

Impact Evaluation of the Imagine Math® Supplemental Digital Math Program in Idaho Schools

Connor J. Rooney, MS, Michael A. Cook, PhD,
Steven M. Ross, PhD

September 2026



JOHNS HOPKINS
UNIVERSITY



Impact Evaluation of the Imagine Math[®] Supplemental Digital Math Program in Idaho Schools

Connor J. Rooney, MS, Michael A. Cook, PhD,
Steven M. Ross, PhD

Center for Research and Reform in Education
Johns Hopkins University School of Education
2800 N. Charles Street
Baltimore, MD 21218
<https://education.jhu.edu/crre>

Preferred citation: Rooney, C. J., Cook, M. A., Ross, S. M.
(2026). *Impact evaluation of the Imagine Math[®] supplemental digital math program in Idaho schools*. Center for Research and Reform in Education, Johns Hopkins University.

Notice of Trademark: “Johns Hopkins University” and “JHU” are registered trademarks.

Copyright © 2026 Johns Hopkins University. No part of this report may be reproduced, distributed, or transmitted in any form or by any means, including photocopying, recording, website display, or other electronic or mechanical methods, without the prior written permission of the Johns Hopkins University Center for Research and Reform in Education. For permission requests, please contact CRRE.



Contents

EXECUTIVE SUMMARY	4
INTRODUCTION.....	6
Overview of Imagine Learning's Imagine Math®	6
Overview of the Evaluation	6
METHOD	7
Research Design	7
Participants	8
Measures	10
Analytical Approach	11
RESULTS	16
Imagine Math® Mathematics Achievement Impacts	17
Subgroup Analyses.....	18
Fidelity Analyses	20
DISCUSSION	22
REFERENCES.....	25
APPENDIX A: Descriptive Analyses	26
APPENDIX B: Full HLM Covariate Estimates.....	29
APPENDIX C: Supplemental Subgroup HLM Analyses	32
APPENDIX D: Supplemental Fidelity HLM Analyses	34
APPENDIX E: ITT Analyses	37
APPENDIX F: Supplemental Sensitivity HLM Analyses	42



EXECUTIVE SUMMARY

The Idaho Department of Education (IDE) partnered with Imagine Learning and the Center for Research and Reform in Education (CRRE) at Johns Hopkins University to conduct a mixed-methods evaluation of Imagine Learning's Imagine Math® (Imagine Math) supplemental digital math program during the 2024-2025 school year. This report presents the quantitative components of the study, focused on student mathematics achievement in Grades 4-8 and analyses of program usage. The qualitative components of the evaluation – focused on teacher experiences, instructional integration, and implementation supports gathered from teacher surveys and focus groups – are described in a separate companion report.

- The quantitative study used a student-level quasi-experimental design (QED), comparing Imagine Math users to matched non-users within grade level using propensity score matching (PSM) on prior achievement and demographic characteristics.
- The matched analytic sample included 34,452 students across 581 schools, with strong baseline equivalence on pretest scores and demographic characteristics.
- Data sources included statewide administrative records provided by IDE (demographics, enrollment), Imagine Math usage data (time in program, lessons completed, lessons passed), and Idaho Standards Achievement Test (ISAT) mathematics scores from spring 2024 and spring 2025.
- Impact analyses used two-level hierarchical linear modeling (HLM) to estimate adjusted differences in ISAT mathematics performance between Imagine Math users and matched comparison students, controlling for prior achievement, demographic characteristics, and grade level, and accounting for clustering of students within schools.
- In the primary analyses, students who completed at least one Imagine Math lesson were included in the treatment group. A supplementary intent-to-treat (ITT) analysis, which included students with any program exposure, was also conducted and is presented in Appendix E.
- Primary impact results were statistically significant, with Imagine Math students scoring significantly higher on the spring 2025 ISAT mathematics assessment than matched comparison students after adjusting for prior achievement, demographics, and school-level clustering.
- Grade-specific analyses indicated positive, statistically significant effects for Grade 4 (+6.35 points, $p = .005$, $ES = 0.06$) and Grade 7 (+9.24 points, $p = .01$, $ES = 0.06$), with smaller, non-significant differences for other grades.
- Subgroup analyses showed positive treatment effects for all student groups examined – including special education (SPED), English Language Learners (ELLs), White, Hispanic, and Other Race students. These marginal effects represent the adjusted difference in ISAT scores between Imagine Math users and matched non-users within each subgroup.
- SPED students were the only subgroup that demonstrated a significantly larger treatment effect than their peers. On average, SPED Imagine Math users scored 13.81 points higher than matched comparison SPED students ($ES = 0.10$), a



treatment effect that was 9.95 points larger than the effect observed for non-SPED students ($p < .001$, $ES = 0.07$).

- Fidelity analyses showed that students who completed 30 or more Imagine Math lessons in 2024-2025 achieved notably higher ISAT mathematics scores than both matched comparison students (+10.69 points, $p < .001$) and treated students who did not reach this usage threshold (+12.50 points, $p < .001$).
- Continuous-usage models demonstrated parallel patterns: each additional hour of Imagine Math engagement was associated with a 0.52- to 0.54-point increase in adjusted ISAT scores, each additional lesson completed with a 0.15- to 0.16-point increase, and each additional lesson passed with a 0.36- to 0.43-point increase. Together, these results indicate that deeper and sustained program engagement was consistently linked to stronger academic outcomes.
- Sensitivity analyses – including alternative model specifications and the exclusion of schools with extreme pretest means – confirmed the robustness of findings across analytic decisions.



INTRODUCTION

Overview of Imagine Learning's Imagine Math®

Imagine Learning's Imagine Math® (Imagine Math) program is a supplemental, adaptive digital mathematics solution designed to support and extend classroom instruction. As described in Imagine Learning's program materials, Imagine Math provides students with a personalized learning pathway built from a placement assessment, adaptive practice, scaffolded feedback, and access to live certified math teachers. The program is designed to strengthen conceptual understanding and procedural fluency through interactive tasks, embedded supports, and real-time adjustments based on student performance. Imagine Math includes content spanning Pre-K through high school Algebra and Geometry.

Each Imagine Math lesson consists of multiple engaging activities that provide a variety of opportunities for students to deepen their conceptual understanding of mathematics including core math foundations, problem solving, fractions, and measurement. Imagine Learning recommends that students use Imagine Math for 60 to 90 minutes per week, which typically results in completing two to three lessons. Students are also encouraged to maintain at least a 70% lesson-pass rate for lessons classified as on-grade-level or below-grade-level. Imagine Learning communicates these standard recommendations for weekly and cumulative program use to states and school districts choosing to implement Imagine Math.

Following a statewide request for proposals (RFP), Imagine Learning was selected in 2022 as Idaho's provider of digital, adaptive mathematics programming for grades K-12. Imagine Math was subsequently made available to all K-12 Idaho students. Imagine Math was purchased at the state level and offered as an optional resource to all Idaho districts. As a result, implementation varies across districts and schools depending on local priorities, scheduling decisions, and teacher discretion. Imagine Learning provides statewide support – including initial professional development, ongoing technical assistance, and regular check-ins with district leaders – to encourage consistent usage aligned with Imagine Learning's recommended dosage guidelines. However, because adoption is optional, the depth and consistency of implementation naturally differs across classrooms and districts.

Overview of the Evaluation

The present evaluation examined student achievement outcomes in 2024-2025, the fourth year of a five-year implementation cycle. The study drew on statewide administrative data, including Imagine Math usage records and end-of-year Idaho Standards Achievement Test (ISAT) mathematics scores for Grades 4-8 (grades with available and consistent pretest data) students. Analyses compared ISAT performance between students who used Imagine Math and those who did not, examined associations between usage levels and achievement, and assessed variation in outcomes across student subgroups.



The quantitative evaluation is guided by the following research questions:

1. How did performance on the ISAT mathematics assessment differ between students who used Imagine Math and students who did not use Imagine Math?
2. What was the association between Imagine Math usage levels (e.g., time in program, lessons completed, lessons passed) and student performance on the ISAT mathematics assessment?
3. How did student ISAT mathematics assessment performance differ by student subgroups (e.g., special education, English Language Learners, race/ethnicity)?

METHOD

Research Design

This study used a student-level quasi-experimental design (QED) to estimate the impact of Imagine Math on mathematics achievement during the 2024-2025 school year. Because Imagine Math was made optionally available to educators across Idaho, usage may not have been assigned at the school level; students within the same school often differed substantially in whether they used the program and in the extent of their engagement. Some students completed lessons, others logged limited time in individual activities without completing lessons, and many had no recorded usage at all. As a result, neither schools nor classrooms could be meaningfully classified as “treatment” or “comparison,” making student-level matching the appropriate analytic approach. At the same time, students were nested within schools, and outcomes may be correlated within those clusters; therefore, all impact analyses were estimated using two-level hierarchical linear models (HLMs) to account for school-level clustering while isolating the relationship between Imagine Math usage and student achievement.

This report establishes two treatment definitions, each aligned to its own matching and analytic framework:

- **Primary Treatment Definition.** Developed in consultation with Imagine Learning, this definition classifies students as treatment if they completed at least one Imagine Math lesson during the 2024-2025 school year. Comparison students completed zero lessons, though they may have logged time in the program (approximately 9% of the matched comparison group under this definition had Imagine Math usage in 2024-2025, with an average of 40 minutes of usage across the school year). This definition reflects meaningful instructional engagement and is used for all main analyses presented in the Results section, as well as all descriptive information in the Participants section and all analytical diagnostics in the Analytical Approach section.
- **Intent-to-Treat (ITT) Definition.** As a classical ITT alternative, this definition classifies students as treatment if they had any recorded Imagine Math usage time in 2024-2025, regardless of whether they completed a lesson. Comparison students had no usage at all. This definition reflects exposure rather than engagement and is modeled across Appendix E.



All matching and outcome analyses were conducted separately and consistently within each treatment definition, ensuring that comparisons were internally valid and aligned to the operational meaning of each definition.

We received statewide administrative data from IDE, including demographic characteristics and ISAT mathematics scores for Grades 4-8 students. Imagine Learning provided program participation indicators and Imagine Math usage records. These records were merged using IDE-assigned student IDs, enabling precise identification of Imagine Math users and non-users across schools. The merged dataset formed the basis for all matching and outcome analyses.

To strengthen causal inference, we employed propensity score matching (PSM) within grade level, using prior ISAT mathematics achievement (spring 2024) and demographic characteristics to construct comparison groups that were academically and demographically similar to Imagine Math users. Additional details on the matching procedures are provided in the Analytical Approach section.

The primary outcome measure for the evaluation was the spring 2025 ISAT mathematics assessment, administered statewide. This outcome supports both overall and grade-specific estimates of Imagine Math's impact. Analyses proceeded in several layers to address the full set of research questions: (a) full-sample impacts, (b) grade-level impacts, (c) subgroup impacts, (d) fidelity and usage associations, and (e) sensitivity analyses.

Participants

Idaho's public school system serves approximately 317,600 K-12 students across 795 public schools. Public school enrollment across Idaho is predominately White (74%), with smaller representation of Hispanic/Latino students (19%). Statewide, 35.1% of students qualify for free or reduced-price lunch (NCES Idaho State Dashboard).

The sample characteristics reported in this section reflect the student analytic sample after PSM was conducted within grade using prior ISAT mathematics achievement and demographic characteristics. Matching was performed separately for each treatment definition, and comparison students were selected from the full statewide pool of eligible students within the same grade. The final analytic sample also accounted for small levels of attrition between the spring 2024 ISAT pretest and the spring 2025 ISAT mathematics assessment. The matched analytic sample corresponding to the Primary Treatment definition (students completing at least one Imagine Math lesson) included 581 schools encompassing 34,452 students with complete pretest and outcome data.



Schools included in the analysis



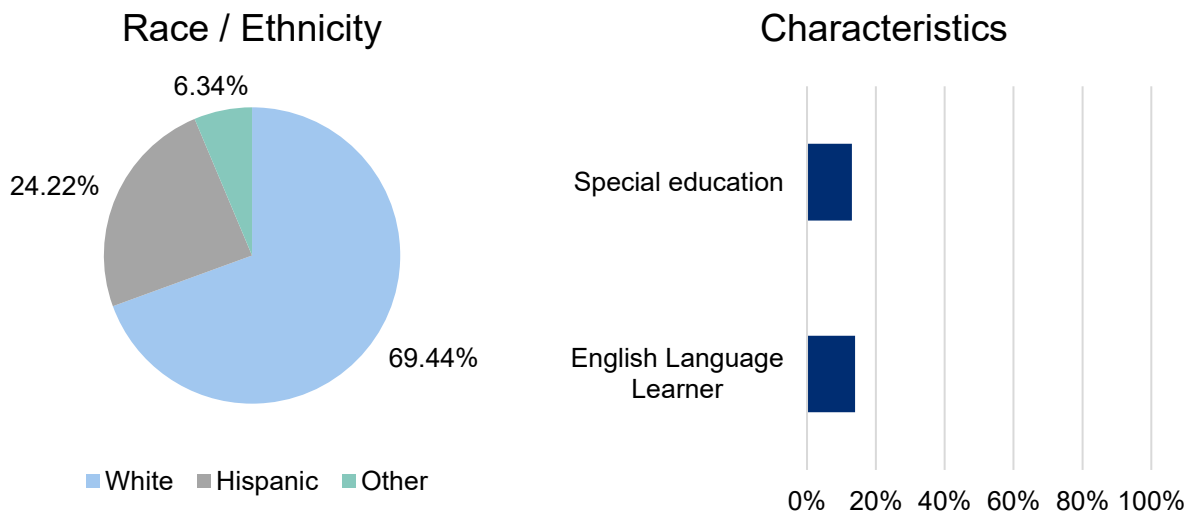
581 Schools

Matched students with complete data



17,226 Imagine Math students
17,226 Comparison students

Demographic snapshot of student participants in the analytic sample:



The demographic makeup of the matched analytic sample closely reflects the overall composition of Idaho public school students. As in the state population, the majority of students in the sample are White (69.44%), with additional representation of Hispanic/Latino students (24.22%) and smaller proportions of Other Race students (6.34%). A more detailed breakdown of demographic characteristics and baseline equivalence is provided in the Analytical Approach section. Table 1 displays demographic characteristics of the analytic sample by study condition.

Table 1

Demographics of Study Participants in the Analytic Sample (Primary Treatment Definition)

Characteristic	All N = 34,452	Imagine Math group (n = 17,226)	Comparison group (n = 17,226)
White	69.44%	68.71%	70.17%
Hispanic	24.22%	24.39%	24.05%
Other race/ethnicity	6.34%	6.90%	5.78%
English Language Learner	13.58%	14.01%	13.15%
Special education	12.93%	12.61%	13.25%



Overall, the treatment and matched-comparison groups were highly similar in demographic composition. Both groups included comparable proportions of White, Hispanic, and other race/ethnicity students, as well as similar percentages of English Language Learners (ELLs) and students receiving special education services (SPED). None of these differences were statistically significant, indicating that the PSM procedure successfully produced balanced groups. To ensure further comparability, all demographic characteristics listed in Table 1 were included as covariate adjustments in the analytic models, allowing estimated program effects to reflect differences attributable to Imagine Math usage rather than underlying demographic factors. Additional details on the matching procedures used to establish this balanced analytic sample are provided in the Analytical Approach section.

Measures

To address the quantitative research questions, the CRRE study team processed and analyzed data gathered by IDE for the 2023-2024 and 2024-2025 school years. The study team received statewide student-level administrative and assessment data from IDE, as well as Imagine Math usage records. Table 2 summarizes the quantitative research questions and data sources.

Table 2

Quantitative Research Questions with Data Sources and Measures

Research question	Student achievement data	Imagine Math usage data
1. How did performance on the ISAT mathematics assessment differ between students who used Imagine Math and students who did not use Imagine Math?	✓	
2. What was the association between Imagine Math usage levels (e.g., time in program, lessons completed, lessons passed) and student performance on the ISAT mathematics assessment?	✓	✓
3. How did student ISAT mathematics assessment performance differ by student subgroups (e.g., special education, English Language Learners, race/ethnicity)?	✓	

Data sources and measures for the quantitative study included student math achievement, as described below.



ISAT Mathematics Assessment. IDE provided student-level ISAT mathematics scores for the spring 2024 and spring 2025 test administrations. The ISAT mathematics assessment is Idaho's statewide standardized test used to measure students' mastery of grade-level mathematics content. ISAT mathematics tests are administered annually in Grades 3-8 and are not vertically scaled, meaning that scale scores cannot be compared across grade levels. A Grade 5 score, for example, is not directly comparable to a Grade 6 score; each reflects performance relative only to the standards assessed at that grade.

The ISAT data included overall scale scores, proficiency level categories corresponding to the scale scores (Level 1 – Level 4), and test administration dates. For this evaluation, the spring 2024 ISAT mathematics score served as the pretest for students in Grades 4-8, providing a consistent baseline measure of prior achievement. The spring 2025 ISAT mathematics score served as the primary outcome measure for all impact analyses. Because ISAT is administered statewide and consistently across grades, it provides a reliable indicator of mathematics achievement for estimating program impacts.

Imagine Math Usage Data. Imagine Learning provided detailed student-level usage records for Imagine Math for the 2024-2025 school year. These data included time spent in the program, lessons completed, and lessons passed, as well as student-level identifiers aligned to IDE's research IDs. Usage data from the 2024-2025 school year served as the basis for establishing the treatment definitions documented in the Research Design section, as this year aligned with the evaluation period and reflected students' most current engagement with Imagine Math. Usage measures from 2024-2025 were also used to quantitatively analyze associations between program engagement levels (e.g., time in program, lessons completed, lessons passed) and ISAT performance.

Demographic Measures. IDE provided student-level demographic data including rostering data (school and grade enrollment), race/ethnicity, English Language Learner (ELL) status and special education (SPED) status.¹ These student characteristic variables were used to describe the matched analytic sample and assess baseline equivalence between the treatment and comparison groups. Moreover, demographic measures were used as covariates in all adjusted impact models to ensure that estimated program effects reflected differences attributable to Imagine Math rather than underlying demographic factors.

Analytical Approach

Data for student outcomes were analyzed using a combination of descriptive statistics, matched-sample comparisons, and multilevel modeling. HLMs with a two-level model of

¹ The administrative file provided by IDE included school level eligibility information for state meal programs (e.g., CEP, PROV, NSLP). Because these data reflect school level program status rather than student level economic disadvantage, they could not be used to construct a reliable economic disadvantage indicator for these analyses. Analyses and references to economic disadvantage have therefore been excluded from this version of the report. IDE has been contacted to determine whether student-level economic disadvantage data can be made available; if so, a revised report will incorporate those measures.



students nested within schools were used to determine the impact of Imagine Math on math achievement outcomes, controlling for student-level demographic variables including race/ethnicity, ELL status, SPED status, pretest scores, and student grade level, and accounting for the clustering of students within schools. All quantitative analyses were completed in Stata (version 19).

Student-Level Baseline Equivalence

Because Imagine Math was optionally available to Idaho educators, usage varied widely across and within schools. Some students completed lessons, others logged only limited time in individual activities, and many had no recorded usage at all. This uneven engagement meant that schools could not be classified as “treatment” or “comparison,” so baseline equivalence needed to be established at the student level. All matching procedures therefore focused on individual students to ensure that comparisons reflected actual patterns of program use.

To strengthen causal inference, we used PSM within grades to create treatment and comparison groups that were similar on key baseline characteristics. Among the 120,065 Grades 4-8 students in the sample, 112,421 had complete pretest and outcome data, and were enrolled in 595 distinct schools; this constituted the unmatched sample used to assess baseline equivalence and perform matching. Prior to matching, the treatment group included proportionally more Hispanic and ELL students, while the comparison group (students who completed no Imagine Math lessons under the Primary Treatment definition) included proportionally more White students, although none of these differences were statistically significant and all were below the What Works Clearinghouse (WWC) 0.25 SD threshold for baseline equivalence. Matching was therefore used to further tighten covariate balance before estimating treatment effects.

Matching was performed separately for each grade level (4-8) using prior ISAT mathematics achievement (spring 2024) and two demographic characteristics (Hispanic ethnicity and ELL status). These demographic variables were selected because they represented two of the largest minority subgroups in the student sample and therefore captured meaningful demographic variation relevant to math achievement. This grade-specific approach ensured that students were compared to peers experiencing similar curricular expectations and assessment conditions.

The matching workflow followed a consistent procedure across grade levels. For each grade band, we estimated a logistic propensity score model using the `psmatch2` command in Stata and implemented 1:1 nearest-neighbor matching without replacement, pairing each treatment student with the closest comparison student based on propensity score.

Diagnostics were conducted to confirm balance, and students who were successfully matched were marked in the dataset and retained for analysis. All treatment students with complete pretest and outcome ISAT data were included in the matched analytic



sample, ensuring that impact estimates reflected the full population of eligible Imagine Math users.

The matching process produced samples that were demographically and academically comparable across treatment and comparison groups. As reported in the Participants section, none of the demographic differences between groups were statistically significant, indicating that the PSM procedure successfully achieved baseline equivalence. Table 3 presents the student-level baseline equivalence estimates before and after matching.

Table 3
Student-level Baseline Equivalence: Pretest and Demographics

Variable	Full sample: Imagine Math mean (SD)	Full sample: Comparison mean (SD)	SMD (Full sample)	Matched sample: Imagine Math mean (SD)	Matched sample: Comparison mean (SD)	SMD (Matched sample)
Pretest (ISAT scale score 2024)	2475 (123.31)	2479 (224.90)	-0.02	2475 (123.31)	2475 (127.41)	-0.001
Hispanic	0.24 (0.43)	0.19 (0.39)	0.12	0.24 (0.43)	0.24 (0.43)	0.01
White	0.69 (0.46)	0.74 (0.44)	-0.12	0.69 (0.46)	0.70 (0.46)	-0.03
Other race	0.07 (0.25)	0.07 (0.25)	0.004	0.07 (0.25)	0.06 (0.23)	0.05
ELL status	0.14 (0.35)	0.11 (0.32)	0.09	0.14 (0.35)	0.13 (0.34)	0.03
SPED status	0.13 (0.33)	0.13 (0.34)	-0.02	0.13 (0.33)	0.13 (0.34)	-0.02
Student <i>n</i>	17,226	95,195	112,421	17,226	17,226	34,452

Note. Demographic characteristics are presented as mean proportions (e.g., 0.50 = 50%). SMD = standardized mean difference.

Standardized mean differences (SMDs) for the pretest and all demographic variables remained below the WWC standard threshold, indicating strong baseline equivalence between the matched students.

Analytical Models

Descriptive Analyses. Descriptive analyses, outlined in Appendix A, summarize unadjusted patterns in spring 2025 ISAT mathematics performance for treatment and comparison students. These include:

- Mean ISAT scale scores and standard deviations for the full sample and for individual grade levels 4-8
- Unadjusted comparisons of treatment and comparison groups (*t*-tests)
- Supplemental descriptives for key subgroups (race/ethnicity, ELL status, and



SPED status)

- Pretest and posttest ISAT proficiency levels for Grades 4-8

These descriptive statistics provided initial context for understanding achievement patterns prior to adjusting for baseline differences and other covariates.

Primary Impact Model. Because Imagine Math usage occurred at the student level, students were nested within schools but not assigned by school. To account for this structure, primary impact estimates were obtained using a two-level HLM, with students (Level 1) nested within schools (Level 2). This analytical structure was necessary because students within the same school may share instructional conditions, demographic compositions, or other unobserved characteristics that make their achievement scores more similar to one another than to students in different schools. HLM accounts for this non-independence by estimating both fixed effects (relationships consistent across all schools) and random effects (variation attributable to between-school differences). These models also produce an intraclass correlation coefficient (ICC), which quantifies the proportion of total variance in ISAT scores that lies between schools rather than between students within schools. All HLMs adjusted for baseline mathematics achievement and demographic characteristics.

The continuous ISAT pretest covariate was grand-mean-centered, meaning it was centered around the overall sample mean rather than school-specific means. Grand-mean-centering improves interpretability of the intercept, reduces multicollinearity among covariates, and ensures that the treatment coefficient reflects the average adjusted difference between treatment and comparison students across the full analytic sample.

In addition to an aggregate model, separate models were estimated for each grade (Grades 4-8). The full sample model included dummy variables for grade level, allowing the analysis to adjust for systematic differences in achievement across grades and ensuring that the estimated treatment effect was not confounded by grade-specific performance patterns. Grade-level indicators function as fixed effects, capturing average achievement differences attributable to grade rather than to treatment status.

The model included:

- *Outcome:* Spring 2025 ISAT mathematics scale score
- *Covariates:* Prior achievement (spring 2024 ISAT), race/ethnicity, ELL status, SPED status, and grade-level dummy variables
- *Level-2 structure:* Random intercepts for schools, allowing each school to have its own baseline level of achievement after accounting for covariates

A simplified representation of the model is:

$$ISAT_{ij} = \beta_0 + \beta_1(Treatment_{ij}) + \beta_2(Pretest_{ij}^{GC}) + \beta_3(Race/Ethnicity_{ij}) + \beta_4(ELL_{ij}) + \beta_5(SPED_{ij}) + \sum_{g=4}^8 \beta_g(GradeDummy_{ij}) + u_{0j} + e_{ij}$$



where β_0 represents the average spring 2025 ISAT mathematics score for a comparison group student with a grand-centered pretest equal to zero (i.e., the sample mean). β_1 captures the adjusted difference in ISAT performance between treatment and comparison students, holding all other covariates constant (i.e., the treatment effect). β_2 reflects the association between prior achievement (spring 2024 ISAT) and the spring 2025 ISAT outcome, using the grand-centered pretest score. The coefficients β_3 through β_5 represent the adjusted relationships between student demographic characteristics – race/ethnicity, ELL status, and SPED status – and ISAT performance. The grade-level dummy coefficients, β_g , capture systematic differences in achievement across Grades 4-8 relative to the reference grade. The term u_{0j} is the random intercept for school j , representing school-specific deviations from the overall intercept after accounting for covariates, and e_{ij} is the student-level residual.

This model estimates the adjusted association between Imagine Math and spring 2025 ISAT achievement while appropriately modeling school-level clustering and controlling for baseline differences.

Subgroup Analyses. To examine whether program impacts varied across student characteristics, we estimated subgroup models using the same two-level HLM structure as the primary impact analysis. Subgroup effects were tested by introducing interaction terms between treatment status and the subgroup indicator (e.g., SPED status $\times$ treatment, ELL status $\times$ treatment). This approach allows the treatment effect to differ across student groups while holding constant the full set of covariates and the school-level random intercept.

These analyses are introduced in the Results section and expanded in Appendix C. Subgroup models included:

- Grade-specific HLMs (Grades 4-8)
- Supplemental subgroup HLMs (race/ethnicity, ELL status, SPED status)

Together, these models assessed whether Imagine Math impacts varied across grade levels and student groups, using a consistent modeling framework that isolated subgroup-specific differences in treatment effects.

Fidelity Analyses. In addition to estimating overall program impacts, we conducted a series of fidelity analyses to examine whether higher levels of Imagine Math engagement were associated with stronger mathematics achievement outcomes. Fidelity was operationalized in three ways: (a) full fidelity, defined as completing 30 or more lessons during the 2024-2025 school year; (b) relaxed fidelity, defined as completing 24 or more lessons; and (c) continuous usage, modeled using total hours spent in the program, Imagine Math lessons completed, or Imagine Math lessons passed. Each fidelity definition was analyzed using two parallel modeling approaches to provide a comprehensive view of how engagement relates to achievement:

- *Fidelity vs. matched comparison students (excluding non-fidelity treatment students).* These models compared fidelity students to matched comparison students, restricting the analytic sample to students who met the fidelity threshold



and their matched counterparts.

- *Fidelity vs. non-fidelity treatment students.* These models compared fidelity students to treatment students who used Imagine Math but did not meet the fidelity threshold.

The objective of these fidelity modeling approaches was to isolate the association between high engagement and achievement relative to students who either did not complete Imagine Math lessons (matched comparison) or to other treatment students who completed lessons but did not achieve fidelity. Modeling fidelity as a continuous usage parameter allowed us to estimate the marginal association between additional time in the program, or engagement with program lessons, and ISAT performance.

Across all fidelity analyses, we used the same two-level HLM structure described previously, adjusting for baseline achievement, demographic characteristics, and grade level. In each specification, the fidelity indicator – whether defined as full fidelity, relaxed fidelity, or continuous usage – replaced the treatment variable. The main Results section presents the full-fidelity and continuous usage models. The relaxed fidelity specifications and additional details for the continuous usage fidelity models are provided in Appendix D.

Sensitivity Analyses. To assess the robustness of the primary findings, we estimated several supplemental HLMS, all reported in Appendix F, including model outputs, diagnostics, and interpretive notes. These analyses were designed to evaluate whether the estimated associations between Imagine Math usage and ISAT mathematics performance remained stable under alternative assumptions, sample definitions, and model structures.

The sensitivity analyses included:

- Covariate-stripped impact models
- Models including school fixed effects
- Models that centered the pretest score within schools
- Models excluding students in schools with the top or bottom 10% of pretest means
- ITT Analyses, presented in full in Appendix E, which include
 - Matching diagnostics and results for the ITT definition
 - Descriptive statistics for the ITT matched sample
 - Primary ITT impact models
 - ITT subgroup models
 - ITT fidelity analyses

RESULTS

This section presents the quantitative evaluation findings for Imagine Math. We begin by reporting the primary impact estimates from the HLMS as described in the Analytical Approach section, followed by grade-specific HLM analyses.

Supplemental analyses for the Primary Treatment definition are presented in the



Appendices. Appendix A provides descriptive summaries of unadjusted spring 2025 ISAT mathematics outcomes. Appendix B reports the full covariate estimates from the primary HLMs described in the following Results section. Appendix C includes complete covariate outputs for the subgroup HLMs examining race/ethnicity, ELL status, and SPED status. Appendix D contains supplemental fidelity analyses. Appendix E documents the ITT analyses, and Appendix F provides sensitivity checks supporting the robustness of the Primary Treatment findings.

Imagine Math® Mathematics Achievement Impacts

Key Findings

- After adjusting for pretest achievement, demographics, and school-level clustering, Imagine Math students scored significantly higher than matched comparison students on the spring 2025 ISAT mathematics assessment. The estimated treatment effect of 5.21 points ($p < .001$) corresponds to an effect size (ES) of 0.04 SDs.
- Grade-specific HLMs showed statistically significant positive effects for Grades 4 and 7, with Imagine Math students scoring 6.35 points higher in Grade 4 ($p = .005$; ES = 0.06) and 9.24 points higher in Grade 7 ($p = .01$; ES = 0.06) than matched comparison peers. Treatment-comparison differences in Grades 5, 6, and 8 were small and not statistically significant.
- Across all student subgroups, adjusted marginal treatment effects – estimated as the difference in outcomes between Imagine Math users and matched non-users within each subgroup, after controlling for pretest scores and demographics – were statistically significant and positive. However, only the interaction estimate for SPED students was statistically significant, indicating that this group experienced significantly larger treatment effects (averaging 9.95 points higher; ES = 0.07) than their non-SPED peers.
- Students who completed 30 or more Imagine Math lessons demonstrated substantially higher adjusted ISAT mathematics scores than both matched comparison students (ES = 0.08) and treated non-fidelity students (ES = 0.10). Continuous-usage analyses showed the same overall pattern: students who engaged more deeply with Imagine Math tended to experience stronger academic outcomes.

This section presents results from the primary impact analysis estimating the association between engagement with Imagine Math and students' spring 2025 ISAT mathematics scores. These models provide the main estimates of program impacts on ISAT performance and are summarized in Table 4, with full covariate results reported in Appendix B.

**Table 4****Primary HLM Estimates for Spring 2025 ISAT Mathematics Achievement**

Variable	Estimate	Standard error	p-value	Effect size ²
Imagine Math	5.21***	1.17	<.001	0.04
Constant	2502.14***	2.22	<.001	
Student <i>n</i>	34,452			
Number of schools (clusters)	581			
ICC	.02			

Note: *** $p < .001$. ICC = Intraclass Correlation

The regression estimates represent the adjusted average difference in spring 2025 ISAT mathematics scores between treatment and comparison students, controlling for pretest achievement, demographic characteristics, and grade level (Grades 4-8). As shown in Table 4, Imagine Math students scored 5.21 points higher on the spring 2025 ISAT mathematics assessment than matched comparison students, and this difference was statistically significant ($p < .001$). Because ISAT scores are not vertically scaled across grades, this pooled estimate should be interpreted as an adjusted average of the differences between Imagine Math and comparison students within each grade, rather than as a directly comparable 5.21-point difference on a common cross-grade scale. The corresponding effect size of 0.04 SDs indicates a small but positive impact of Imagine Math on mathematics achievement.

Subgroup Analyses

Because the ISAT mathematics assessment is not vertically scaled, scores cannot be directly compared across grade levels. To examine whether program impacts varied by grade, we estimated grade-specific HLMs using the same specifications as the primary impact analysis. These models allowed treatment effects to be interpreted within each grade's unique achievement scale, providing a more precise view of grade-level variation in program impacts. Results from the grade-specific models for Grades 4-8 are summarized in Table 5.

Table 5**Grade-Specific HLM Estimates for Spring 2025 ISAT Mathematics Achievement**

Grade	Treatment estimate	Standard error	p-value	Effect size	Constant	Student <i>n</i>	Number of schools (clusters)	ICC
4	6.35*	2.28	.005	0.06	2510.11***	10,006	409	.08
5	2.45	1.96	.21	0.02	2503.23***	8,648	390	.04
6	3.77	3.09	.22	0.02	2502.13***	7,214	296	.06
7	9.24*	3.71	.01	0.06	2505.82***	4,522	227	.03

² Effect sizes throughout the report are reported as Hedges' *g*, a bias-corrected standardized mean difference that represents the magnitude of the adjusted treatment effect in standard deviation units.



8	-1.25	3.39	.71	-0.009	2499.02***	4,062	228	.03
---	-------	------	-----	--------	------------	-------	-----	-----

Note. * $p < .05$; *** $p < .001$.

As shown in Table 5, grade-specific HLM results indicate that treatment effects were positive and statistically significant for Grades 4 and 7, with estimates of 6.35 points ($p = .005$) and 9.24 points ($p = .01$), respectively. These correspond to small effect sizes of 0.06 SDs, suggesting modest yet practically significant advantages for Imagine Math students in those grades. For Grades 5 and 6, treatment estimates were positive but not statistically significant, and the Grade 8 model yielded a small, non-significant negative estimate. Taken together, these results show that, after adjusting for pretest achievement, demographic characteristics, and school-level clustering, Imagine Math students in Grades 4 and 7 demonstrated significantly higher adjusted mathematics performance than their matched comparison peers, while differences in other grades were small and not statistically meaningful.

Next, to examine whether program impacts varied across student characteristics, we estimated HLMs that included treatment $\times$ subgroup interaction terms. These models test whether the adjusted difference in mathematics achievement between Imagine Math and comparison students differed by subgroup. For the race/ethnicity analyses, White students served as the referent category, with a separate interaction term estimating differences for Hispanic students. Table 6 summarizes the adjusted treatment effects for each subgroup and the corresponding differences in treatment effects, while full subgroup model outputs, including all covariate estimates and constants, are provided in Appendix C.

Table 6
Adjusted HLM Estimates of Imagine Math Treatment Effects by Student Subgroup

Subgroup	Marginal treatment effect	Standard error	p -value	Effect size	Difference in treatment effects ³	p -value
Imagine Math x non-SPED	3.86*	1.22	.002	0.03	—	—
Imagine Math x SPED	13.81***	2.39	<.001	0.10	+9.95***	<.001
Imagine Math x White	4.45*	1.30	.001	0.03	—	—
Imagine Math x Hispanic	6.62***	1.87	<.001	0.05	+2.17	.27
Imagine Math x non-ELL	4.92***	1.23	<.001	0.04	—	—
Imagine Math x ELL	6.94*	2.39	.004	0.05	+2.02	.41

³ Difference in treatment effects represents the treatment $\times$ subgroup interaction (subgroup treatment effect minus comparison group treatment effect).



Note. Marginal treatment effects represent the adjusted treatment effect of Imagine Math within each subgroup. * $p < .05$; *** $p < .001$. ICC = .02 across all models. Student $n = 34,452$; number of school clusters = 581.

Across all subgroup models, Imagine Math participation was associated with statistically significant, positive impacts on mathematics achievement for every group of students. The marginal treatment effects shown in Table 6 represent the adjusted difference in ISAT scores between Imagine Math users and matched non-users within each subgroup (for example, ELL Imagine Math users compared with ELL non-users). The difference-in-effects column shows whether those subgroup-specific impacts were significantly larger or smaller than the effects for their counterparts (for example, whether ELL students benefited more than non-ELL students). Although all racial/ethnic groups and both ELL and non-ELL students demonstrated significant positive treatment effects, none of the differences between groups were statistically significant, indicating that these groups benefited similarly from Imagine Math. In contrast, SPED students showed a significant positive difference-in-effect, meaning they experienced a larger adjusted treatment effect than their non-SPED peers. Overall, Imagine Math participation was associated with meaningful improvements in mathematics achievement for all student groups (effect sizes ranging from 0.03 to 0.10), with particularly strong gains among SPED students.

Fidelity Analyses

To assess whether higher levels of engagement with Imagine Math was associated with larger gains in mathematics achievement, we estimated a series of fidelity models using the same two-level HLM structure applied in the primary impact analyses. In this section, we begin with a focus on the full fidelity definition – students who completed 30 or more lessons during 2024-2025 – and present results from two comparisons in Table 7: high-fidelity students versus matched comparison students and high-fidelity students versus treated non-fidelity students. Among the matched sample of 17,226 treatment students, 4,516 (26.22%) achieved this full fidelity threshold.

Table 7

Panel A: Adjusted Fidelity HLM Estimates, Full Fidelity vs. Matched Comparison Students

Variable	Estimate	Standard error	p -value	Effect size
Full Imagine Math fidelity	10.69***	1.91	<.001	0.08
Constant	2498.61***	2.80	<.001	
Student n	21,742			
Number of schools (clusters)	580			
ICC	.02			

Note. *** $p < .001$.



Panel B: Adjusted Fidelity HLM Estimates, Full Fidelity vs. Treated Non-Fidelity Students

Variable	Estimate	Standard error	p-value	Effect size
Full Imagine Math fidelity	12.50***	1.34	<.001	0.10
Constant	2506.55***	2.99	<.001	
Student <i>n</i>	17,226			
Number of schools (clusters)	224			
ICC	.02			

Note. *** $p < .001$.

Across both types of comparisons, students who met the full fidelity threshold demonstrated consistently strong and statistically significant gains in mathematics achievement. When compared with matched comparison students (Panel A), high-fidelity students scored an adjusted 10.69 points higher ($p < .001$; ES = 0.08), and this advantage increased to 12.5 points ($p < .001$; ES = 0.10), when compared with treated students who did not meet the fidelity threshold (Panel B). Although effect sizes are modest, as expected in large-scale educational evaluations, the pattern across panels is highly consistent: students who completed 30 or more Imagine Math lessons achieved meaningfully higher ISAT mathematics scores than both reference groups, reinforcing the conclusion that program engagement that meets developer recommendations is associated with stronger academic outcomes.

In Table 8, we present treatment effect results from analyses that model fidelity as a continuous usage parameter of additional time using the program or additional lesson engagement.

Table 8

Panel A: Adjusted HLM Estimates, Imagine Math Usage, Matched Sample

Variable	Estimate	Standard error	p-value
Imagine Math hours	0.52***	0.07	<.001
Imagine Math lessons completed	0.15***	0.02	<.001
Imagine Math lessons passed	0.36***	0.03	<.001

Note: *** $p < .001$. ICC = .02-.03 across all models. Student *n* = 34,452; number of school clusters = 581.

Panel B: Adjusted HLM Estimates, Imagine Math Usage, Treatment Sample

Variable	Estimate	Standard error	p-value
Imagine Math hours	0.54***	0.07	<.001
Imagine Math lessons completed	0.16***	0.02	<.001
Imagine Math lessons passed	0.43***	0.02	<.001

Note: *** $p < .001$. ICC = .02-.03 across all models. Student *n* = 17,226; number of school clusters = 224.



Continuous-usage models were estimated to examine how different dimensions of program engagement – hours of use, lessons completed, and lessons passed – related to ISAT achievement. As shown in Table 8, each measure showed a small but statistically significant positive association with achievement. For example, in the matched sample, each additional hour of Imagine Math usage was associated with a 0.52-point increase in adjusted ISAT mathematics scores (0.54 points in the treatment-only sample). This means that a student who used Imagine Math for 10 hours would be expected to score roughly 5 points higher than a similar student who did not use the program. Likewise, completing 10 lessons would correspond to about a 1.5-point gain, and passing 10 lessons to about a 4-point gain. These results indicate that greater engagement with Imagine Math was consistently linked to higher mathematics achievement.

The matched-sample models include both Imagine Math users and the small portion of comparison students who recorded limited usage under the Primary Treatment definition. Including these students allows the analysis to capture the full spectrum of program engagement – from students with little to no exposure to those with sustained use – rather than restricting the sample only to active users. In contrast, the treatment-only models estimate the same relationships solely among students who used Imagine Math, providing a check on whether the positive associations hold within the user population itself. Across both samples, the results were highly consistent, reinforcing that incremental increases in program engagement were reliably associated with higher adjusted ISAT mathematics scores.

Supplemental fidelity analyses, including parallel models using the relaxed fidelity threshold (24+ lessons completed), are reported in Appendix D. Sensitivity analyses presented in Appendix F further tested the robustness of these findings using alternative HLM specifications and models excluding schools with extreme pretest distributions. Across all alternative specifications, the estimated treatment effects remained stable, supporting the consistency of the primary and grade-specific results.

DISCUSSION

The present study examined the impact of the Imagine Math supplemental curriculum during the 2024-2025 school year across Idaho schools. Using a student-level QED that included grade-specific propensity score matching, this quantitative evaluation compared mathematics achievement outcomes for Imagine Math users to those of demographically and academically similar non-users statewide. A separate CRRE report will incorporate teacher survey and focus group data to capture educators' perceptions of program usability, instructional integration, and implementation conditions. Together, these quantitative and qualitative components were designed to provide a comprehensive understanding of how Imagine Math functioned during the fourth year of statewide availability.



Imagine Math® Impacts on Student Math Achievement

Across all primary and sensitivity analyses, results consistently indicated that students who used Imagine Math scored significantly higher on the spring 2025 ISAT mathematics assessment than matched comparison students. The adjusted treatment effect for the full sample was +5.21 points (ES = 0.04), representing a modest but positive association between Imagine Math usage and mathematics achievement during the 2024-2025 school year. Because ISAT scores are not vertically scaled across grades, the pooled 5.21-point estimate should be interpreted cautiously as an adjusted average of within-grade differences rather than as a directly comparable 5.21-point difference across grades.

Grade-specific models revealed positive effects in Grade 4 (+6.35 points; ES = 0.06) and Grade 7 (+9.24 points; ES = 0.06), indicating that Imagine Math users in these grades achieved higher adjusted ISAT scores than their matched peers. Differences in Grades 5, 6, and 8 were small and not statistically significant, suggesting that impacts varied somewhat across grade levels but were generally positive.

Subgroup analyses showed that Imagine Math was associated with positive impacts for every group of students examined. The marginal treatment effects in these models represented the adjusted difference in ISAT mathematics scores between Imagine Math users and matched non-users within each subgroup (for example, ELL Imagine Math users compared with ELL non-users). All subgroups – including SPED, ELL, White, Hispanic, and Other Race students – demonstrated statistically significant positive marginal treatment effects, indicating that Imagine Math users outperformed similar non-users within each group.

The difference-in-effects estimates compared whether any subgroup benefited more than its counterpart. SPED students were the only group with a statistically significant difference-in-effect, meaning their treatment effect was significantly larger than that of non-SPED students (9.95 points higher, ES = 0.07). These findings indicate that SPED students experienced a larger adjusted treatment effect than their non-SPED peers, while all other groups showed similar benefits.

Fidelity analyses provided additional insight into the relationship between program engagement and achievement. Students who completed 30 or more Imagine Math lessons demonstrated substantially higher adjusted mathematics scores across both modeled comparisons, including +10.69 points (ES = 0.08) relative to matched comparison students and +12.50 points (ES = 0.10) relative to treated non-fidelity students. The continuous-usage models showed a similar pattern: higher levels of engagement were consistently associated with small, incremental increases in adjusted ISAT mathematics scores. These associations were correlational, but together they suggest that students who engaged more deeply with Imagine Math tended to experience stronger academic outcomes.

Several contextual factors warrant consideration when interpreting these results. Imagine Math is a supplemental, digital program offered as an optional resource to educators statewide, and districts and schools vary substantially in whether, and how



consistently, they adopt and implement it. Within schools, students often differed in both whether they used the program and the extent of their engagement, reflecting differences in scheduling decisions, teacher discretion, and local priorities. This variability in implementation means that this evaluation captures the impact of Imagine Math as it was actually used across Idaho, not under conditions of uniform adoption or structured dosage.

Given this context, the positive impacts observed for Imagine Math users should be interpreted as evidence of the program's effectiveness under real-world, heterogeneous implementation conditions. The stronger associations observed for high-usage students highlight the importance of consistent engagement, but they do not imply causality, nor do they assume uniform instructional integration across districts or classrooms.

Additionally, the ISAT mathematics assessment captures grade-level proficiency rather than growth, which may limit sensitivity to incremental gains associated with supplemental program use. These quantitative findings should be interpreted alongside the qualitative results presented in a companion CRRE report, which provides essential context regarding implementation conditions, teacher experiences, and factors that shaped program use across Idaho schools. As implementation continues to evolve, future research should examine successive cohorts of students to assess whether impacts strengthen over time and how usage patterns change across districts and grade levels.



REFERENCES

Imagine Learning (2026). *Imagine Math®*.

<https://www.imaginelearning.com/programs/math/math/>

Imagine Learning (2026). *Usage recommendations for Imagine Math 3+*.

<https://help.imaginelearning.com/hc/en-us/articles/360057794374-Usage-recommendations-for-Imagine-Math-3>

National Center for Education Statistics (2024). *Digest State Dashboard, Idaho*.

<https://nces.ed.gov/programs/digest-dashboard/state/idaho>



APPENDIX A: Descriptive Analyses

Table A1

*Overall Spring 2024 and Spring 2025 ISAT Mathematics Descriptive Statistics
(Treatment vs. Comparison)*

Group	N	Spring 2024 ISAT scale score mean (SD)	Spring 2025 ISAT scale score mean (SD)	ISAT gain score mean ⁴ (SD)
Imagine Math	17,226	2475.07 (123.31)	2506.01 (128.22)	30.94 (70.63)
Comparison	17,226	2475.23 (127.41)	2502.91 (143.95)	27.68 (83.08)
Total	34,452	2475.15 (125.15)	2504.46 (136.32)	29.31 (77.12)

Table A2

Unadjusted Spring 2025 ISAT Mathematics Scores by Grade Level

Grade	Group	N	Spring 2024 ISAT scale score mean (SD)	Spring 2025 ISAT scale score mean (SD)	ISAT gain score mean (SD)
4	Imagine Math	5,003	2437.23 (92.88)	2484.66 (95.90)	47.44 (52.14)
	Comparison	5,003	2438.45 (92.97)	2482.30 (125.67)	43.85 (90.18)
5	Imagine Math	4,324	2478.94 (104.28)	2507.24 (121.09)	28.30 (66.55)
	Comparison	4,324	2479.09 (103.25)	2504.12 (119.51)	25.03 (62.97)
6	Imagine Math	3,607	2494.68 (142.37)	2513.33 (151.72)	18.66 (78.67)
	Comparison	3,607	2495.70 (141.53)	2513.06 (156.36)	17.37 (85.32)
7	Imagine Math	2,261	2491.02 (148.74)	2519.55 (140.28)	28.53 (88.83)
	Comparison	2,261	2487.62 (175.86)	2510.67 (189.79)	22.95 (96.64)
8	Imagine Math	2,031	2507.48 (135.61)	2527.86 (144.59)	20.38 (73.14)
	Comparison	2,031	2507.45 (135.70)	2524.53 (145.82)	17.08 (75.98)

Note: Because ISAT is not vertically scaled, scores should not be compared across grades.

⁴ Gain score represents the change in a student's ISAT mathematics scale score from spring 2024 to spring 2025 (posttest – pretest), providing an indicator of year-to-year growth in mathematics achievement.

**Table A3***Independent t-tests Comparing Treatment and Comparison Students*

Grade	Outcome	Mean difference (T – C)	t-statistic	p-value
4	Spring 2025 ISAT	2.37	1.06	.29
	Gain score	3.59	2.44*	.01
5	Spring 2025 ISAT	3.12	1.21	.23
	Gain score	3.27	2.35*	.02
6	Spring 2025 ISAT	0.27	0.07	.94
	Gain score	1.29	0.67	.50
7	Spring 2025 ISAT	8.98	1.81	.07
	Gain score	5.59	2.02*	.04
8	Spring 2025 ISAT	3.34	0.73	.46
	Gain score	3.30	1.41	.16
All	Spring 2025 ISAT	3.10	2.11*	.03
	Gain score	3.26	3.92***	<.001

Note: * $p < .05$; *** $p < .001$. Tests compare unadjusted means only; they do not adjust for covariates or school-level clustering.

Table A4*Subgroup Unadjusted Descriptive ISAT Mathematics Statistics*

Subgroup	Category	Treatment N	Comparison N	Treatment ISAT spring 2024 mean (SD)	Comparison ISAT spring 2024 mean (SD)	Treatment ISAT spring 2025 mean (SD)	Comparison ISAT spring 2025 mean (SD)
Race/Ethnicity	White	11,836	12,088	2490.77 (121.77)	2489.46 (120.13)	2522.92 (129.21)	2518.84 (135.32)
	Hispanic	4,202	4,142	2435.63 (113.99)	2434.92 (130.26)	2463.41 (112.99)	2458.28 (141.02)
	Other race	1,188	996	2458.17 (136.20)	2470.09 (160.79)	2488.18 (131.61)	2495.12 (205.09)
ELL status	ELL	2,414	2,266	2424.97 (117.65)	2421.93 (124.06)	2452.52 (114.66)	2444.11 (129.43)
	Non-ELL	14,812	14,960	2483.24 (122.28)	2483.30 (125.96)	2514.72 (128.20)	2511.81 (143.95)
SPED status	SPED	2,173	2,283	2362.24 (214.32)	2358.98 (225.99)	2385.93 (217.58)	2372.11 (271.68)
	Non-SPED	15,053	14,943	2491.36 (93.11)	2492.99 (92.38)	2523.34 (97.98)	2522.89 (97.98)



Table A5
Spring 2024 and Spring 2025 ISAT Proficiency Level Distributions

Grade	Group	Year	Level 1 (%)	Level 2 (%)	Level 3 (%)	Level 4 (%)
4	Imagine Math	2024	24.93%	21.81%	29.66%	23.61%
		2025	20.87%	27.02%	27.66%	24.45%
	Comparison	2024	24.63%	22.03%	29.64%	23.71%
		2025	19.91%	28.60%	26.92%	24.57%
5	Imagine Math	2024	21.35%	28.38%	27.84%	22.43%
		2025	27.87%	26.23%	21.46%	24.44%
	Comparison	2024	21.44%	28.54%	27.59%	22.43%
		2025	30.11%	26.83%	18.80%	24.26%
6	Imagine Math	2024	32.27%	26.23%	20.02%	21.49%
		2025	32.16%	27.59%	20.24%	20.02%
	Comparison	2024	32.16%	26.20%	20.04%	21.60%
		2025	31.74%	28.33%	19.79%	20.13%
7	Imagine Math	2024	39.63%	28.31%	17.65%	14.42%
		2025	36.36%	26.67%	20.43%	16.54%
	Comparison	2024	39.63%	28.26%	17.65%	14.46%
		2025	35.38%	28.13%	20.74%	15.75%
8	Imagine Math	2024	41.11%	25.46%	18.96%	14.48%
		2025	42.54%	22.85%	17.43%	17.18%
	Comparison	2024	41.11%	25.46%	18.96%	14.48%
		2025	44.07%	23.73%	15.31%	16.89%
All	Imagine Math	2024	29.40%	25.66%	24.35%	20.59%
		2025	29.58%	26.40%	22.40%	21.62%
	Comparison	2024	29.32%	25.76%	24.28%	20.64%
		2025	29.83%	27.46%	21.21%	21.50%



APPENDIX B: Full HLM Covariate Estimates

Table B1

Full HLM Coefficient Estimates for Primary Impact Model, Full Sample

Variable	Estimate	Standard error	p-value	Effect size
Imagine Math	5.21***	1.17	<.001	0.04
ISAT spring 2024 scale score	0.87***	0.004	<.001	0.006
Hispanic	-1.40	1.89	.46	-0.01
White	4.60*	1.71	.007	0.03
SPED status	-28.20***	1.29	<.001	-0.21
ELL status	-3.22*	1.49	.03	-0.02
Grade 4	11.50***	1.75	<.001	0.08
Grade 5	-1.88	1.75	.28	-0.01
Grade 6	-5.58***	1.55	<.001	-0.04
Grade 7	4.79*	1.63	.003	0.04
Constant	2502.14***	2.22	<.001	
Student <i>n</i>	34,452			
Number of schools (clusters)	581			
ICC	.02			

Note. * $p < .05$; *** $p < .001$. "Other Race" was omitted from the model due to collinearity. Dummy grade variables (Grades 4-7) were included to adjust for systematic differences in achievement across grades; Grade 8 served as the reference category.

Table B2

Full HLM Coefficient Estimates for Primary Impact Model, Grade 4

Variable	Estimate	Standard error	p-value	Effect size
Imagine Math	6.35*	2.28	.005	0.06
ISAT spring 2024 scale score	0.86***	0.009	<.001	0.008
Hispanic	4.65	3.39	.17	0.04
White	7.40*	3.02	.01	0.07
SPED status	-23.69***	2.30	<.001	-0.21
ELL status	-3.02	2.89	.30	-0.03
Constant	2510.11***	3.23	<.001	
Student <i>n</i>	10,006			
Number of schools (clusters)	409			
ICC	.08			

Note. * $p < .05$; *** $p < .001$. "Other Race" was omitted from the model due to collinearity.

**Table B3***Full HLM Coefficient Estimates for Primary Impact Model, Grade 5*

Variable	Estimate	Standard error	p-value	Effect size
Imagine Math	2.45	1.96	.21	0.02
ISAT spring 2024 scale score	0.94***	0.007	<.001	0.008
Hispanic	-0.57	3.31	.86	-0.005
White	2.49	2.92	.39	0.02
SPED status	-23.56***	2.21	<.001	-0.20
ELL status	-2.63	2.80	.35	-0.02
Constant	2503.23***	3.02	<.001	
Student <i>n</i>	8,648			
Number of schools (clusters)	390			
ICC	.04			

Note. *** $p < .001$. "Other Race" was omitted from the model due to collinearity.

Table B4*Full HLM Coefficient Estimates for Primary Impact Model, Grade 6*

Variable	Estimate	Standard error	p-value	Effect size
Imagine Math	3.77	3.09	.22	0.02
ISAT spring 2024 scale score	0.88***	0.007	<.001	0.006
Hispanic	-2.50	4.49	.58	-0.02
White	3.87	4.08	.34	0.03
SPED status	-42.37***	3.01	<.001	-0.28
ELL status	-3.22	3.41	.35	-0.02
Constant	2502.13***	4.32	<.001	
Student <i>n</i>	7,214			
Number of schools (clusters)	296			
ICC	.06			

Note. *** $p < .001$. "Other Race" was omitted from the model due to collinearity.

Table B5*Full HLM Coefficient Estimates for Primary Impact Model, Grade 7*

Variable	Estimate	Standard error	p-value	Effect size
Imagine Math	9.24*	3.71	.01	0.06
ISAT spring 2024 scale score	0.82***	0.009	<.001	0.005
Hispanic	-9.97	5.98	.10	-0.06
White	2.87	5.55	.61	0.02
SPED status	-29.46***	4.21	<.001	-0.18
ELL status	-3.40	4.28	.43	-0.02
Constant	2505.82***	5.69	<.001	



Student n	4,522
Number of schools (clusters)	227
ICC	.03

Note. * $p < .05$; *** $p < .001$. "Other Race" was omitted from the model due to collinearity.

Table B6

Full HLM Coefficient Estimates for Primary Impact Model, Grade 8

Variable	Estimate	Standard error	p -value	Effect size
Imagine Math	-1.25	3.39	.71	-0.009
ISAT spring 2024 scale score	0.89***	0.01	<.001	0.006
Hispanic	-3.75	4.99	.45	-0.03
White	2.94	4.63	.53	0.02
SPED status	-17.24***	3.71	<.001	-0.12
ELL status	-3.56	3.60	.32	-0.02
Constant	2499.02***	4.78	<.001	
Student n	4,062			
Number of schools (clusters)	228			
ICC	.03			

Note. *** $p < .001$. "Other Race" was omitted from the model due to collinearity.



APPENDIX C: Supplemental Subgroup HLM Analyses

Table C1
Spring 2025 ISAT HLM Estimates by SPED Subgroups

Variable	Estimate	Standard error	p-value	Effect size
Imagine Math x non-SPED	3.86*	1.22	.002	0.03
Imagine Math x SPED	13.81***	2.39	<.001	0.10
ISAT spring 2024 scale score	0.87***	0.004	<.001	0.006
Hispanic	-1.34	1.89	.48	-0.01
White	4.70*	1.70	.006	0.03
ELL status	-3.13*	1.49	.04	-0.02
Grade 4	11.46***	1.75	<.001	0.08
Grade 5	-1.88	1.75	.28	-0.01
Grade 6	-5.61***	1.55	<.001	-0.04
Grade 7	4.83*	1.63	.003	0.04
Constant	2502.73***	2.23	<.001	
Student <i>n</i>	34,452			
Number of schools (clusters)	581			
ICC	.02			

Note. * $p < .05$; *** $p < .001$. Estimates represent adjusted treatment effects for SPED and non-SPED students based on the interaction term between treatment and SPED. Across tables in Appendix C, the treatment coefficient listed first in the table (e.g., 3.86) reflects the effect for the reference category (e.g., non-SPED students), and the model interaction term adjusts the effect for the non-reference category (e.g., SPED students), yielding a marginal effect for that category (e.g., 13.81). Grade 8 served as the reference category for the grade dummy variables. "Other Race" was omitted from the model due to collinearity.

Table C2
Spring 2025 ISAT HLM Estimates by Race/Ethnicity Subgroups

Variable	Estimate	Standard error	p-value	Effect size
Imagine Math x White	4.45*	1.30	.001	0.03
Imagine Math x Hispanic	6.62***	1.87	<.001	0.05
Imagine Math x Other race	7.86*	3.35	.02	0.06
ISAT spring 2024 scale score	0.87***	0.004	<.001	0.006
SPED status	-28.21***	1.29	<.001	-0.21
ELL status	-3.23*	1.49	.03	-0.02
Grade 4	11.47***	1.75	<.001	0.08
Grade 5	-1.92	1.75	.27	-0.01



Grade 6	-5.58***	1.55	<.001	-0.04
Grade 7	4.79*	1.63	.003	0.04
Constant	2507.14***	1.63	<.001	
Student <i>n</i>	34,452			
Number of schools (clusters)	581			
ICC	.02			

Note. * $p < .05$; *** $p < .001$. Grade 8 served as the reference category for the grade dummy variables.

Table C3
Spring 2025 ISAT HLM Estimates by ELL Subgroups

Variable	Estimate	Standard error	<i>p</i> -value	Effect size
Imagine Math x non-ELL	4.92***	1.23	<.001	0.04
Imagine Math x ELL	6.94*	2.39	.004	0.05
ISAT spring 2024 scale score	0.87***	0.004	<.001	0.006
SPED status	-28.19***	1.29	<.001	-0.21
Hispanic	-1.39	1.89	.46	-0.01
White	4.61*	1.71	.007	0.03
Grade 4	11.49***	1.75	<.001	0.08
Grade 5	-1.90	1.75	.28	-0.01
Grade 6	-5.59***	1.55	<.001	-0.04
Grade 7	4.79*	1.63	.003	0.04
Constant	2502.29***	2.23	<.001	
Student <i>n</i>	34,452			
Number of schools (clusters)	581			
ICC	.02			

Note. * $p < .05$; *** $p < .001$. Grade 8 served as the reference category for the grade dummy variables. "Other Race" was omitted from the model due to collinearity.



APPENDIX D: Supplemental Fidelity HLM Analyses

The fidelity analyses presented in this appendix examined how varying levels of engagement with Imagine Math related to student mathematics achievement, using the Primary Treatment definition. Across all modeling approaches – including full fidelity (≥ 30 lessons completed; modeled in the Results section), relaxed fidelity (≥ 24 lessons completed), and continuous usage (hours spent using Imagine Math, lessons completed, and lessons passed) – results consistently showed positive, statistically significant associations between deeper program engagement and higher adjusted ISAT mathematics scores. Full covariate estimates are excluded from these supplemental appendix models because they were consistent with all other models, including strong pretest effects and expected demographic associations

Of the 17,226 treatment students in the matched analytic sample, 4,516 students (26.22%) met the full fidelity threshold and 5,509 students (31.98%) met the relaxed fidelity threshold. Students meeting the relaxed fidelity threshold scored 9.57 points higher than matched comparison students and 10.93 points higher than treated non-fidelity peers, indicating that even moderate levels of sustained engagement were associated with meaningfully stronger outcomes.

Relaxed Fidelity Analyses

Table D1

Panel A: Adjusted Fidelity HLM Estimates, Relaxed Fidelity vs. Matched Comparison Students

Variable	Estimate	Standard error	p-value	Effect size
Relaxed Imagine Math fidelity	9.57***	1.78	<.001	0.07
Constant	2498.46***	2.75	<.001	
Student <i>n</i>	22,735			
Number of schools (clusters)	580			
ICC	.02			

Note. *** $p < .001$.

Panel B: Adjusted Fidelity HLM Estimates, Relaxed Fidelity vs. Treated Non-Fidelity Students

Variable	Estimate	Standard error	p-value	Effect size
Relaxed Imagine Math fidelity	10.93***	1.27	<.001	0.09
Constant	2506.35***	2.98	<.001	
Student <i>n</i>	17,226			
Number of schools (clusters)	224			



ICC .02

Note. *** $p < .001$.

Continuous Usage Analyses

Continuous-usage models further demonstrated a clear dose-response pattern consistent with the full fidelity and relaxed fidelity models. Regardless of whether fidelity was modeled categorically or continuously, high-engagement students demonstrated substantially stronger achievement outcomes than matched comparison students and treated non-fidelity peers, underscoring the robustness and consistency of the relationship between sustained Imagine Math usage and ISAT mathematics performance.

Table D2

Panel A: Adjusted HLM Estimates, Imagine Math Usage (in Hours), Matched Sample

Variable	Estimate	Standard error	p -value	Effect size
Imagine Math hours	0.52***	0.07	<.001	0.004
Constant	2502.94***	2.20	<.001	
Student n	34,452			
Number of schools (clusters)	581			
ICC	.03			

Note. *** $p < .001$.

Panel B: Adjusted HLM Estimates, Imagine Math Usage (in Hours), Treatment Sample

Variable	Estimate	Standard error	p -value	Effect size
Imagine Math hours	0.54***	0.07	<.001	0.004
Constant	2504.96***	3.02	<.001	
Student n	17,226			
Number of schools (clusters)	224			
ICC	.02			

Note. *** $p < .001$.

Table D3

Panel A: Adjusted HLM Estimates, Imagine Math Usage (in Lessons Completed), Matched Sample

Variable	Estimate	Standard error	p -value	Effect size
Imagine Math lessons completed	0.15***	0.02	<.001	0.001
Constant	2503.16***	2.20	<.001	
Student n	34,452			
Number of schools (clusters)	581			
ICC	.02			



Note. *** $p < .001$.

Panel B: Adjusted HLM Estimates, Imagine Math Usage (in Lessons Completed), Treatment Sample

Variable	Estimate	Standard error	p -value	Effect size
Imagine Math lessons completed	0.16***	0.02	<.001	0.001
Constant	2505.70***	2.99	<.001	
Student n	17,226			
Number of schools (clusters)	224			
ICC	.02			

Note. *** $p < .001$.

Table D4

Panel A: Adjusted HLM Estimates, Imagine Math Usage (in Lessons Passed), Matched Sample

Variable	Estimate	Standard error	p -value	Effect size
Imagine Math lessons passed	0.36***	0.03	<.001	0.003
Constant	2503.16***	2.20	<.001	
Student n	34,452			
Number of schools (clusters)	581			
ICC	.03			

Note. *** $p < .001$.

Panel B: Adjusted HLM Estimates, Imagine Math Usage (in Lessons Passed), Treatment Sample

Variable	Estimate	Standard error	p -value	Effect size
Imagine Math lessons passed	0.43***	0.02	<.001	0.003
Constant	2505.11***	3.03	<.001	
Student n	17,226			
Number of schools (clusters)	224			
ICC	.03			

Note. *** $p < .001$.



APPENDIX E: ITT Analyses

This appendix presents a set of ITT analyses to complement the primary impact analyses. The impact analyses in the main report focused on contrasts between students who completed at least one Imagine Math lesson during the 2024-2025 school year and a matched comparison group that completed no lessons. By contrast, the following analyses used a “pure” treatment definition: students were classified as treatment if they had any amount of Imagine Math program usage during the 2024-2025 school year, regardless of whether they completed any lessons. These ITT analyses are important because they estimate program impacts under the broadest, least restrictive definition of intervention exposure, thereby providing a conservative check on the robustness of the main findings.

The unmatched ITT sample included 87,282 comparison students and 25,139 treatment students, enrolled in 595 distinct schools, with complete pretest and outcome data. Although baseline equivalence was generally strong in the unmatched sample, treatment students had higher proportions of Hispanic and ELL students, while comparison students had higher proportions of White students, making student-level matching important. After 1:1 propensity score matching, the ITT analytic sample included 25,139 treatment students and 25,139 matched comparison students, enrolled in 588 distinct schools, with excellent balance across all demographic and pretest variables (see Table E1).

Table E1

Student-Level Baseline Equivalence: Pretest and Demographics, ITT Sample

Variable	Full sample: Imagine Math mean (SD)	Full sample: Comparison mean (SD)	SMD (Full Sample)	Matched sample: Imagine Math mean (SD)	Matched sample: Comparison mean (SD)	SMD (Matched sample)
Pretest (ISAT scale score 2024)	2472 (142.5)	2481 (228.71)	-0.04	2472 (142.5)	2472 (151.4)	-0.001
Hispanic	0.25 (0.43)	0.19 (0.39)	0.16	0.25 (0.43)	0.25 (0.43)	0.002
White	0.69 (0.47)	0.75 (0.44)	-0.15	0.68 (0.47)	0.69 (0.46)	-0.02
Other race	0.07 (0.26)	0.07 (0.25)	0.01	0.07 (0.26)	0.06 (0.24)	0.03
ELL status	0.15 (0.36)	0.11 (0.31)	0.13	0.15 (0.36)	0.14 (0.35)	0.02
SPED status	0.13 (0.34)	0.13 (0.34)	0.004	0.13 (0.34)	0.14 (0.34)	-0.009
Student <i>n</i>	25,139	87,282	112,421	25,139	25,139	50,278

Note. Demographic characteristics are presented as mean proportions (e.g., 0.50 = 50%).

Across the full analytic sample and grades 4-8, treatment and comparison students



displayed highly similar pretest, posttest, and gain score averages. Differences between groups were small and inconsistent in direction, reinforcing that unadjusted grade-level achievement patterns were comparable under the ITT definition (see Tables E2-E3).

Table E2
Overall Spring 2024 and Spring 2025 ISAT Mathematics Descriptive Statistics (Treatment vs. Comparison), ITT Sample

Group	N	Spring 2024 ISAT scale score mean (SD)	Spring 2025 ISAT scale score mean (SD)	ISAT gain score mean (SD)
Imagine Math	25,139	2472.26 (142.50)	2500.11 (151.36)	27.85 (79.06)
Comparison	25,139	2472.47 (1151.42)	2500.31 (163.51)	27.84 (84.66)
Total	50,278	2472.37 (147.03)	2500.21 (157.55)	27.84 (81.91)

Table E3
Unadjusted Spring 2025 ISAT Mathematics Scores by Grade Level, ITT Sample

Grade	Group	N	Spring 2024 ISAT scale score mean (SD)	Spring 2025 ISAT scale score mean (SD)	ISAT gain score mean (SD)
4	Imagine Math	6,400	2428.18 (117.17)	2474.52 (128.82)	46.34 (67.02)
	Comparison	6,400	2430.77 (115.34)	2476.70 (136.99)	45.93 (79.64)
5	Imagine Math	6,070	2469.46 (128.67)	2495.70 (144.38)	26.24 (69.47)
	Comparison	6,070	2471.03 (128.17)	2496.47 (149.08)	25.43 (84.07)
6	Imagine Math	5,324	2488.03 (145.90)	2506.34 (155.65)	18.30 (74.84)
	Comparison	5,324	2487.78 (145.96)	2506.75 (158.08)	18.97 (95.69)
7	Imagine Math	3,904	2490.01 (160.00)	2512.90 (164.78)	22.89 (110.35)
	Comparison	3,904	2485.08 (209.43)	2508.52 (212.63)	23.44 (80.04)
8	Imagine Math	3,441	2514.67 (159.65)	2531.35 (170.27)	16.68 (73.44)
	Comparison	3,441	2514.60 (159.770)	2531.75 (170.60)	17.15 (76.10)

Note: Because ISAT is not vertically scaled, scores should not be compared across grades.

Across the ITT analyses (illustrated in Tables E4-E7), results tended to be directionally positive but statistically non-significant, contrasting the primary findings reported in the main Results section, which were positive and statistically significant. The one



exception was the SPED differential treatment effect, which remained statistically significant in the ITT subgroup models, consistent with the main matched-sample findings. The additional sensitivity models included for the Primary Treatment sample (modeled in Appendix F) were excluded for the ITT sample because these yielded similar results to the main impact model (directionally positive but statistically non-significant). Grade-specific ITT models were also excluded because no grade-level treatment effects reached statistical significance.

Table E4

Full HLM Coefficient Estimates for Primary Impact Model, ITT Sample

Variable	Estimate	Standard error	p-value	Effect size
Imagine Math	0.73	1.07	.50	0.005
ISAT spring 2024 scale score	0.89***	0.003	<.001	0.006
Hispanic	-5.01*	1.62	.002	-0.03
White	0.80	1.48	.59	0.005
SPED status	-29.68***	1.12	<.001	-0.19
ELL status	-3.50*	1.25	.005	-0.02
Grade 4	14.23***	1.53	<.001	0.09
Grade 5	-1.06	1.51	.48	-0.007
Grade 6	-4.08*	1.30	.002	-0.03
Grade 7	3.97*	1.32	.003	0.03
Constant	2502.85***	1.94	<.001	
Student <i>n</i>	50,278			
Number of schools (clusters)	588			
ICC	.02			

Note. * $p < .05$; *** $p < .001$. "Other Race" was omitted from the model due to collinearity. Dummy grade variables (Grades 4-7) were included to adjust for systematic differences in achievement across grades; Grade 8 served as the reference category.

The adjusted ITT HLM model produced a positive but non-significant treatment estimate (+0.73 points, $p = .50$). Covariate patterns were consistent with all other models, including strong pretest effects and expected demographic associations.

Subgroup analyses (Table E5) showed largely non-significant treatment effects across race/ethnicity, and ELL status. The SPED subgroup, however, demonstrated a statistically significant differential treatment effect (+7.19 points, $p = .001$), consistent with the main matched-sample findings. This suggests that SPED students may benefit more from Imagine Math than their non-SPED peers, even under the broad ITT treatment definition.

**Table E5**

Adjusted HLM Estimates of Imagine Math Treatment Effects by Student Subgroup, ITT Sample

Subgroup	Marginal treatment effect	Standard error	<i>p</i> -value	Effect size	Difference in treatment effects	<i>p</i> -value
Imagine Math x non-SPED	-0.34	1.11	.76	-0.002	—	—
Imagine Math x SPED	6.85*	2.08	.001	0.04	7.19*	.001
Imagine Math x White	0.41	1.18	.73	0.003	—	—
Imagine Math x Hispanic	1.77	1.67	.29	0.01	1.37	.43
Imagine Math x non-ELL	0.55	1.11	.62	0.003	—	—
Imagine Math x ELL	1.75	2.08	.40	0.01	1.21	.57

Note. Marginal treatment effects represent the adjusted treatment effect of Imagine Math within each subgroup. * $p < .05$; *** $p < .001$. ICC = .02 across all models. Student $n = 50,278$; number of school clusters = 588.

Fidelity models using the ITT sample (Tables E6-E7) showed positive and statistically significant effects for full fidelity across both comparisons: full fidelity students scored 9.38 points higher than matched comparison students and 12.94 points higher than treated non-fidelity students. These results mirror the fidelity findings under the Primary Treatment definition. Because the definition of fidelity (≥ 30 lessons completed) remained unchanged, the consistency of these effects is expected: students who reached high levels of engagement demonstrated meaningfully stronger achievement outcomes, regardless of how treatment was defined. Relaxed fidelity models (24+ lessons), although not included in this appendix, produced impacts similar in magnitude to those observed under the Primary Treatment definition, further reinforcing the robustness of the fidelity findings.

Table E6

Panel A: Adjusted Fidelity HLM Estimates, Full Fidelity vs. Matched Comparison Students, ITT Sample

Variable	Estimate	Standard error	<i>p</i> -value	Effect size
Full Imagine Math fidelity	9.38***	2.02	<.001	0.06
ISAT spring 2024 scale score	0.90***	0.004	<.001	0.006
Constant	2500.34***	1.44	<.001	
Student n	29,655			
Number of schools (clusters)	585			



ICC .03

Note. *** $p < .001$.

Panel B: Adjusted Fidelity HLM Estimates, Full Fidelity vs. Treated Non-Fidelity Students, ITT Sample

Variable	Estimate	Standard error	p -value	Effect size
Full Imagine Math fidelity	12.94***	1.47	<.001	0.09
ISAT spring 2024 scale score	0.88***	0.004	<.001	0.006
Constant	2502.97***	2.66	<.001	
Student n	25,139			
Number of schools (clusters)	291			
ICC	.02			

Note. *** $p < .001$.

Continuous-usage models also produced parallel patterns to the Primary Treatment definition, demonstrating that incremental engagement was associated with stronger outcomes: each additional hour of Imagine Math usage was associated with a 0.65-point increase in adjusted ISAT mathematics scores, each additional lesson completed with a 0.17-point increase, and each additional lesson passed with a 0.38-point increase. For the ITT continuous usage analyses, the matched comparison sample was excluded because this specification represents a purist contrast between any program usage and no usage, rather than the matched-sample framework used in the primary model. Despite this difference in analytic design, the direction and magnitude of the continuous-usage effects remained consistent with the primary findings, reinforcing the robustness of the relationship between sustained engagement and mathematics achievement.

Table E7

Adjusted HLM Estimates, Imagine Math Usage, ITT Treatment Sample

Variable	Estimate	Standard error	p -value	Effect size
Imagine Math hours	0.65***	0.07	<.001	0.004
Imagine Math lessons completed	0.17***	0.02	<.001	0.001
Imagine Math lessons passed	0.38***	0.03	<.001	0.003

Note: *** $p < .001$. ICC = .02 across all models. Student n = 25,139; number of school clusters = 291.



APPENDIX F: Supplemental Sensitivity HLM Analyses

To assess the robustness of the primary impact findings, we conducted a series of supplemental sensitivity analyses using the Primary Treatment definition. These models were designed to test whether the estimated treatment effects remained stable under reasonable alternative assumptions, including changes to the covariate structure, adjustments to the handling of school-level differences, and restrictions to the analytic sample. Each analysis modified one key aspect of the original HLM specification to evaluate whether results were sensitive to particular modeling choices or data conditions. Across all sensitivity analyses, the direction and magnitude of the treatment effect were highly consistent with the primary HLM results, providing additional evidence for the credibility and stability of the study's conclusions.

Table F1

Covariate-Stripped HLM Estimates for Spring 2025 ISAT Mathematics Achievement

Variable	Estimate	Standard error	p-value	Effect size
Imagine Math	6.73***	1.23	<.001	0.05
ISAT spring 2024 scale score	0.90***	0.003	<.001	0.007
Constant	2503.60***	0.87	<.001	
Student <i>n</i>	34,452			
Number of schools (clusters)	581			
ICC	.03			

Note: *** $p < .001$.

This covariate-stripped HLM model shows that Imagine Math participation remained a significant positive predictor of spring 2025 ISAT mathematics achievement even when all demographic covariates were removed. The adjusted treatment effect of +6.73 points (ES = 0.05) indicates that the program's impact is robust to model specification, consistent with the fully adjusted primary impact model. In other words, the positive association between Imagine Math use and mathematics performance persists even under the simplest analytic conditions.

Table F2

School Fixed-Effects Models for Spring 2025 ISAT Mathematics Achievement

Outcome	Treatment Estimate	Standard error	p-value	N students	N schools
ISAT spring 2025 scale score	7.22***	1.53	<.001	34,452	581

Note: *** $p < .001$.

School fixed-effects models included a dummy indicator for each school to fully absorb between-school differences in achievement. This specification tested whether treatment effects persisted when comparisons were made strictly within schools. Because school fixed effects replaced the random intercept, individual school coefficients are not shown



here; only the treatment effect and model summary statistics are reported. Under school fixed-effects, the treatment effect remained positive and statistically significant.

Table F3

School-Centered Pretest Models for Spring 2025 ISAT Mathematics Achievement

Outcome	Treatment estimate	Standard error	p-value	N students	N schools
ISAT spring 2025 scale score	6.19***	1.50	<.001	34,452	581

Note: *** $p < .001$.

Centering pretest scores within schools adjusted the model to estimate treatment effects relative to each school's achievement distribution rather than the overall sample distribution. This specification tested whether treatment effects depend on between-school differences in pretest achievement. Results showed that treatment estimates remained stable and significant under this alternative centering approach.

Table F4

HLM Estimates Excluding Schools in the Top and Bottom 10% of Pretest Means: Spring 2025 ISAT Mathematics Achievement

Variable	Estimate	Standard error	p-value	Effect size
Imagine Math	4.32*	1.32	.001	0.03
ISAT spring 2024 scale score	0.87***	0.004	<.001	0.006
Constant	2502.21***	2.47	<.001	
Student n	27,924			
Number of schools (clusters)	443			
ICC	.03			

Note: * $p < .05$; *** $p < .001$.

The sensitivity analysis excluding schools with extreme pretest averages – those in the top and bottom 10% of baseline achievement – showed that the estimated treatment effect was reduced in magnitude but remained positive and statistically significant. After removing these schools, Imagine Math students scored 4.32 points higher than matched comparison students ($p = .001$), indicating that the primary findings were not driven by schools with unusually high or low pre-intervention performance. This consistency reinforces the robustness of the overall HLM results and suggests that the positive treatment effect persists even under more restrictive sample conditions.