size
2	*i-Ready*	1,356	11,673	371.81	13.97	0.06 (0.21)	0.004
	Comparison	1,330	11,673	371.75	13.82		
3	*i-Ready*	1,410	14,679	395.59	15.84	0.17 (0.24)	0.011
	Comparison	1,376	14,679	395.41	15.87		
4	*i-Ready*	1,475	12,757	412.50	18.32	-0.25 (0.30)	-0.014
	Comparison	1,425	12,757	412.75	18.16		
5	*i-Ready*	1,500	12,566	427.51	20.59	-0.27 (0.33)	-0.013
	Comparison	1,438	12,566	427.77	20.48		

Note. SD = standard deviation of variables; SE = standard error; Effect Size = Hedge's *g*.

Table A.3. Mathematics Baseline Equivalence Statistics for i-Ready (Treatment) and Comparison Bottom 20th Percentile Striving Learner Groups, by Grade

Grade	Group	Schools	Students	Diagnostic Mean	Diagnostic SD	Adj Mean Diff (SE)	Effect Size
2	*i-Ready*	1,309	9,114	368.19	13.79	0.01 (0.23)	0.001
	Comparison	1,283	9,121	368.17	13.62		
3	*i-Ready*	1,314	9,606	388.45	15.25	0.19 (0.26)	0.012
	Comparison	1,295	9,598	388.26	15.31		
4	*i-Ready*	1,427	10,068	407.69	17.73	-0.20 (0.31)	-0.011
	Comparison	1,351	10,031	407.89	17.55		
5	*i-Ready*	1,447	9,956	422.46	20.29	-0.31 (0.35)	-0.015
	Comparison	1,372	9,969	422.77	20.17		

Note. SD = standard deviation of variables; SE = standard error; Effect Size = Hedge's *g*.

Exploratory Analytic Samples

For our exploratory research questions focused on Black or African American striving learners and those of Hispanic origin, we modified our sample to include only students from our matched sample with the necessary demographic data. To be included in the Black or African American analysis, a student needed to have complete race data. To be included in the Hispanic origin analysis, student needed to have complete Hispanic origin data. We determined this was preferred to imputation because (a) we had large enough sample sizes of students and schools with available data to have sufficient power, and (b) internal analysis suggested demographic data were missing at random. Though we did not conduct additional matching, we established baseline equivalence for the analyses for which we found significant interactions.

We conducted a post-hoc examination of baseline equivalence for the only significant interaction—Hispanic origin by treatment at grade 4. Particularly, we sought to confirm baseline equivalence was achieved for students of Hispanic origin in the *i-Ready* and comparison groups. Table A.4 summarizes these findings. As shown, the adjusted mean difference between these two groups is minimal, with an effect size of – 0.009. Therefore, our students of Hispanic origin were considered baseline equivalent.

Table A.4. Mathematics Baseline Equivalence Statistics for Grade 4 Students of Hispanic Origin of the i-Ready (Treatment) and Comparison Groups

Group	Students	*i-Ready* Mean	*i-Ready* SD	Adj Mean Diff	Effect Size
i-Ready	9,753	413.02	10.08	-0.16 (0.41)	-0.009
Comparison	7,725	413.05	17.93		

Appendix B
Analytic Model Descriptions

Benchmark Model

Level 1 of the benchmark model was specified as:

$$Yij = \beta 0j + \beta 1j(BASELINEij) + eij$$

where Yij is the spring Diagnostic score for student i in school j. β0j is the adjusted mean outcome for students in school j. β1j is the regression slope of the student's baseline (fall) Diagnostic score for school j. eij is the random error in the outcome associated with student i in school j not accounted for in the model.

Level 2 of the model was specified as:

$$\beta 0j = \gamma 00 + \gamma 01(TREATj) + \Sigma \gamma q(URBANICITYj) + \gamma 02(\%FRL) + \gamma 03(\%HMR) + \gamma 05(ENROLL) + u0j$$

$$\beta 1j = \gamma 10$$

where γ00 is the adjusted comparison group grand mean of the outcome, γ01 is the adjusted mean difference in the outcome between school study groups, and TREAT is an indicator variable coded as 1 for schools in the *i-Ready* treatment group and 0 for schools in the *i-Ready* comparison group. Σγq is a vector of indicator variables for school urbanicity (city, suburb, town, rural). γ02 - γ04 are regression slopes of the school-level covariates. ENROLL is number of students enrolled at the grade level in the analysis. u0j is the random error in the achievement outcome associated with school j.

We conducted two sensitivity analyses for each grade to examine the robustness of the findings of the benchmark impact model for the samples of striving learners. The first sensitivity analyses included a school-level grand mean centered baseline covariate. The second included student-level Diagnostic domain level scores to account for the stratification and matching of students within fall Diagnostic placement profiles. Both analyses yielded results consistent with our benchmark model.

We conducted two sensitivity analyses for each grade to examine the robustness of the findings of the benchmark impact model for the samples of striving learners. The first sensitivity analyses included a school-level grand mean centered baseline covariate. The second included student-level Diagnostic domain level scores to account for the stratification and matching of students within fall Diagnostic placement profiles. Both analyses yielded results consistent with our benchmark model.

Sensitivity Analysis 1. The first sensitivity analysis examined the robustness of the findings to including a school-level grand mean centered baseline covariate. Level 1 of the model had the same specification as the benchmark model. Level 2 of the models was specified as follows:

$$\beta 0_j = \gamma 00 + \gamma 01(TREAT_j) + \gamma 02(BASELINE_{.j} - BASELINE..)_j + \Sigma \gamma q\ (URBANCITY_j) + \gamma 03(\%FRL) + \gamma 04(\%HMR) + \gamma 05(ENROLL) + u0_j$$

where γ00 is the adjusted comparison group grand mean of the outcome, γ01 is the adjusted mean difference in the outcome between school study groups, and TREAT is an indicator variable coded as 1 for schools in the *i-Ready* treatment group and 0 for schools in the comparison group. γ02 is the regression slope of the school-level baseline Diagnostic score (grand mean centered). Σγq and γ03 - γ05 are regression slopes of the school-level covariates specified as described in the benchmark model. ENROLL is number of students enrolled at the grade level in the analysis. $u0_j$ is the random error in the achievement outcome associated with school $_j$.

Results of this model were consistent with the benchmark model findings for all grades: *i-Ready* had statistically significant impacts on student Mathematics achievement, and impact estimates were similar to those from the benchmark model.

Sensitivity Analysis 2. The second sensitivity analysis we conducted examined the robustness of the findings to including student level Diagnostic domain level scores to account for the stratification and matching of students within fall *i-Ready* domain placement profiles. Level 2 of this model as the same specification as Level 1 of the benchmark model. Level 1 of this sensitivity analysis model is specified as follows:

$$Y_{ij} = \beta0_j + \beta1_j(\text{BASELINE}_{ij}) + \beta2_j(\text{DOMAIN1}_{ij}) + \beta3_j(\text{DOMAIN2}_{ij}) + \beta4_j(\text{DOMAIN3}_{ij}) + e_{ij}$$

where Y_{ij} is the spring Diagnostic score for student $_i$ in school $_j$. $\beta0_j$ is the adjusted mean outcome for students in school $_j$. $\beta1_j$ is the regression slope of the student's baseline (fall) Diagnostic score for school $_j$. $B2_j$ – $\beta4_j$ are regression slopes of the baseline (fall) Diagnostic domain scores for student $_i$ in school $_j$. The mathematics domain scores at grades 2–5 include Algebra and Algebraic Thinking, Number and Operations, Geometry, and Measurement and Data. e_{ij} is the random error in the outcome associated with student $_i$ in school $_j$ not accounted for in the model.

Results of this model were consistent with the benchmark model findings for all grades. *i-Ready* had statistically significant impacts on student Mathematics achievement, and impact estimates were similar to those from the benchmark model.

Exploratory Models

We analyzed two additional models to address the exploratory research questions. For these models, an interaction term was added at Level 1. To address our research question focused on Black or African American striving learners, we defined Level 1 as:

$$Y_{ij} = \beta0_j + \beta1_j(\text{BASELINE}_{ij}) + \beta2_j(\text{BLACK OR AFRICAN AMERICAN}_{ij}) + e_{ij}$$

Level 2 of the model was specified as:

$$\beta0_j = \gamma00 + \gamma01(\text{TREAT}_j) + \gamma02(\text{URBANICITY}_j) + \gamma03(\%\text{FRL}) + \gamma04(\%\text{HMR}) + \gamma05(\text{ENROLL}) + u0_j$$

$$\beta1_j = \gamma10$$
$$\beta2_j = \gamma20 + \gamma21(\text{TREAT}_j)$$

To address our research question focused on striving learners of Hispanic origin, we defined Level 1 as:

$$Y_{ij} = \beta 0_j + \beta 1_j(BASELINE_{ij}) + \beta 2_j(HISPANIC_{ij}) + e_{ij}$$

Level 2 of the model was specified as:

$$\beta 0_j = \gamma 00 + \gamma 01(TREAT_j) + \gamma 02(URBANICITY_j) + \gamma 03(\%FRL) + \gamma 04(\%HISTORICALLY\ MARGINALIZED\ RACE) + \gamma 05(ENROLL) + u0_j$$

$$\beta 1_j = \gamma 10$$

$$\beta 2_j = \gamma 20 + \gamma 21(TREAT_j)$$

Level 1 of the model has the same specification as the benchmark model except for the addition of an indicator variable for student membership in either the Black or African American or Hispanic origin group. Level 2 of the model also had the same specification of the benchmark model except for the addition of the cross-level interaction of student race/ethnicity and treatment group status.

The two exploratory models were run separately with each of the four analytic samples.

Baseline Difference Model

We used a baseline difference model to provide a model-based estimate of the difference between students in the treatment and comparison groups on the baseline (fall Diagnostic) score separately for each grade level.

$$Y_{ij} = \beta 0_j + e_{ij}$$

where Y_{ij} is the fall Diagnostic score for student $_i$ in school $_j$. $\beta 0_j$ is the adjusted mean outcome for students in school $_j$. e_{ij} is the random error in the outcome associated with student $_i$ in school $_j$ not accounted for in the model.

Level 2 of the model was specified as follows:

$$\beta 0_j = \gamma 00 + \gamma 01(TREAT_j) + u0_j$$

where γ00 is the adjusted comparison group grand mean of the fall baseline score, γ01 is the adjusted mean difference in the baseline score between school study groups, and TREAT is an indicator variable coded as 1 for schools in the *i-Ready* treatment group and 0 for schools in the comparison group. $u0_j$ is the random error in the achievement outcome associated with school $_j$.

Appendix C
Confirmatory Impact HLM Coefficients for Striving Learners

Table C.1. Grade 2 Mathematics HLM Results for Striving Learners

Covariates	Coef.	SE	z	*p*-value	95% Conf. Interval	
Student-Level Covariates						
Fall 2018 *i-Ready Diagnostic* Score	0.74	0.01	83.22	<0.001	0.72	0.76
School-Level Covariates						
Treatment Group Membership	4.97	0.39	12.89	<0.001	4.21	5.73
Urbanicity – City*	-0.71	0.70	-1.02	0.306	-2.08	0.65
Urbanicity – Suburban*	-0.67	0.65	-1.03	0.305	-1.95	0.61
Urbanicity – Town*	-0.97	0.89	-1.09	0.274	-2.72	0.77
Percent Free or Reduced Lunch	-8.29	0.95	-8.68	<0.001	-10.16	-6.42
Percent HMR	-4.16	0.88	-4.75	<0.001	-5.88	-2.44
Grade-level Enrollment	0.01	0.01	0.99	0.322	-0.01	0.02
Intercept						
Intercept	134.48	3.42	39.27	<0.001	127.77	141.19

Notes: *Rural is the reference group for urbanicity; HMR = historically marginalized race

Table C.2. Grade 3 Mathematics HLM Results for Striving Learners

Covariates	Coef.	SE	z	*p*-value	95% Conf. Interval	
Student-Level Covariates						
Fall 2018 *i-Ready Diagnostic* Score	0.90	0.01	125.55	<0.001	0.88	0.91
School-Level Covariates						
Treatment Group Membership	4.08	0.39	10.53	<0.001	3.32	4.84
Urbanicity – City*	-1.50	0.70	-2.14	0.033	-2.87	-0.12
Urbanicity – Suburban*	-1.35	0.67	-2.03	0.043	-2.65	-0.04
Urbanicity – Town*	-2.47	0.89	-2.78	0.005	-4.20	-0.73
Percent Free or Reduced Lunch	-6.75	0.95	-7.09	<0.001	-8.61	-4.88
Percent HMR	-2.33	0.89	-2.63	0.009	-4.07	-0.59
Grade-level Enrollment	0.02	0.00	3.22	0.001	0.01	0.03
Intercept						
Intercept	76.60	2.95	25.98	<0.001	70.82	82.38

Notes: *Rural is the reference group for urbanicity; HMR = historically marginalized race

Table C.3. Grade 4 Mathematics HLM Results for Striving Learners

Covariates	Coef.	SE	z	*p*-value	95% Conf. Interval	
Student-Level Covariates						
Fall 2018 *i-Ready Diagnostic* Score	0.90	0.01	127.92	<0.001	0.89	0.91
School-Level Covariates						
Treatment Group Membership	5.46	0.39	13.90	<0.001	4.69	6.24
Urbanicity – City*	-0.46	0.68	-0.67	0.502	-1.79	0.88
Urbanicity – Suburban*	0.24	0.65	0.37	0.712	-1.03	1.51
Urbanicity – Town*	-1.56	0.93	-1.68	0.094	-3.37	0.26
Percent Free or Reduced Lunch	-4.91	0.99	-4.96	<0.001	-6.85	-2.97
Percent HMR	-4.62	0.90	-5.11	<0.001	-6.39	-2.85
Grade-level Enrollment	0.01	0.00	2.67	0.008	0.00	0.02
Intercept						
Intercept	70.77	3.03	23.36	<0.001	64.84	76.71

Notes: *Rural is the reference group for urbanicity; HMR = historically marginalized race

Table C.4. Grade 5 Mathematics HLM Results for Striving Learners

Covariates	Coef.	SE	z	*p*-value	95% Conf. Interval	
Student-Level Covariates						
Fall 2018 *i-Ready Diagnostic* Score	0.87	0.01	139.15	<0.001	0.85	0.88
School-Level Covariates						
Treatment Group Membership	3.64	0.38	9.67	<0.001	2.90	4.38
Urbanicity – City*	-1.14	0.65	-1.75	0.081	-2.42	0.14
Urbanicity – Suburban*	-0.59	0.62	-0.94	0.346	-1.81	0.64
Urbanicity – Town*	-0.74	0.90	-0.82	0.41	-2.50	1.02
Percent Free or Reduced Lunch	-5.11	0.93	-5.48	<0.001	-6.93	-3.28
Percent HMR	-1.23	0.84	-1.46	0.146	-2.88	0.42
Grade-level Enrollment	0.01	0.00	2.11	0.035	0.00	0.02
Intercept						
Intercept	81.01	2.78	29.14	<0.001	75.57	86.46

Notes: *Rural is the reference group for urbanicity; HMR = historically marginalized race

Appendix D
Confirmatory Impact HLM Coefficients for Striving Learners at the Bottom 20th Percentile

Table D.1. Grade 2 Mathematics HLM Results for Striving learners at the Bottom 20th Percentile

Covariates	Coef.	SE	z	*p*-value	95% Conf. Interval	
Student-Level Covariates						
Fall 2018 *i-Ready Diagnostic* Score	0.71	0.01	68.28	<0.001	0.69	0.73
School-Level Covariates						
Treatment Group Membership	4.91	0.42	11.69	<0.001	4.09	5.73
Urbanicity – City*	-0.96	0.77	-1.25	0.212	-2.46	0.55
Urbanicity – Suburban*	-0.95	0.72	-1.31	0.189	-2.36	0.47
Urbanicity – Town*	-1.28	0.97	-1.31	0.189	-3.19	0.63
Percent Free or Reduced Lunch	-8.47	1.05	-8.06	<0.001	-10.54	-6.41
Percent HMR	-3.61	0.95	-3.78	<0.001	-5.48	-1.73
Grade-level Enrollment	0.00	0.01	0.81	0.416	-0.01	0.02
Intercept						
Intercept	144.82	3.96	36.59	<0.001	137.06	152.58

Notes: *Rural is the reference group for urbanicity; HMR = historically marginalized race

Table D.2. Grade 3 Mathematics HLM Results for Striving learners at the Bottom 20th Percentile

Covariates	Coef.	SE	z	*p*-value	95% Conf. Interval	
Student-Level Covariates						
Fall 2018 *i-Ready Diagnostic* Score	0.86	0.01	89.14	<0.001	0.84	0.88
School-Level Covariates						
Treatment Group Membership	4.43	0.46	9.58	<0.001	3.52	5.33
Urbanicity – City*	-1.69	0.85	-2.00	0.046	-3.34	-0.03
Urbanicity – Suburban*	-1.08	0.81	-1.34	0.180	-2.67	0.50
Urbanicity – Town*	-2.75	1.06	-2.59	0.010	-4.84	-0.67
Percent Free or Reduced Lunch	-6.71	1.15	-5.84	<0.001	-8.97	-4.46
Percent HMR	-2.46	1.06	-2.32	0.020	-4.53	-0.38
Grade-level Enrollment	0.02	0.01	2.93	0.003	0.01	0.03
Intercept						
Intercept	89.20	3.89	22.93	<0.001	81.57	96.82

Notes: *Rural is the reference group for urbanicity; HMR = historically marginalized race

Table D.3. Grade 4 Mathematics HLM Results for Striving learners at the Bottom 20th Percentile

Covariates	Coef.	SE	z	*p*-value	95% Conf. Interval	
Student-Level Covariates						
Fall 2018 *i-Ready Diagnostic* Score	0.91	0.01	107.19	<0.001	0.89	0.92
School-Level Covariates						
Treatment Group Membership	5.64	0.44	12.94	<0.001	4.79	6.50
Urbanicity – City*	-0.60	0.76	-0.79	0.430	-2.08	0.89
Urbanicity – Suburban*	-0.17	0.73	-0.23	0.819	-1.59	1.26
Urbanicity – Town*	-1.13	1.03	-1.10	0.270	-3.15	0.88
Percent Free or Reduced Lunch	-5.22	1.10	-4.73	<0.001	-7.38	-3.06
Percent HMR	-4.42	1.00	-4.42	<0.001	-6.38	-2.46
Grade-level Enrollment	0.01	0.01	2.19	0.028	0.00	0.02
Intercept						
Intercept	68.54	3.58	19.15	<0.001	61.53	75.56

Notes: *Rural is the reference group for urbanicity; HMR = historically marginalized race

Table D.4. Grade 5 Mathematics HLM Results for Striving learners at the Bottom 20th Percentile

Covariates	Coef.	SE	z	*p*-value	95% Conf. Interval	
Student-Level Covariates						
Fall 2018 *i-Ready Diagnostic* Score	0.85	0.01	116.42	<0.001	0.84	0.87
School-Level Covariates						
Treatment Group Membership	3.54	0.42	8.43	<0.001	2.72	4.37
Urbanicity – City*	-1.24	0.73	-1.68	0.092	-2.68	0.20
Urbanicity – Suburban*	-0.78	0.71	-1.10	0.270	-2.16	0.60
Urbanicity – Town*	-0.77	1.01	-0.76	0.446	-2.74	1.21
Percent Free or Reduced Lunch	-5.37	1.05	-5.13	<0.001	-7.42	-3.32
Percent HMR	-0.83	0.94	-0.89	0.375	-2.67	1.01
Grade-level Enrollment	0.01	0.00	2.36	0.018	0.00	0.02
Intercept						
Intercept	87.24	3.22	27.12	<0.001	80.93	93.54

Notes: *Rural is the reference group for urbanicity; HMR = historically marginalized race

Appendix E
Exploratory HLM with Black or African American by Treatment Interaction

Table E.1. HLM Results for Grade 2 Mathematics with Black or African American by Treatment Interaction

Covariates	Coef.	SE	z	*p*-value	95% Conf. Interval	
Student-Level Covariates						
Fall 2018 *i-Ready Diagnostic* Score	0.74	0.01	62.84	< 0.001	0.72	0.76
Black or African American	-3.02	0.69	-4.35	< 0.001	-4.38	-1.66
Black or African American by Treatment	-0.31	0.88	-0.36	0.721	-2.03	1.40
School-Level Covariates						
Treatment Group Membership	4.49	0.54	8.37	< 0.001	3.44	5.54
Urbanicity – City*	-1.35	0.91	-1.48	0.137	-3.14	0.43
Urbanicity – Suburban*	-0.77	0.85	-0.91	0.365	-2.43	0.89
Urbanicity – Town*	-0.85	1.16	-0.74	0.460	-3.12	1.41
Percent Free or Reduced Lunch	-8.32	1.32	-6.32	< 0.001	-10.90	-5.74
Percent HMR	-1.45	1.16	-1.25	0.212	-3.73	0.83
Grade-level Enrollment	0.01	0.01	1.33	0.183	0.00	0.02
Intercept						
Intercept	135.15	4.54	29.80	< 0.001	126.26	144.04

Notes: *Rural is the reference group for urbanicity; HMR = historically marginalized race

Table E.2. HLM Results for Grade 3 Mathematics with Black or African American by Treatment Interaction

Covariates	Coef.	SE	z	*p*-value	95% Conf. Interval	
Student-Level Covariates						
Fall 2018 *i-Ready Diagnostic* Score	0.90	0.01	97.98	< 0.001	0.88	0.92
Black or African American	-3.74	0.61	-6.10	< 0.001	-4.95	-2.54
Black or African American by Treatment	0.56	0.79	0.72	0.475	-0.98	2.11
School-Level Covariates						
Treatment Group Membership	3.49	0.53	6.52	< 0.001	2.44	4.54
Urbanicity – City*	-0.61	0.92	-0.66	0.507	-2.40	1.19
Urbanicity – Suburban*	-0.92	0.86	-1.08	0.282	-2.61	0.76
Urbanicity – Town*	-3.60	1.15	-3.14	< 0.001	-5.85	-1.35
Percent Free or Reduced Lunch	-8.27	1.26	-6.54	< 0.001	-10.74	-5.79
Percent HMR	0.50	1.17	0.43	0.669	-1.80	2.80
Grade-level Enrollment	0.02	0.01	3.25	< 0.001	0.01	0.03
Intercept						
Intercept	75.90	3.81	19.93	<0.001	68.43	83.36

Notes: *Rural is the reference group for urbanicity; HMR = historically marginalized race

Table E.3. HLM Results for Grade 4 Mathematics with Black or African American by Treatment Interaction

Covariates	Coef.	SE	z	*p*-value	95% Conf. Interval	
Student-Level Covariates						
Fall 2018 *i-Ready Diagnostic* Score	0.91	0.01	98.92	< 0.001	0.89	0.93
Black or African American	-3.58	0.69	-5.21	< 0.001	-4.93	-2.23
Black or African American by Treatment	0.67	0.87	0.77	0.442	-1.04	2.39
School-Level Covariates						
Treatment Group Membership	4.52	0.56	8.13	< 0.001	3.43	5.61
Urbanicity – City*	-0.09	0.92	-0.10	0.923	-1.89	1.72
Urbanicity – Suburban*	0.78	0.87	0.90	0.369	-0.92	2.49
Urbanicity – Town*	-2.21	1.19	-1.86	0.064	-4.53	0.12
Percent Free or Reduced Lunch	-5.63	1.35	-4.16	< 0.001	-8.28	-2.98
Percent Historically marginalized race	-1.18	1.22	-0.97	0.333	-3.57	1.21
Grade-level Enrollment	0.02	0.01	3.03	< 0.001	0.01	0.03
Intercept						
Intercept	66.41	3.97	16.71	< 0.001	58.62	74.20

Notes: *Rural is the reference group for urbanicity; HMR = historically marginalized race

Table E.4. HLM Results for Grade 5 Mathematics with Black or African American by Treatment Interaction

Covariates	Coef.	SE	z	*p*-value	95% Conf. Interval	
Student-Level Covariates						
Fall 2018 *i-Ready Diagnostic* Score	0.86	0.01	108.23	< 0.001	0.85	0.88
Black or African American	-2.83	0.67	-4.24	< 0.001	-4.14	-1.52
Black or African American by Treatment	0.29	0.86	0.34	0.733	-1.39	1.97
School-Level Covariates						
Treatment Group Membership	4.17	0.53	7.85	< 0.001	3.13	5.21
Urbanicity – City*	-1.15	0.86	-1.34	0.181	-2.84	0.54
Urbanicity – Suburban*	-0.98	0.81	-1.21	0.225	-2.58	0.61
Urbanicity – Town*	-0.30	1.13	-0.26	0.792	-2.51	1.91
Percent Free or Reduced Lunch	-4.46	1.25	-3.55	< 0.001	-6.92	-2.00
Percent Historically marginalized race	0.16	1.12	0.15	0.884	-2.02	2.35
Grade-level Enrollment	0.01	0.00	1.72	0.086	0.00	0.02
Intercept						
Intercept	81.84	3.59	22.80	< 0.001	74.80	88.87

Notes: *Rural is the reference group for urbanicity; HMR = historically marginalized race

Appendix F. Impacts and Baseline to Outcome Gains for Black or African American by Treatment Interactions

Table F.1 summarizes the achievement differences between of Black or African American *i-Ready* and comparison striving learners by grade. We did not find a statistically significant interaction at any grade, so the differences between groups is not statistically significantly different. The adjusted mean differences between *i-Ready* and comparison groups ranged from just above 4 points (grade 3) and more than 5 points (grade 4). The effect sizes ranged from 0.16 (grade 3) to 0.20 (grade 4) and improvement indices ranged from 6.36 (grade 3) to 7.93 (grade 4).

Table F.1. Differences between Treatment and Comparison Black or African American groups, by Grade.

Grade	Group	Students	Diagnostic Mean	Diagnostic SD	Adj. Mean Difference	Effect Size	Improvement Index
2	*i-Ready*	1,994	406.65	21.34	4.18	0.19	7.53
	Comparison	1,261	402.48	22.30			
3	*i-Ready*	2,504	429.25	24.35	4.05	0.16	6.36
	Comparison	1,735	425.20	25.75			
4	*i-Ready*	2,306	443.48	25.54	5.19	0.20	7.93
	Comparison	1,580	438.29	26.66			
5	*i-Ready*	2,114	452.03	24.35	4.46	0.17	6.75
	Comparison	1,676	447.56	27.42			

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

Table F.2 provides the mean change in the mathematics Diagnostic between baseline and outcome for Black or African American and non-Black or African American *i-Ready* and Comparison groups by grade. They are illustrated in Figures F.1–F.4. The means were generated by the interaction models presented in Tables E.1 through E.4 of Appendix E. Please note that these interactions were not statistically significant – indicating the impact of *i-Ready* did not differ between Black or African American and non-Black or African American student groups (i.e., we expect *i-Ready* to be equally as advantageous to both Black or African American and non-Black or African American striving learners). In addition, although gain scores provide a good approximation of the achievement growth, they do not provide model-based estimates of group differences or impacts.

Table F.2. Baseline to Outcome Change in Mathematics Diagnostic Performance for Black or African American and Non-Black or African American Striving Learners in i-Ready (Treatment) Schools Compared to Comparison Schools, by Grade.

Grade	Group	Group	Students	Diagnostic Baseline Mean	Diagnostic Outcome Mean	Baseline to Outcome Gain
2	Black or African American	*i-Ready*	1,994	371.51	406.65	35.15
		Comparison	1,261	370.92	402.48	31.56
	Not Black or African American	*i-Ready*	4,351	372.21	409.99	37.77
		Comparison	6,262	372.46	405.50	33.04
3	Black or African American	*i-Ready*	2,504	395.69	429.25	33.56
		Comparison	1,735	393.89	425.20	31.30
	Not Black or African American	*i-Ready*	7,873	396.20	432.43	36.23
		Comparison	5,594	396.65	428.94	32.29
4	Black or African American	*i-Ready*	2,306	412.73	443.48	30.74
		Comparison	1,580	410.89	438.29	27.39
	Not Black or African American	*i-Ready*	6,554	413.17	446.39	33.22
		Comparison	4,804	414.12	441.87	27.76
5	Black or African American	*i-Ready*	2,114	427.68	452.03	24.35
		Comparison	1,676	425.89	447.56	21.68
	Not Black or African American	*i-Ready*	6,828	428.42	454.57	26.15
		Comparison	5,049	429.21	450.40	21.19

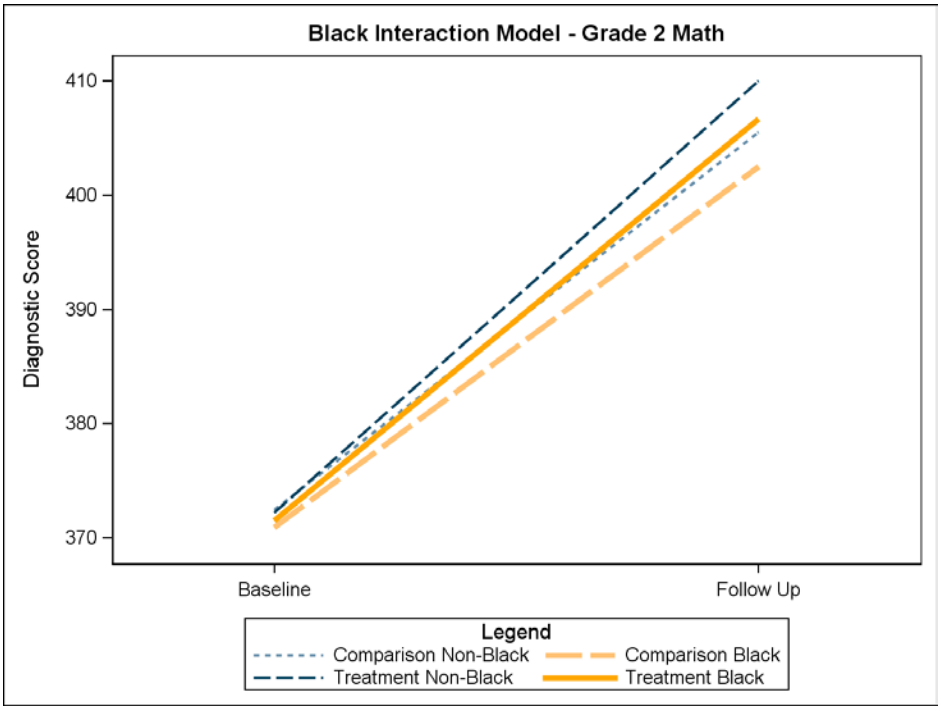

Figure F.1. Gains in Mathematics Diagnostic achievement between baseline and outcome for the i-Ready (treatment) and comparison groups for Black or African American and non-Black or African American groups at grade 2.

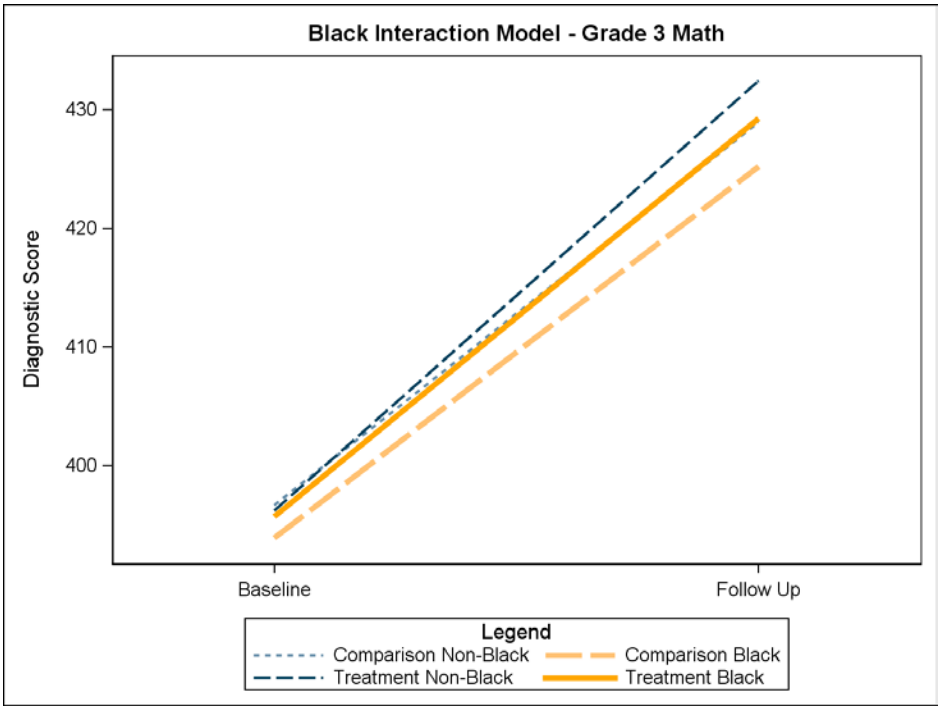

Figure F.2. Gains in Mathematics Diagnostic achievement between baseline and outcome for the i-Ready (treatment) and comparison groups for Black or African American and non-Black or African American groups at grade 3.

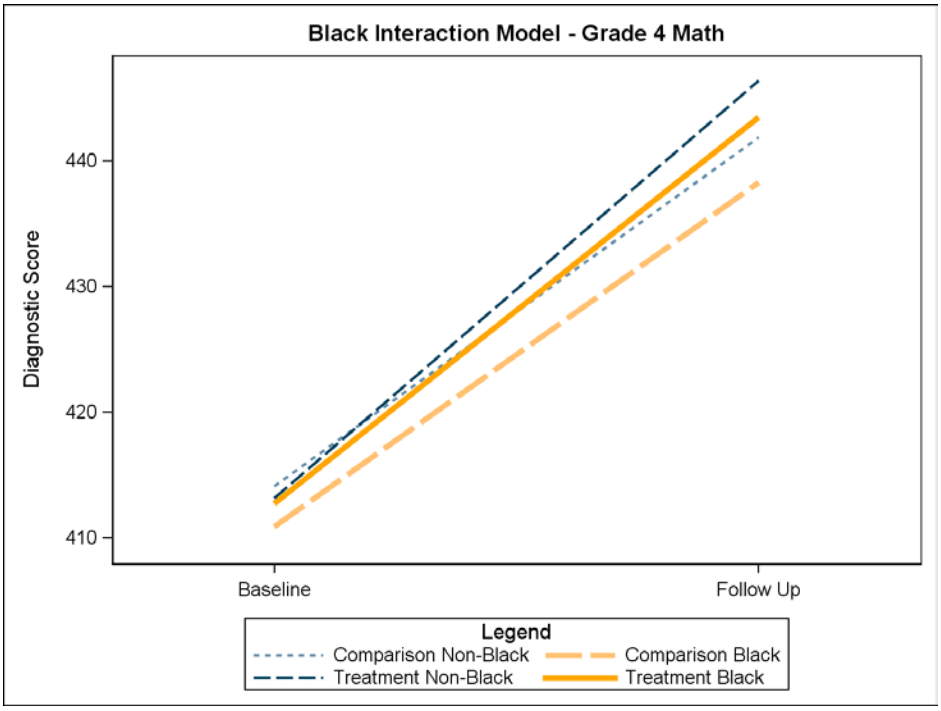

Figure F.3. Gains in Mathematics Diagnostic achievement between baseline and outcome for the i-Ready (treatment) and comparison groups for Black or African American and non-Black or African American groups at grade 4.

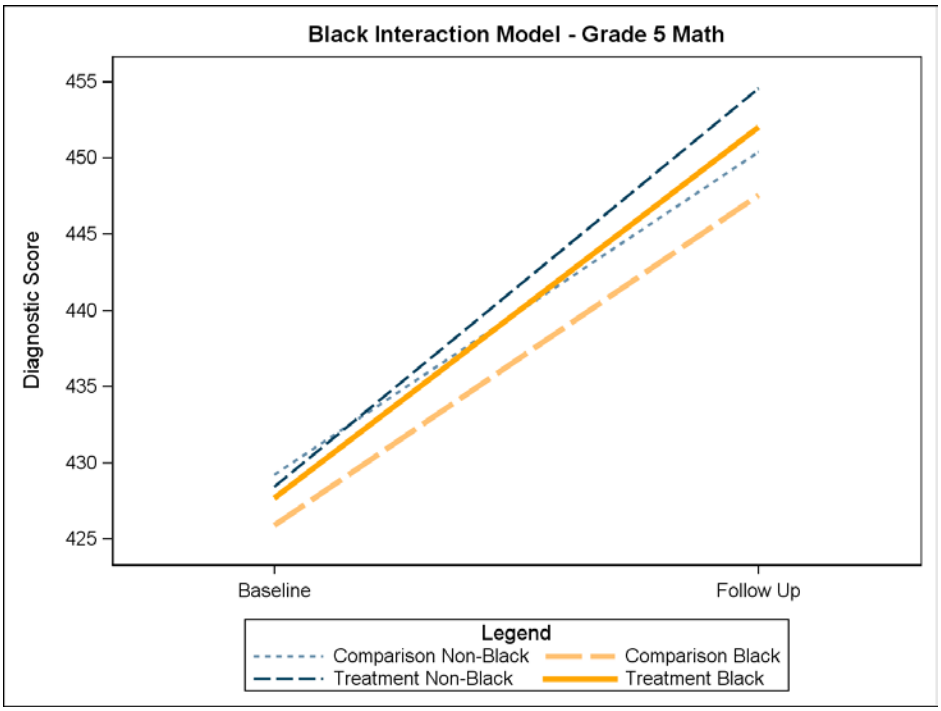

Figure F.4. Gains in Mathematics Diagnostic achievement between baseline and outcome for the i-Ready (treatment) and comparison groups for Black or African American and non-Black or African American groups at grade 5.

Appendix G
Exploratory HLM with Hispanic Origin by Treatment Interaction

Table G.1. HLM Results for Grade 2 Mathematics with Hispanic Origin by Treatment Interaction

Covariates	Coef.	SE	z	*p*-value	95% Conf. Interval	
Student-Level Covariates						
Fall 2018 *i-Ready Diagnostic* Score	0.73	0.01	66.78	< 0.001	0.71	0.75
Hispanic Origin	-1.47	0.54	-2.74	0.01	-2.53	-0.42
Hispanic Origin by Treatment	0.99	0.71	1.40	0.16	-0.40	2.38
School-Level Covariates						
Treatment Group Membership	4.57	0.52	8.82	< 0.001	3.55	5.59
Urbanicity – City*	-0.49	0.86	-0.57	0.57	-2.19	1.20
Urbanicity – Suburban*	-0.26	0.81	-0.32	0.75	-1.84	1.33
Urbanicity – Town*	-0.59	1.08	-0.54	0.59	-2.71	1.53
Percent Free or Reduced Lunch	-9.81	1.21	-8.14	< 0.001	-12.18	-7.45
Percent HMR	-1.91	1.11	-1.72	0.09	-4.08	0.27
Grade-level Enrollment	0.01	0.01	1.66	0.10	0.00	0.02
Intercept						
Intercept	139.37	4.21	33.13	< 0.001	131.13	147.62

Notes: *Rural is the reference group for urbanicity; HMR = historically marginalized race

Table G.2. HLM Results for Grade 3 Mathematics with Hispanic Origin by Treatment Interaction

Covariates	Coef.	SE	z	*p*-value	95% Conf. Interval	
Student-Level Covariates						
Fall 2018 *i-Ready Diagnostic* Score	0.89	0.01	104.95	< 0.001	0.87	0.91
Hispanic Origin	-0.45	0.48	-0.92	0.36	-1.40	0.50
Hispanic Origin by Treatment	1.17	0.64	1.82	0.07	-0.09	2.42
School-Level Covariates						
Treatment Group Membership	3.82	0.51	7.46	< 0.001	2.82	4.83
Urbanicity – City*	-0.72	0.88	-0.82	0.41	-2.45	1.00
Urbanicity – Suburban*	-0.87	0.83	-1.05	0.30	-2.50	0.76
Urbanicity – Town*	-3.15	1.09	-2.90	< 0.001	-5.27	-1.02
Percent Free or Reduced Lunch	-7.90	1.20	-6.58	< 0.001	-10.25	-5.55
Percent HMR	-1.58	1.13	-1.39	0.16	-3.80	0.64
Grade-level Enrollment	0.02	0.01	3.01	< 0.001	0.01	0.03
Intercept						
Intercept	78.68	3.54	22.22	< 0.001	71.74	85.62

Notes: *Rural is the reference group for urbanicity; HMR = historically marginalized race

Table G.3. HLM Results for Grade 4 Mathematics with Hispanic Origin by Treatment Interaction

Covariates	Coef.	SE	z	*p*-value	95% Conf. Interval	
Student-Level Covariates						
Fall 2018 *i-Ready Diagnostic* Score	0.91	0.01	106.72	< 0.001	0.90	0.93
Hispanic Origin	-0.86	0.54	-1.59	0.11	-1.92	0.20
Hispanic Origin by Treatment	1.61	0.72	2.25	0.03	0.21	3.02
School-Level Covariates						
Treatment Group Membership	4.29	0.53	8.08	< 0.001	3.25	5.33
Urbanicity – City*	-0.41	0.88	-0.47	0.64	-2.13	1.30
Urbanicity – Suburban*	0.50	0.83	0.60	0.55	-1.13	2.13
Urbanicity – Town*	-1.48	1.12	-1.32	0.19	-3.68	0.72
Percent Free or Reduced Lunch	-6.37	1.25	-5.09	< 0.001	-8.83	-3.92
Percent HMR	-2.09	1.15	-1.81	0.07	-4.35	0.17
Grade-level Enrollment	0.02	0.01	3.17	0.00	0.01	0.03
Intercept						
Intercept	65.56	3.70	17.72	< 0.001	58.31	72.81

Notes: *Rural is the reference group for urbanicity; HMR = historically marginalized race

Table G.4. HLM Results for Grade 5 Mathematics with Hispanic Origin by Treatment Interaction

Covariates	Coef.	SE	z	*p*-value	95% Conf. Interval	
Student-Level Covariates						
Fall 2018 *i-Ready Diagnostic* Score	0.86	0.01	114.62	< 0.001	0.85	0.88
Hispanic Origin	-0.48	0.54	-0.89	0.37	-1.54	0.57
Hispanic Origin by Treatment	1.39	0.72	1.94	0.05	-0.01	2.79
School-Level Covariates						
Treatment Group Membership	3.25	0.52	6.27	< 0.001	2.23	4.26
Urbanicity – City*	-1.37	0.83	-1.66	0.10	-2.99	0.25
Urbanicity – Suburban*	-1.19	0.79	-1.51	0.13	-2.74	0.35
Urbanicity – Town*	-0.03	1.10	-0.02	0.98	-2.17	2.12
Percent Free or Reduced Lunch	-5.30	1.19	-4.46	< 0.001	-7.63	-2.97
Percent HMR	-0.56	1.08	-0.52	0.61	-2.67	1.56
Grade-level Enrollment	0.01	0.00	2.00	0.05	0.00	0.02
Intercept						
Intercept	82.83	3.39	24.45	< 0.001	76.19	89.46

Notes: *Rural is the reference group for urbanicity; HMR = historically marginalized race

Appendix H. Impacts and Baseline to Outcome Gains for Hispanic Origin by Treatment Interactions

Table H.1 summarizes the differences in achievement for striving learners of Hispanic origin in the *i-Ready* and comparison groups. We did not find a statistically significant interaction at grades 2,3, or 5 so the difference between these two groups is not statistically significantly different. For these grades, we noted adjusted mean differences between 4.64 (grade 5) and 5.56 (grade 2) points between the *i-Ready* and comparison groups, and effect sizes ranged from 0.19 (grade 5) to 0.25 (grade 2). The improvement indices fell between 7.53 (grade 5) and 9.87 (grade 2). We found a positive significant interaction at grade 4 – indicating differences in the impact of *i-Ready* on students of Hispanic origin compared to students not of Hispanic origin. We noted an adjusted mean difference of almost 6 points between these two groups and an effect size of 0.22. This is considered a large effect for an education intervention based on Kraft's guidelines (2019). The improvement index was 8.71. This suggests that a student of Hispanic origin in the comparison group in this study would be expected to improve by more than eight percentiles if they were to use *i-Ready* with fidelity.

Table H.1. Differences between i-Ready (Treatment) and Comparison Hispanic Origin groups, by Grade.

Grade	Group	Students	Diagnostic Mean	Diagnostic SD	Adj. Mean Difference	Effect Size	Improvement Index
2	*i-Ready*	3,154	408.56	22.48	5.56	0.25	9.87
	Comparison	2,267	403.00	21.02			
3	*i-Ready*	3,986	431.99	24.25	4.99	0.21	8.32
	Comparison	2,849	427.00	23.75			
4	*i-Ready*	3,401	445.42	27.08	5.91*	0.22	8.71
	Comparison	2,528	439.52	25.60			
5	*i-Ready*	3,592	453.14	24.25	4.64	0.19	7.53
	Comparison	2,421	448.50	25.28			

*Notes: * = statistically significant interaction;* SD = standard deviation of Diagnostic scores; Adj Mean Diff = adjusted mean difference between *i-Ready* and Comparison groups; SE = standard error of the adjusted mean difference; and Effect Size = Hedge's *g*.

Table H.2 provides the mean change in the mathematics Diagnostic between baseline and outcome for students of Hispanic origin and students not of Hispanic origin *i-Ready* and comparison groups by grade. They are illustrated in Figures H.1– I.4. The means were generated by the interaction models presented in Tables G.1– G.4 of Appendix G. Please note that these interactions were not statistically significant at grades 2–4, indicating the impact of *i-Ready* did not differ between Hispanic origin and non-Hispanic origin student groups (i.e., we expect *i-Ready* to be equally as advantageous to both striving learners of Hispanic origin and those not of Hispanic origin). We did find a positive statistically significant interaction at grade 4, indicating students of Hispanic origin saw greater benefit using *i-Ready* than students not of Hispanic origin. As shown, striving learners of Hispanic origin showed gains that surpasseds the gains of the comparison groups and of the non-Hispanic origin treatment group. Although gain scores provide a good approximation of the achievement growth, they do not provide model-based estimates of group differences or impacts.

Table I.2. Baseline to Outcome Change in Mathematics i-Ready Diagnostic Performance for Hispanic Origin and Non-Hispanic Origin Students in Treatment Schools Compared to Comparison Schools, by grade.

Grade	Group	Group	Students	*i-Ready Baseline* Mean	*i-Ready* Outcome Mean	*Baseline to Outcome Gain*
2	Hispanic Origin	*i-Ready*	3,154	370.95	408.56	37.61
		Comparison	2,267	371.18	403.00	31.82
	Not Hispanic Origin	*i-Ready*	5,968	372.15	409.04	36.89
		Comparison	4,508	372.13	404.47	32.34
3	Hispanic Origin	*i-Ready*	3,986	394.52	431.99	37.47
		Comparison	2,849	394.84	427.00	32.17
	Not Hispanic Origin	*i-Ready*	7,519	396.29	431.27	34.98
		Comparison	5,955	396.08	427.45	31.37
4	Hispanic Origin	*i-Ready*	3,401	411.78	445.42	33.64
		Comparison	2,528	412.42	439.52	27.10
	Not Hispanic Origin	*i-Ready*	6,352	413.34	444.67	31.33
		Comparison	5,197	413.49	440.38	26.89
5	Hispanic Origin	*i-Ready*	3,592	426.66	453.14	26.47
		Comparison	2,421	427.42	448.50	21.08
	Not Hispanic Origin	*i-Ready*	6,022	428.45	452.23	23.78
		Comparison	5,299	428.08	448.98	20.90

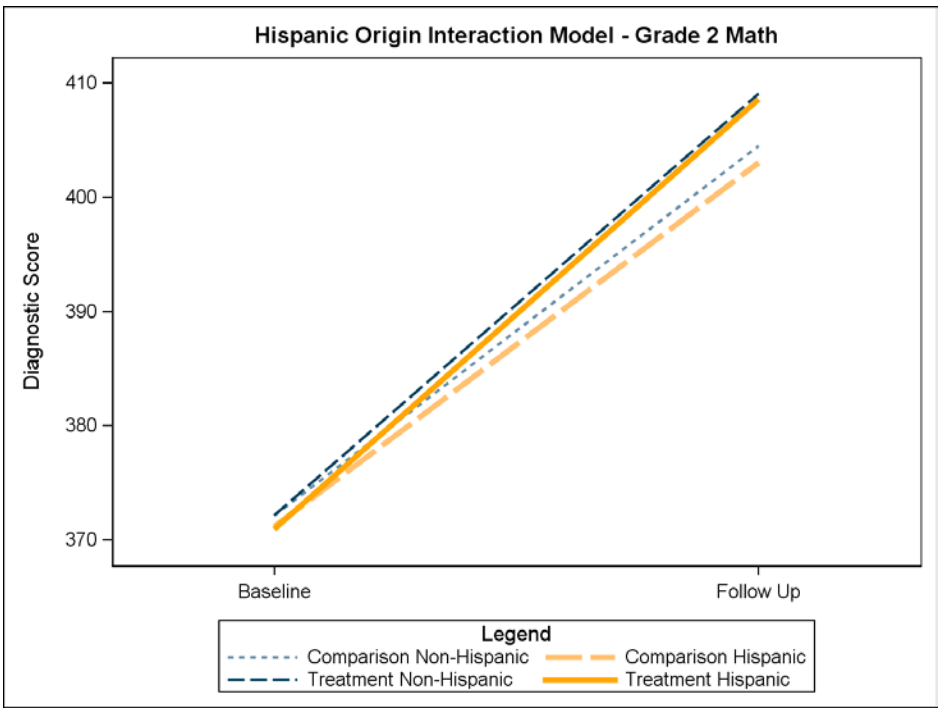

Figure H.1. Gains in Mathematics Diagnostic achievement between baseline and outcome for the i-Ready (treatment) and comparison groups for Hispanic Origin and non-Hispanic Origin groups at grade 2.

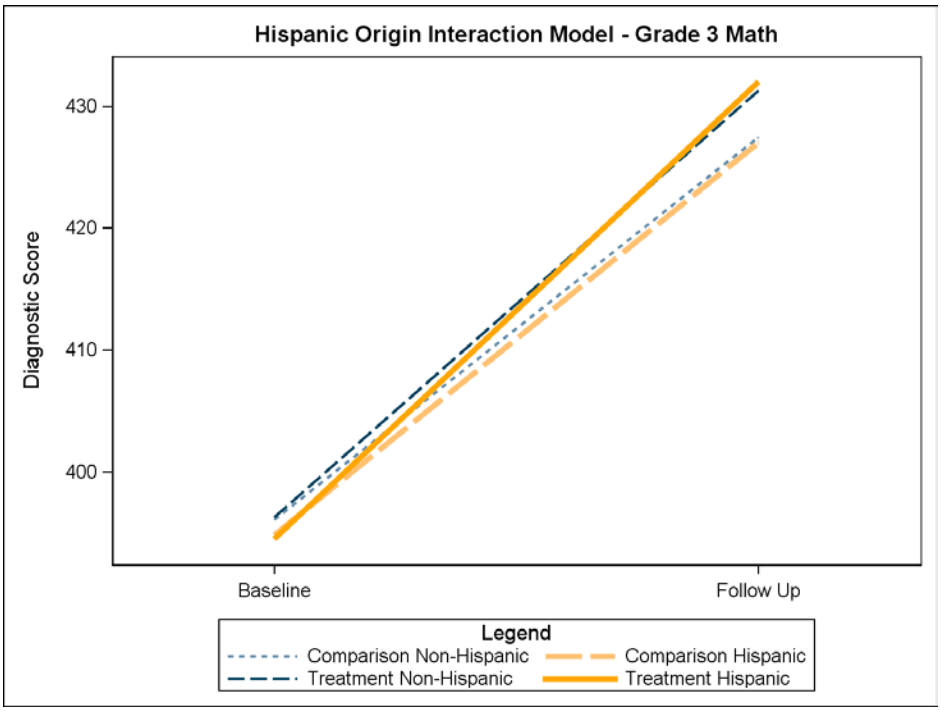

Figure H.2. Gains in Mathematics Diagnostic achievement between baseline and outcome for the i-Ready (treatment) and comparison groups for Hispanic Origin and non-Hispanic Origin groups at grade 3.

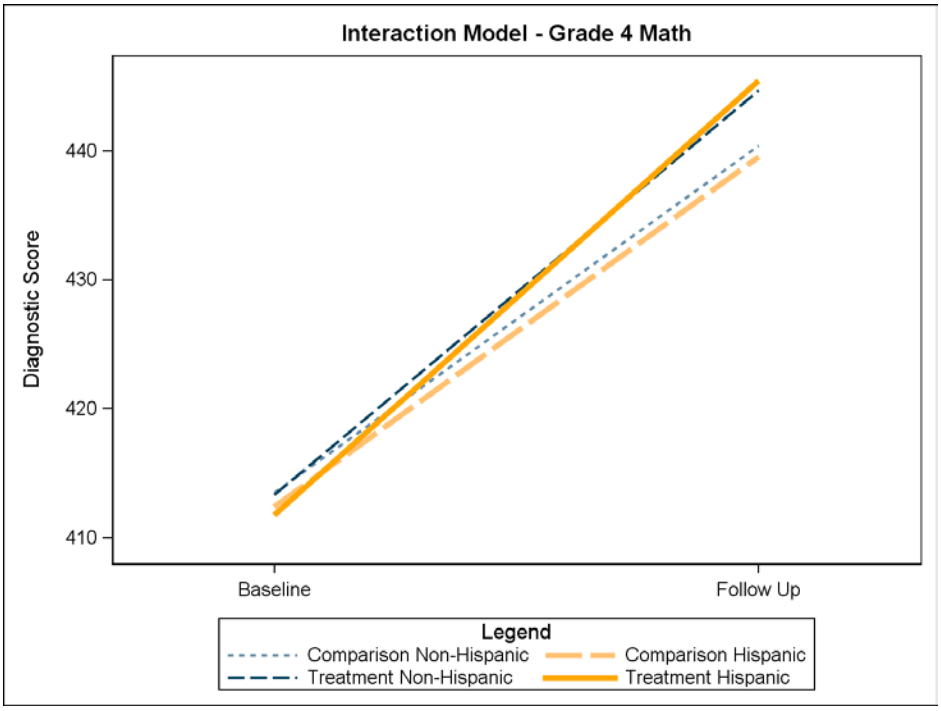

Figure H.3. Gains in Mathematics Diagnostic achievement between baseline and outcome for the i-Ready (treatment) and comparison groups for Hispanic Origin and non-Hispanic Origin students at grade 4.

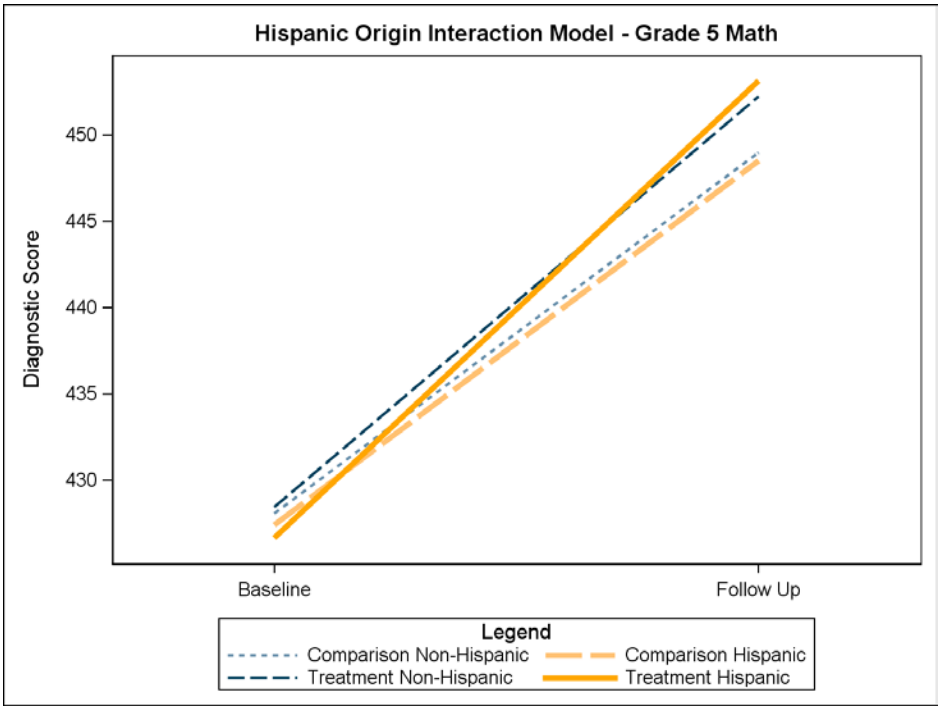

Figure H.4. Gains in Mathematics Diagnostic achievement between baseline and outcome for the i-Ready (treatment) and comparison groups for Hispanic Origin and non-Hispanic Origin groups at grade 5.

Appendix I
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 baseline analytic model for all four 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.