i-Ready Personalized Instruction Longitudinal Impact Study – Mathematics

Final Report

Prepared for: Curriculum Associates

Date November 5, 2025

Authors: Caroline R.H. Wiley
Meng Fan
Arthur Thacker
Dea Mulolli
Hye-Jeong Choi

i-Ready Personalized Instruction Longitudinal Impact Study – Mathematics

Table of Contents

Table of Contents (Continued)

List of Tables

List of Appendix Tables

List of Figures

i-Ready Personalized Instruction Longitudinal Impact Study – Mathematics

Summary

This study examined the long-term impact of *i-Ready Personalized Instruction* on students' Grade 5 math achievement. We followed nearly 10,000 students across 13 districts in 10 states who used both the *i-Ready Diagnostic* assessment and *i-Ready Personalized Instruction* from Grade 3 through Grade 5 and compared them with matched peers who used the *i-Ready Diagnostic* assessment only. Using a quasi-experimental design, we matched grade 3 students on the fall *i-Ready Diagnostic* assessment in SY2021-22 and key demographics to estimate the effect of *i-Ready Personalized Instruction* on statewide assessment outcomes. After matching, groups were well-balanced and the analysis accounted for students being nested within schools.

While the impact on student achievement after one year and after two years of usage was positive but statistically non-significant, students who consistently used *i-Ready Personalized Instruction* with fidelity across all three years (grades 3 through 5) scored 0.24 standard deviation units higher than students who did not use the program. This represents a statistically significant and educationally meaningful effect (Hedges' $g = 0.24$, $p = 0.02$). The 0.24 treatment effect translates to gains of approximately 5 to 24 scale score points across the participating states, depending on each state's scaling properties.

The practical significance of these findings becomes evident when translating the treatment effect to state assessment scales. These translated scores provide educators, policymakers, and other stakeholders with concrete benchmarks for interpreting the program's impact in familiar terms. When sustained over multiple years, these gains represent meaningful progress that can move students to higher performance levels (e.g., Basic to Proficient) or help close achievement gaps.

Introduction

The Human Resources Research Organization (HumRRO) contracted with Curriculum Associates to evaluate the longitudinal impact *i-Ready Personalized Instruction* has on student mathematics achievement. *i-Ready Personalized Instruction* is a system of personalized lessons designed to fill students' knowledge gaps and help every student reach grade-level proficiency through engaging instruction for students of all abilities and ages. This personalized instruction is powered by data from *i-Ready*'s comprehensive diagnostic assessment system. The *i-Ready Diagnostic* assessments are adaptive assessments that provide a detailed picture of student performance and growth while pinpointing each student's ability level, identifying skill development needs, and delivering personalized learning paths with actionable insights for teachers.

Research Questions

The primary research question addressed:

1. What long-term impact on mathematics achievement (as measured by state summative assessments) does *i-Ready Personalized Instruction* (*PI*) have for students who used the *i-Ready Diagnostic* assessments and *i-Ready Personalized Instruction* for three years compared to those who used only the *i-Ready Diagnostic* assessments and how does implementation fidelity affect these results?
   a. The study also examined short-term (1 year of use) and intermediate-term (2 years of use) impacts.

Methods

This study employed a cluster-level Quasi-Experimental Design (QED) to evaluate the effects of *i-Ready Personalized Instruction* in public elementary schools on state summative mathematics assessments. The study examined long-term (3-year usage), intermediate-term (2-year usage), and short-term (1-year usage) effects by analyzing three distinct student cohorts:

- Cohort A (Grade 3 in SY2021-22):
  - Long-term: Grade 3 students followed through Grade 5 in SY2023-24
  - Intermediate-term: Grade 3 students followed through Grade 4 in SY2022-23
  - Short-term: Grade 3 students followed through end of Grade 3 in SY2021-22

- Cohort B (Grade 3 in SY2022-23):
  - Intermediate-term: Grade 3 students followed through Grade 4 in SY2023-24
  - Short-term: Grade 3 students followed through end of Grade 3 in SY2022-23

- Cohort C (Grade 3 in SY2023-24):
  - Short-term: Grade 3 students followed through end of Grade 3 in SY2023-24

We included three distinct cohorts to examine the three types of distal effects. Cohort A served as our primary focus for analyses because it uniquely provides long-term (3-year) program impact data. Cohort A is the only cohort that offers comprehensive insight into the sustained *i-Ready Personalized Instruction* effect across the elementary progression from grade 3 through grade 5. All primary findings and conclusions in this study are based on Cohort A analyses.

Findings for Cohorts B and C are considered supplemental and are provided in appendix tables referenced throughout the report.

Treatment and Comparison Group Definitions

Across all cohorts, participating students in both treatment and comparison groups attended schools that met the following requirements:

- Used the *i-Ready Diagnostic* assessment in grades 3, 4, and 5.

- Minimum of 10 eligible students per grade level.

Participating students in the *i-Ready Personalized Instruction* usage group met the additional requirement of using *i-Ready Personalized Instruction* in grades 3, 4, and 5. Additionally, Curriculum Associates provided recommended student-level usage guidelines, which they operationalized into a fidelity of implementation metric defined as an average *i-Ready Personalized Instruction* usage rate of at least 30 minutes per week for at least 18 distinct weeks, with a lesson pass rate of at least 70% for the school year.

Because implementation fidelity is critical to ensure that observed outcomes can be accurately attributed to the intended intervention rather than to variations in how the program was delivered, we defined four distinct *i-Ready Personalized Instruction* usage groups for Cohort A:

- Condition X: Ignored implementation fidelity and only required that students used *i-Ready Personalized Instruction.*

- Condition 1: Required that students used *i-Ready Personalized Instruction* for at least 1 year with fidelity.

- Condition 2: Required that students used *i-Ready Personalized Instruction* for at least 2 years with fidelity.

- Condition 3: Required that students used *i-Ready Personalized Instruction* all 3 years with fidelity.

Table 1 provides an overview of the research design and treatment conditions for each cohort.

Table 1. i-Ready Personalized Instruction (PI) Treatment Group Design by Cohort

Cohort	School Year	Usage Duration	PI Condition
Cohort A	SY2021-22 – SY2023-24	Long-term (3-yr usage): Grade 3 to 5	Condition X: *PI* users (any fidelity) Condition 1: *PI* + 1 or more years of fidelity Condition 2: *PI* + 2 or more years of fidelity Condition 3: *PI* + 3 years fidelity
		Intermediate-term (2-yr usage): Grade 3 to 4	Condition X: *PI* users (any fidelity) Condition 1: *PI* + 1 or more years of fidelity Condition 2: *PI* + 2 or more years of fidelity
		Short-term (1-yr usage): Grade 3 in SY2021-22	Condition X: *PI* users (any fidelity) Condition 1: *PI* + 1 year fidelity
Cohort B	SY2022-23 – SY2023-24	Intermediate-term (2-yr usage): Grade 3 to 4	Condition X: *PI* users (any fidelity) Condition 1: *PI* + 1 or more years of fidelity Condition 2: *PI* + 2 or more years of fidelity
		Short-term (1-yr usage): Grade 3 in SY2022-23	Condition X: *PI* users (any fidelity) Condition 1: *PI* + 1 year of fidelity
Cohort C	SY2023-24	Short-term (1-yr usage): Grade 3 in SY2023-24	Condition X: *PI* users (any fidelity) Condition 1: *PI* + 1 year of fidelity

Note. Cohort A = primary analyses. Cohorts B & C = supplemental analyses.

Data Sources

Curriculum Associates recruited schools based on preliminary school usage data and entered into data sharing agreements with participating districts that provided student-level state assessment scores for SY2021-22, SY2022-23, and SY2023-24. Following initial data cleaning and processing by Curriculum Associates, the prepared datasets were transmitted to HumRRO for analysis. The final data sources included *i-Ready Diagnostic* assessment scores, *i-Ready Personalized Instruction* usage data, state assessment scores, and student demographic information.

Eligible Sample

For Cohort A, we received data from just over 9,900 grade 3 students from 13 districts across 10 states for SY2021-22. Table 2 presents the eligible sample sizes for Cohort A across different usage durations and treatment conditions after removing ineligible students. The comparison group (No *i-Ready Personalized Instruction* usage) remained consistent across all analyses, drawing from four states with 61 schools and 2,110 students. The treatment group composition (*i-Ready Personalized Instruction* usage) varied by fidelity requirements: the least restrictive condition (Condition X) included the largest sample (3,552 students), while the most stringent condition (Condition 3) had the smallest treatment group (784 students). Descriptive statistics for the eligible samples of Cohorts B and C can be found in Appendix Tables A1-A2.

Table 2. Descriptive Statistics for the Eligible Cohort A Sample

Usage Duration	PI Condition	Group	States	N of States	N of Schools	N of Students
Long	Condition X	No *PI* Usage	CO, NY, OH & WI	4	61	2,110
		PI Usage	CO, CT, IL, KY, MO, NM, NY & WI	8	73	3,552
	Condition 1	No *PI* Usage	CO, NY, OH & WI	4	61	2,110

Usage Duration	*PI* Condition	Group	States	N of States	N of Schools	N of Students
		PI Usage	CO, CT, IL, KY, MO, NM, NY & WI	8	73	2,968
	Condition 2	No *PI* Usage	CO, NY, OH & WI	4	61	2,110
		PI Usage	CO, CT, IL, KY, MO, NM, NY & WI	8	71	1,931
	Condition 3	No *PI* Usage	CO, NY, OH & WI	4	61	2,110
		PI Usage	CO, CT, IL, KY, MO, NM, NY & WI	8	60	784
Intermediate	Condition X	No *PI* Usage	CO, NY, OH & WI	4	61	2,110
		PI Usage	CO, CT, IL, KY, MO, NY & WI	7	49	2,412
	Condition 1	No *PI* Usage	CO, NY, OH & WI	4	61	2,110
		PI Usage	CO, CT, IL, KY, MO, NY & WI	7	48	1,749
	Condition 2	No *PI* Usage	CO, NY, OH & WI	4	61	2,110
		PI Usage	CO, CT, IL, KY, MO, NY & WI	7	42	816
Short	Condition X	No *PI* Usage	CO, NY, OH & WI	4	61	2,110
		PI Usage	CO, CT, IL, KY, MO, NY & WI	7	49	2,412
	Condition 1	No *PI* Usage	CO, NY, OH & WI	4	61	2,110
		PI Usage	CO, CT, IL, KY, MO, NY & WI	7	45	1,103

Table 3 shows students' demographic characteristics before matching for Cohort A's long-term usage condition. The most substantial difference was in racial composition: *i-Ready Personalized Instruction* groups were predominantly White students (70-80%) compared to 47% in the comparison group, representing large effect sizes (Hedges' g = 0.64 to 0.71). English Learner (EL) status also showed notable differences, with the comparison group having higher EL rates (Hedges' g = -0.21 to -0.28). Economically Disadvantaged (ECD) status was moderately imbalanced, with more economically disadvantaged students in the comparison group (Hedges' g = -0.18 to -0.32). Students with Disabilities (SWD) were similarly distributed across all groups, with minimal differences (Hedges' g = 0.04 to 0.06). Similar demographic patterns were observed for the intermediate-term and short-term usage durations (see Appendix Tables B1-B2).

Table 3. Cohort A Covariate Descriptives Before Matching (Long-Term)

PI Condition	Variables	PI Usage		No PI Usage		Difference	Effect Size (Hedges' *g*)
		N	Percent	N	Percent		
Condition X	White	3,552	80.4%	2,110	47.4%	33.0%	0.71
	EL	3,552	7.0%	2,110	14.9%	-7.9%	-0.26
	SWD	3,552	21.1%	2,110	16.8%	4.3%	0.11
	ECD	3,552	41.9%	2,110	55.3%	-13.3%	-0.27
Condition 1	White	2,968	80.5%	2,110	47.4%	33.1%	0.70
	EL	2,968	6.9%	2,110	14.9%	-8.0%	-0.26
	SWD	2,968	20.0%	2,110	16.8%	3.2%	0.08
	ECD	2,968	42.8%	2,110	55.3%	-12.4%	-0.25
Condition 2	White	1,931	80.0%	2,110	47.4%	32.6%	0.68
	EL	1,931	6.5%	2,110	14.9%	-8.4%	-0.27
	SWD	1,931	19.0%	2,110	16.8%	2.1%	0.06
	ECD	1,931	44.4%	2,110	55.3%	-10.8%	-0.22
Condition 3	White	784	79.0%	2,110	47.4%	31.6%	0.64
	EL	784	5.9%	2,110	14.9%	-9.1%	-0.27
	SWD	784	19.5%	2,110	16.8%	2.7%	0.07
	ECD	784	46.0%	2,110	55.3%	-9.2%	-0.18

Note. EL = English Learner, SWD = Students with Disabilities, ECD = Economically Disadvantaged

Matching

What Works Clearinghouse (WWC) recognizes matching (i.e., propensity score matching, genetic matching) as an acceptable baseline adjustment strategy (WWC, 2022). We utilized genetic matching with Mahalanobis distance at the student level to ensure balanced comparison groups and reduce selection bias. Genetic matching is an advanced matching technique that uses a genetic algorithm to optimize the balance between treatment and comparison groups across multiple covariates (Diamond & Sekhon, 2013). We employed 1:1 matching and assessed balance using standardized mean differences, with values below 0.25 considered acceptable balance in accordance with WWC guidelines. Students who could not be adequately matched were excluded from the analysis.

Students were matched using genetic matching with Mahalanobis distance on their Year 1[1] fall *i-Ready Diagnostic* assessment and key demographic characteristics (i.e., ECD status), EL status, SWD status, and race/ethnicity (*White* vs. *non-White*). We used *i-Ready Diagnostic* scores as the baseline academic achievement measure since grade 2 state assessments were not available. This approach has been validated by Curriculum Associates (2025), which found high correlations between the *i-Ready Diagnostic* and state assessments in 46 states (above 0.70).

---

[1] Year 1 refers to each cohort's respective start year (Cohort A's start year is SY2021-22, Cohort B's start year is SY2022-23, and Cohort C's start year is SY2023-24)

Before Matching

Table 4 presents baseline equivalence on the grade 3 fall *i-Ready Diagnostic* mathematics assessment before matching for Cohort A across all usage durations and implementation fidelity conditions. Prior to matching, imbalances existed between *i-Ready Personalized Instruction* usage groups and comparison groups, with treatment groups consistently demonstrating higher baseline *i-Ready Diagnostic* assessment scores. Effect sizes ranged from 0.14 to 0.43, indicating small to moderate differences that exceed acceptable thresholds for baseline equivalence.

Table 4. Baseline Equivalence Before Matching (i-Ready Diagnostic assessment)

Usage Duration	PI Condition	No PI N	No PI Mean	No PI SD	PI N	PI Mean	PI SD	Mean Difference	Hedges' g
Long	Condition X	2,110	415.37	30.77	3,552	419.29	26.05	3.92	0.14
	Condition 1	2,110	415.37	30.77	2,968	420.96	25.35	5.59	0.20
	Condition 2	2,110	415.37	30.77	1,931	423.33	24.62	7.96	0.28
	Condition 3	2,110	415.37	30.77	784	427.02	23.41	11.65	0.40
Intermediate	Condition X	2,110	415.37	30.77	2,412	423.57	25.33	8.20	0.29
	Condition 1	2,110	415.37	30.77	1,749	425.44	24.57	10.07	0.36
	Condition 2	2,110	415.37	30.77	816	427.72	23.95	12.36	0.43
Short	Condition X	2,110	415.37	30.77	2,412	423.57	25.33	8.20	0.29
	Condition 1	2,110	415.37	30.77	1,103	425.00	25.28	9.63	0.33

Note. According to WWC Standards Version 5.0, baseline equivalence is considered met if |g| <= 0.25. Differences above this threshold indicate imbalance prior to matching.

After Matching

Genetic matching successfully achieved satisfactory covariate balance across all treatment conditions, with all effect sizes (Hedge's g) well below the WWC threshold of 0.25. Table 5 displays the baseline equivalence of mathematics performance after matching. The matching procedure removed almost all the differences between treatment and comparison groups, with effect sizes below 0.05 across the three usage durations and *i-Ready Personalized Instruction* conditions.

Table 5. Baseline Equivalence After Matching (i-Ready Diagnostic assessment)

Usage Duration	PI Condition	No PI N	No PI Mean	No PI SD	PI N	PI Mean	PI SD	Mean Difference	Hedges' g
Long	Condition X	1,566	422.90	25.60	1,558	422.99	25.58	0.09	0.00
	Condition 1	1,473	424.23	24.88	1,470	424.26	24.87	0.03	0.00
	Condition 2	1,278	426.11	23.29	1,262	425.91	23.32	-0.20	-0.01
	Condition 3	665	428.91	21.43	676	427.86	21.57	-1.05	-0.05
Intermediate	Condition X	1,514	423.59	24.92	1,522	423.63	24.90	0.04	0.00
	Condition 1	1,310	425.84	23.54	1,318	425.72	23.60	-0.12	0.00
	Condition 2	717	428.81	21.86	739	428.11	21.87	-0.69	-0.03
Short	Condition X	1,514	423.59	24.92	1,522	423.63	24.90	0.04	0.00
	Condition 1	976	425.45	23.30	988	425.22	23.27	-0.23	-0.01

Note. According to WWC Standards Version 5.0, baseline equivalence is considered met if |g| <= 0.25. Differences above this threshold require covariate adjustment in the impact model.

The matching procedure successfully achieved covariate balance across all demographic variables. Most notably, the large racial/ethnic disparity observed before matching was effectively reduced, with all effect sizes falling below 0.25. Following matching, all demographic variables showed effect sizes within acceptable ranges, providing a strong foundation for making causal inferences regarding treatment effects. Covariate balance results for all three Cohort A usage durations are presented in Appendix Tables C1-C3.

Impact Model

The outcome measure was standardized z-scores from state summative mathematics assessments from each cohort's final year. This transformation enabled us to compare students' state assessment scores from different states. We employed a three-level Hierarchical Linear Model (HLM) with random intercepts to account for the nested structure of the data, where students (level 1) are nested within schools (level 2), which in turn are nested within states (level 3). At level 1, the model includes the treatment indicator (*i-Ready Personalized Instruction*) and student-level covariates (baseline fall *i-Ready Diagnostic* score, ECD status, EL status, SWD status, and race/ethnicity[2]). This specification allows for robust estimation of the treatment effect while accounting for clustering at both school and state levels. We applied the following HLM model for each treatment condition:

$$Outcome_{ijk} = \beta_0 + \beta_1(iReady\ Personalized\ Instruction_{ijk}) + \beta_2(iReady\ Diagnostic_{ijk})$$
$$+ \beta_3(ECD_{ijk}) + \beta_4(EL_{ijk}) + \beta_5(SWD_{ijk}) + \beta_6(White_{ijk}) + \mu_{0jk} + \Upsilon_{0k} + e_{ijk}$$

where $Outcome_{ijk}$ is the outcome variable for student *i* in school *j* in state *k*. β_0 is the intercept of the model. β_1 is the PI effect. β_{2-6} are the fixed effects of student-level covariables. μ_{0jk} is the random intercept at the school level. Υ_{0k} is the random intercept at the state level. e_{ijk} is the

---

[2] Covariate scales were: *i-Ready Personalized Instruction* (0 = non-user, 1 = user); *i-Ready Diagnostic* = vertical scale scores; ECD, EL, SWD, and White were all binary coded (0 = No, 1 = Yes).

residual error term at the student level. Table 6 presents the baseline and outcome assessments across cohorts and usage terms.

Table 6. Description of Baseline and Outcome Measures

Cohorts	Usage Duration	Baseline Measure	Outcome Measure (z-score)
Cohort A	Long	Grade 3 fall *i-Ready Diagnostic* (SY2021-22)	Grade 5 state assessment (SY2023-24)
	Intermediate	Grade 3 fall *i-Ready Diagnostic* (SY2021-22)	Grade 4 state assessment (SY2022-23)
	Short	Grade 3 fall *i-Ready Diagnostic* (SY2021-22)	Grade 3 state assessment (SY2021-22)
Cohort B	Intermediate	Grade 3 fall *i-Ready Diagnostic* (SY2022-23)	Grade 4 state assessment (SY2023-24)
	Short	Grade 3 fall *i-Ready Diagnostic* (SY2022-23)	Grade 3 state assessment (SY2022-23)
Cohort C	Short	Grade 3 fall *i-Ready Diagnostic* (SY2023-24)	Grade 3 state assessment (SY2023-24)

Results

This section presents the results for Cohort A. Before examining treatment effects, we calculated Intraclass Correlation Coefficients (ICCs) to quantify clustering in our outcome data and validate our analytical approach. ICCs indicate how much variance in student outcomes is attributable to higher-level units (schools and states), which directly impacts treatment effect precision. High ICCs suggest students within the same school or state perform more similarly to each other, necessitating multilevel modeling to avoid biased standard errors.

The unconditional ICC analysis (see Appendix Table D1) revealed meaningful clustering at both levels. School-level ICCs ranged from 0.17 to 0.21 (17-21% of variance), while state-level ICCs ranged from 0.07 to 0.14. After including covariates, ICCs decreased but remained meaningful (school: 0.07-0.17; state: 0.00-0.11). This reduction demonstrates that our covariates explained considerable between-cluster variance, improving treatment effect precision. These ICC values strongly supported our three-level HLM approach, confirming meaningful clustering that would violate independence assumptions of single-level analyses.

Long-term (3-year usage)

Figure 1 (and Table 7) shows that the long-term effects of using *i-Ready Personalized Instruction* for three continuous years (grade 3 through grade 5) substantially increased state summative mathematics assessment scores when implemented with fidelity all three years.

Figure 1. 3-Year Impact of i-Ready Personalized Instruction Usage Across Treatment Fidelity Conditions (Cohort A)
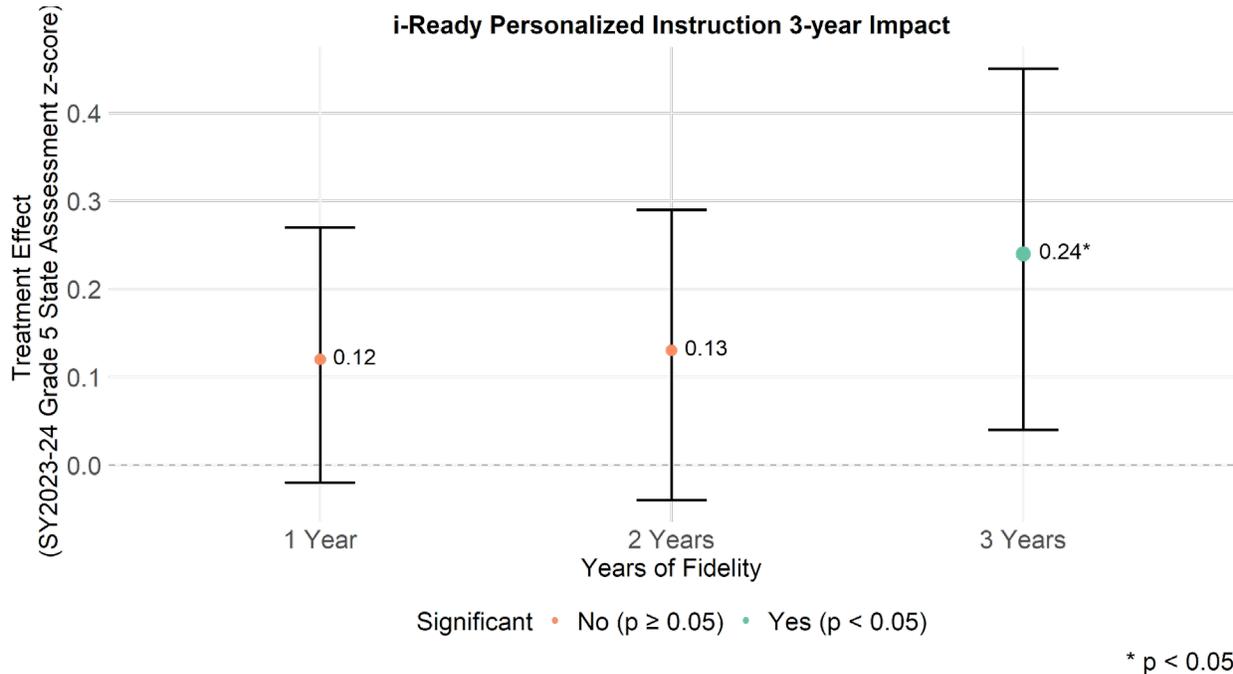
Note. Years of fidelity category indicates whether students *used i-Ready Personalized Instruction* with at least 1 year of fidelity, 2 years of fidelity, or 3 years of fidelity.

Students with any *i-Ready Personalized Instruction* use regardless of fidelity (Condition X) showed no meaningful impact (b = 0.07, p = 0.27), demonstrating that *i-Ready Personalized Instruction* use alone without sustained fidelity is insufficient for achievement gains. Similarly, students meeting fidelity requirements for only one year showed minimal impact (b = 0.12, p = 0.10). The 2-year fidelity condition demonstrated modest improvement (b = 0.13, p = 0.14), though effects remained non-significant. However, students who maintained high fidelity across all 3 years showed a statistically significant and educationally meaningful effect (b = 0.24, p = 0.02). This effect size (Hedges' g = 0.24) exceeds the median effect size (0.10) found in educational research (Kraft, 2019), providing strong evidence of PI's potential when implemented with consistent fidelity over the full study period.

Table 7. Impact of Long-term i-Ready Personalized Instruction Usage (3 years) Across Treatment Fidelity Conditions (Cohort A)

Usage Duration	PI Condition	Coefficient (b)	Standard Error	df	p	Effect Size (Hedges' g)
Long	Condition X	0.07	0.07	96	0.27	0.07
	Condition 1	0.12	0.07	79	0.10	0.12
	Condition 2	0.13	0.08	76	0.14	0.13
	Condition 3	**0.24**	**0.10**	**61**	**0.02**	**0.24**

Note. Full impact model results are presented in the Appendix Table E1.

Further, we transformed the treatment effect (0.24) of the 3-fidelity *i-Ready Personalized Instruction* usage condition to scale scores for each participating state to show the scale score

gains in their more familiar metric. Table 8 shows that the scale score gains range from 4.87 points (New Mexico) to 24.22 points (Connecticut). Most states demonstrated gains between 5 and 10 points. Note that the variation of score gains does not indicate different treatment effects across states. States with larger score ranges and standard deviations show larger scale score gains, but all gains represent the same 0.24 treatment effect.

Table 8. Predicted Mathematics Achievement Gains from Three Years of i-Ready Personalized Instruction with Fidelity

State	Coefficient	State Scale Score Change
CO	0.24	8.24
CT	0.24	24.22
IL	0.24	8.18
KY	0.24	4.97
MO	0.24	10.26
NY	0.24	6.67
OH	0.24	9.52
WI	0.24	12.41
NM	0.24	4.87

Note. All gains represent the same 0.24 treatment effect; variation reflects differences in state scale score ranges and standard deviations.

Intermediate-term (2-year usage) and Short-term (1-year usage)

Intermediate-term usage showed non-significant effects ranging from -0.09 to 0.03 across the four *i-Ready Personalized Instruction* implementation fidelity conditions. Short-term usage also yielded non-significant results (Hedges' g = 0.04 - 0.05), indicating insufficient evidence to detect meaningful effects from shorter usage periods. These findings (Table 9) suggest that sustained, high-fidelity implementation over three years may be necessary to achieve detectable impacts on mathematics achievement. Full impact model results are presented in the Appendix Tables E2-E3.

Table 9. Impact of Intermediate-term (2 year) and Short-Term (1 year) i-Ready Personalized Instruction Usage Across Treatment Fidelity Conditions

Usage Duration	*PI* Condition	Coefficient (b)	Standard Error	df	*p*	Effect Size (Hedges' *g*)
Intermediate	Condition X	-0.09	0.05	48	0.10	-0.09
	Condition 1	-0.02	0.06	34	0.70	-0.02
	Condition 2	0.03	0.08	13	0.71	0.03
Short	Condition X	0.05	0.05	77	0.31	0.05
	Condition 1	0.04	0.06	14	0.56	0.04

Results for Cohorts B and C showed similar patterns of non-significant effects for intermediate-term and short-term *i-Ready Personalized Instruction* usage and are presented in Appendix Tables F1-F3. These consistent findings across all three cohorts strengthen the evidence that longer implementation periods with sustained fidelity are necessary for meaningful achievement gains.

Discussion and Future Directions

The current study examined the longitudinal effects of *i-Ready Personalized Instruction* on students' mathematics achievement on state summative assessments across nine states from grade 3 through grade 5. Employing a rigorous quasi-experimental design with genetic matching to achieve baseline equivalence and three-level hierarchical linear modeling to account for clustering effects, we analyzed three distinct student cohorts to examine long-term, intermediate-term, and short-term usage effects.

Our findings reveal a clear pattern: the effectiveness *of i-Ready Personalized Instruction* depends on both usage duration and fidelity strength (i.e., number of years used with fidelity). Students who maintained consistent usage with fidelity across all 3 years (Grades 3 through 5) demonstrated a statistically significant and educationally meaningful effect (Hedges' g = 0.24, p = 0.02). This effect size exceeds the median impact found in educational intervention research (Kraft, 2019) and represents substantial learning gains. In contrast, students with any level of *i-Ready Personalized Instruction* usage but without sustained fidelity showed negligible effects (Hedges' g = 0.07), while those meeting fidelity requirements for only 1 or 2 years produced minimal, non-significant impacts. These results demonstrate that brief or inconsistent usage of *i-Ready Personalized Instruction* may not be sufficient to produce statistically detectable achievement gains.

The practical significance of these findings becomes evident when translating effect sizes to state assessment scales. The 0.24 treatment effect represents concrete gains ranging from 5 to 24 scale score points across the nine participating states, depending on each state's assessment characteristics and scaling properties. These translated scores provide educators, policymakers, and other stakeholders with concrete benchmarks for understanding program impact in familiar metric units. For context, these gains represent meaningful progress that could move students to higher performance levels (e.g., Basic to Proficient) or help close achievement gaps when sustained over multiple years.

References

Curriculum Associates. (2025). *i-Ready diagnostic correlations: Correlations between i-Ready diagnostic and consortium and state summative assessments*. Curriculum Associates. Retrieved from https://cdn.bfldr.com/LS6J0F7/at/fp638hqx8vx56s6cbtjchp9/iready-correlations.pdf

Diamond, A., & Sekhon, J. S. (2013). Genetic matching for estimating causal effects: A general multivariate matching method for achieving balance in observational studies. *The Review of Economics and Statistics, 95*(3), 932–945

Kraft, M. A. (2019). *Interpreting effect sizes of education interventions*. Harvard University. https://scholar.harvard.edu/files/mkraft/files/kraft_2019_effect_sizes.pdf

What Works Clearinghouse. (2022). *What Works Clearinghouse procedures and standards handbook, version 5.0*. U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance. Retrieved from https://ies.ed.gov/ncee/wwc/handbooks

Appendix A: Eligible Sample Descriptive Statistics – Cohorts B and C

Table A1. Cohort B: Descriptive Statistics for the Eligible Cohort B Sample

Usage Duration	PI Condition	Group	States	N of States	N of Schools	N of Students
Intermediate	Condition X	No *PI* Usage	CO, NY, OH & WI	4	61	2,411
		PI Usage	CO, CT, IL, KY, MO, NM, NY & WI	8	76	4,464
	Condition 1	No *PI* Usage	CO, NY, OH & WI	4	61	2,411
		PI Usage	CO, CT, IL, KY, MO, NM, NY & WI	8	75	3,462
	Condition 2	No *PI* Usage	CO, NY, OH & WI	4	61	2,411
		PI Usage	CO, CT, IL, KY, MO, NM, NY & WI	8	67	1,709
Short	Condition X	No *PI* Usage	CO, NY, OH & WI	4	61	2,411
		PI Usage	CO, CT, IL, KY, MO, NY & WI	7	51	3,113
	Condition 1	No *PI* Usage	CO, NY, OH & WI	4	61	2,411
		PI Usage	CO, CT, IL, KY, MO, NY & WI	7	49	1,904

Table A2. Cohort C: Descriptive Statistics for the Eligible Cohort C Sample

Usage Term	Treatment Condition	Group	States	N of States	N of Schools	N of Students
Short	Condition X	No *PI* Usage	CO, NY & WI	3	27	2,768
		PI Usage	CO, CT, IL, KY, MO, NM, NY & WI	8	38	5,202
	Condition 1	No *PI* Usage	CO, NY & WI	3	27	2,768
		PI Usage	CO, CT, IL, KY, MO, NM, NY & WI	8	26	3,182

Appendix B: Covariate Descriptives Before Matching (Intermediate- and Short-Term) – Cohort A

Table B1. Cohort A: Covariate Descriptives Before Matching (Intermediate-Term)

PI Condition	Variables	PI Usage N	PI Usage Percent	No PI Usage N	No PI Usage Percent	Difference	Effect Size (Hedges' g)
Condition X	White	2,412	72.5%	2,110	47.4%	25.1%	0.51
	ELL	2,412	6.4%	2,110	14.9%	-8.5%	-0.28
	SWD	2,412	22.6%	2,110	16.8%	5.7%	0.14
	ECD	2,412	39.4%	2,110	55.3%	-15.8%	-0.32
Condition 1	White	1,749	73.3%	2,110	47.4%	25.9%	0.53
	ELL	1,749	6.6%	2,110	14.9%	-8.4%	-0.27
	SWD	1,749	21.3%	2,110	16.8%	4.4%	0.11
	ECD	1,749	42.9%	2,110	55.3%	-12.4%	-0.25
Condition 2	White	816	74.0%	2,110	47.4%	26.6%	0.53
	ELL	816	6.3%	2,110	14.9%	-8.7%	-0.26
	SWD	816	22.1%	2,110	16.8%	5.2%	0.14
	ECD	816	45.2%	2,110	55.3%	-10.0%	-0.20

Table B2. Cohort A: Covariate Descriptives Before Matching (Short-Term)

PI Condition	Variables	PI Usage N	PI Usage Percent	No PI Usage N	No PI Usage Percent	Difference	Effect Size (Hedges' g)
Condition X	White	2,412	72.5%	2,110	47.4%	25.1%	0.51
	ELL	2,412	6.4%	2,110	14.9%	-8.5%	-0.28
	SWD	2,412	22.6%	2,110	16.8%	5.7%	0.14
	ECD	2,412	39.4%	2,110	55.3%	-15.8%	-0.32
Condition 1	White	1,103	69.7%	2,110	47.4%	22.3%	0.45
	ELL	1,103	8.1%	2,110	14.9%	-6.9%	-0.21
	SWD	1,103	23.2%	2,110	16.8%	6.4%	0.16
	ECD	1,103	47.6%	2,110	55.3%	-7.7%	-0.15

Appendix C: Covariate Descriptives After Matching – Cohort A

Table C1. Cohort A: Covariate Descriptives After Matching (Long-Term)

PI Condition	Variables	PI Usage		No PI Usage		Difference	Effect Size (Hedges' g)
		N	Percent	N	Percent		
Condition X	White	1,558	62.3%	1,566	62.2%	0.1%	0.00
	ELL	1,558	9.8%	1,566	9.1%	0.7%	0.02
	SWD	1,558	19.6%	1,566	16.0%	3.6%	0.09
	ECD	1,558	42.7%	1,566	45.1%	-2.4%	-0.05
Condition 1	White	1,470	65.5%	1,473	65.6%	-0.1%	0.00
	ELL	1,470	9.3%	1,473	8.4%	0.9%	0.03
	SWD	1,470	18.5%	1,473	16.0%	2.5%	0.07
	ECD	1,470	43.9%	1,473	43.2%	0.7%	0.01
Condition 2	White	1,262	71.9%	1,278	72.4%	-0.4%	-0.01
	ELL	1,262	7.6%	1,278	7.2%	0.4%	0.02
	SWD	1,262	18.0%	1,278	15.0%	3.0%	0.08
	ECD	1,262	44.8%	1,278	40.0%	4.8%	0.10
Condition 3	White	676	78.7%	665	82.9%	-4.2%	-0.11
	ELL	676	5.8%	665	4.1%	1.7%	0.08
	SWD	676	20.0%	665	13.1%	6.9%	0.19
	ECD	676	46.2%	665	36.8%	9.3%	0.19

Table C2. Cohort A: Covariate Descriptives After Matching (Intermediate-Term)

PI Condition	Variables	PI Usage		No PI Usage		Difference	Effect Size (Hedges' g)
		N	Percent	N	Percent		
Condition X	White	1,522	62.9%	1,514	62.8%	0.1%	0.00
	ELL	1,522	9.0%	1,514	8.9%	0.2%	0.01
	SWD	1,522	21.9%	1,514	15.7%	6.3%	0.16
	ECD	1,522	41.3%	1,514	44.5%	-3.2%	-0.06
Condition 1	White	1,318	68.0%	1,310	68.2%	-0.2%	0.00
	ELL	1,318	7.9%	1,310	7.8%	0.1%	0.00
	SWD	1,318	20.7%	1,310	14.9%	5.8%	0.15
	ECD	1,318	43.9%	1,310	42.0%	1.9%	0.04
Condition 2	White	739	74.3%	717	76.8%	-2.6%	-0.06
	ELL	739	6.1%	717	4.9%	1.2%	0.05
	SWD	739	22.2%	717	13.0%	9.2%	0.24
	ECD	739	45.7%	717	38.8%	7.0%	0.14

Table C3. Cohort A: Covariate Descriptives After Matching (Short-Term)

PI Condition	Variables	PI Usage		No PI Usage		Difference	Effect Size (Hedges' g)
		N	Percent	N	Percent		
Condition X	White	1,522	62.9%	1,514	62.8%	0.1%	0.00
	ELL	1,522	9.0%	1,514	8.9%	0.2%	0.01
	SWD	1,522	21.9%	1,514	15.7%	6.3%	0.16
	ECD	1,522	41.3%	1,514	44.5%	-3.2%	-0.06
Condition 1	White	988	68.2%	976	69.0%	-0.7%	-0.02
	ELL	988	8.7%	976	7.5%	1.2%	0.04
	SWD	988	23.1%	976	14.4%	8.6%	0.22
	ECD	988	48.5%	976	44.0%	4.5%	0.09

Appendix D: Intraclass Correlation Coefficients – Cohort A

Table D1. Cohort A: Intraclass Correlation Coefficient for Each Condition

Usage Duration	*PI* Condition	ICC Model	School	State
Long	Condition X	Unconditional	0.17	0.12
		Full	0.07	0.05
	Condition 1	Unconditional	0.20	0.12
		Full	0.11	0.06
	Condition 2	Unconditional	0.19	0.14
		Full	0.13	0.09
	Condition 3	Unconditional	0.18	0.13
		Full	0.14	0.11
Intermediate	Condition X	Unconditional	0.18	0.10
		Full	0.17	0.02
	Condition 1	Unconditional	0.19	0.09
		Full	0.08	0.02
	Condition 2	Unconditional	0.19	0.09
		Full	0.13	0.02
Short	Condition X	Unconditional	0.21	0.07
		Full	0.11	0.00
	Condition 1	Unconditional	0.20	0.09
		Full	0.11	0.02

Appendix E: Full Impact Model Results – Cohort A

Table E1. Cohort A: Long-Term Effects of i-Ready Personalized Instruction on Mathematics Achievement by Fidelity Condition

PI Condition	Variables	Coefficients	Standard Error	df	*p*
Condition X	*i-Ready PI Usage*	0.07	0.07	96	0.27
	i-Ready Diagnostic	**0.03**	0.00	3112	0.00*
	White	**0.10**	0.03	2972	0.00*
	ELL	0.01	0.05	2840	0.82
	SWD	**-0.17**	0.03	3092	0.00*
	ECD	**-0.16**	0.03	3071	0.00*
Condition 1	*i-Ready PI Usage*	0.12	0.07	79	0.10
	i-Ready Diagnostic	**0.03**	0.00	2930	0.00*
	White	**0.11**	0.03	2863	0.00*
	ELL	0.07	0.05	2759	0.15
	SWD	**-0.12**	0.03	2894	0.00*
	ECD	**-0.15**	0.03	2897	0.00*
Condition 2	*i-Ready PI Usage*	0.13	0.08	76	0.14
	i-Ready Diagnostic	**0.03**	0.00	2525	0.00*
	White	**0.11**	0.03	2482	0.00*
	ELL	0.13	0.05	2395	0.02*
	SWD	**-0.10**	0.03	2480	0.00*
	ECD	**-0.16**	0.03	2520	0.00*
Condition 3	*i-Ready PI Usage*	**0.24**	0.10	61	0.02*
	i-Ready Diagnostic	**0.03**	0.00	1321	0.00*
	White	**0.22**	0.05	1330	0.00*
	ELL	**0.16**	0.09	1258	0.06
	SWD	-0.07	0.04	1303	0.14
	ECD	**-0.14**	0.04	1321	0.00*

Table E2. Cohort A: Intermediate-Term Effects of i-Ready Personalized Instruction on Mathematics Achievement by Fidelity Condition

PI Condition	Variables	Coefficients	Standard Error	df	*p*
Condition X	*i-Ready PI Usage*	-0.09	0.05	48	0.10
	i-Ready Diagnostic	**0.03**	0.00	3023	0.00*
	White	**0.06**	0.03	2888	0.02*
	ELL	0.03	0.04	2520	0.47
	SWD	**-0.17**	0.03	3008	0.00*
	ECD	**-0.12**	0.03	2529	0.00*
Condition 1	*i-Ready PI Usage*	-0.02	0.06	34	0.70
	i-Ready Diagnostic	**0.03**	0.00	2619	0.00*
	White	**0.06**	0.03	2467	0.05
	ELL	-0.02	0.05	1965	0.67
	SWD	**-0.11**	0.03	2601	0.00*
	ECD	**-0.11**	0.03	2033	0.00*
Condition 2	*i-Ready PI Usage*	0.03	0.08	13	0.71
	i-Ready Diagnostic	**0.03**	0.00	1435	0.00*
	White	**0.08**	0.04	1370	0.05
	ELL	0.02	0.08	1238	0.79
	SWD	-0.07	0.04	1424	0.07
	ECD	**-0.14**	0.04	1114	0.00*

Table E3. Cohort A: Short-Term Effects of i-Ready Personalized Instruction on Mathematics Achievement by Fidelity Condition

PI Condition	Variables	Coefficients	Standard Error	df	*p*
Condition X	*i-Ready PI Usage*	0.05	0.05	77	0.31
	i-Ready Diagnostic	**0.03**	0.00	3029	0.00*
	White	0.05	0.03	2941	0.06
	ELL	-0.01	0.04	2572	0.89
	SWD	**-0.22**	0.03	3000	0.00*
	ECD	**-0.13**	0.03	2768	0.00*
Condition 1	*i-Ready PI Usage*	0.04	0.06	14	0.57
	i-Ready Diagnostic	**0.03**	0.00	1945	0.00*
	White	0.06	0.03	1812	0.06
	ELL	0.01	0.06	1489	0.89
	SWD	**-0.13**	0.03	1932	0.00*
	ECD	**-0.09**	0.03	1400	0.00*

Appendix F: Full Impact Model Results – Cohorts B and C

Table F1. Cohort B: Intermediate-Term Effects of i-Ready Personalized Instruction on Mathematics Achievement by Fidelity Condition

PI Condition	Variables	Coefficients	Standard Error	df	p
Condition X	i-Ready PI Usage	-0.06	0.06	90	0.31
	i-Ready Diagnostic	**0.03**	0.00	3671	0.00*
	White	**0.11**	0.03	3553	0.00*
	ELL	**0.11**	0.04	3600	0.00*
	SWD	**-0.18**	0.03	3639	0.00*
	ECD	**-0.15**	0.02	3652	0.00*
Condition 1	i-Ready PI Usage	-0.07	0.06	95	0.25
	i-Ready Diagnostic	**0.03**	0.00	3314	0.00*
	White	**0.08**	0.03	3251	0.00*
	ELL	**0.11**	0.04	3214	0.00*
	SWD	**-0.15**	0.03	3292	0.00*
	ECD	**-0.18**	0.02	3299	0.00*
Condition 2	i-Ready PI Usage	-0.03	0.07	82	0.61
	i-Ready Diagnostic	**0.03**	0.00	2516	0.00*
	White	**0.07**	0.03	2490	0.02*
	ELL	0.03	0.05	2378	0.52
	SWD	**-0.14**	0.03	2496	0.00*
	ECD	**-0.14**	0.03	2504	0.00*

Table F2. Cohort B: Short-Term Effects of i-Ready Personalized Instruction on Mathematics Achievement by Fidelity Condition

PI Condition	Variables	Coefficients	Standard Error	df	p
Condition X	i-Ready PI Usage	-0.04	0.05	41	0.44
	i-Ready Diagnostic	**0.03**	0.00	3546	0.00*
	White	**0.06**	0.02	3207	0.02
	ELL	0.01	0.04	3323	0.77
	SWD	**-0.12**	0.03	3527	0.00*
	ECD	**-0.13**	0.02	2880	0.00*
Condition 1	i-Ready PI Usage	-0.01	0.05	57	0.82
	i-Ready Diagnostic	**0.03**	0.00	2823	0.00*
	White	0.03	0.03	2683	0.25
	ELL	0.02	0.05	2397	0.68
	SWD	**-0.08**	0.03	2800	0.01*
	ECD	**-0.13**	0.03	2536	0.00*

Table F3. Cohort C: Short-Term Effects of i-Ready Personalized Instruction on Mathematics Achievement by Fidelity Condition

PI Condition	Variables	Coefficients	Standard Error	df	p
Condition X	*i-Ready PI Usage*	-0.06	0.06	42	0.29
	i-Ready Diagnostic	**0.03**	0.00	4087	0.00*
	White	**0.07**	0.02	3729	0.00*
	ELL	0.04	0.03	4068	0.24
	SWD	**-0.09**	0.02	4055	0.00*
	ECD	**-0.11**	0.02	3543	0.00*
Condition 1	*i-Ready PI Usage*	0.01	0.06	31	0.87
	i-Ready Diagnostic	**0.03**	0.00	3248	0.00*
	White	**0.07**	0.02	2822	0.01*
	ELL	0.05	0.03	3231	0.14
	SWD	**-0.10**	0.03	3235	0.00*
	ECD	**-0.09**	0.03	2601	0.00*