

EXECUTIVE SUMMARY

3P Learning contracted with LearnPlatform by Instructure, a third-party edtech research company, to examine the relationship between usage of Mathletics and student math achievement. LearnPlatform designed the study to satisfy Level II requirements (Moderate Evidence) according to the Every Student Succeeds Act (ESSA).

Study Sample, Measures, and Methods

This study occurred during the 2022–23 school year and included 1,885 grade 3–5 students from eight elementary schools in one district. The grade 3–5 sample comprised 1,561 students who used Mathletics and 324 students who did not use the program. According to demographic data provided by the district, the racial breakdown of students in the sample was as follows: White (57%), Hispanic (19%), African American (12%), and multi-racial (10%). Females made up 49% of the group, while males accounted for 51%. In addition, the percentages of students enrolled in each of the grades was as follows: third grade (32%), fourth grade (34%), and fifth grade (35%).

Researchers used two measures to provide insights into Mathletics implementation and evidence about potential impacts of Mathletics on student learning outcomes: Mathletics usage data and NWEA MAP® scaled scores for mathematics.

Researchers used a variety of quantitative analytic approaches to answer the research questions. First, researchers used descriptive statistics to examine participant characteristics and support implementation analyses. Researchers then used two-level multilevel modeling analysis (i.e., students nested in schools) to examine how Mathletics use related to student math end-of-year outcomes controlling for their beginning-of-year scores. The analyses included student-level covariates to control for potential selection bias (e.g., grade-level, gender). Researchers also used two-level multilevel modeling analysis to examine whether there were any differences between students who used Mathletics during the 2022–23 school year and a statistically similar sample of students who did not use the program. In addition, researchers calculated standardized effect sizes (Hedges' *g*) to determine the magnitude of changes in student outcomes.

Student Outcomes

- ✓ Grades 3–5 students who completed between 251–724 total activities (high use) in Mathletics had higher end-of-year NWEA MAP® achievement compared to students who completed between 1–105 total activities (low use; $p < .001$).
- ✓ Grades 3–5 students who completed between 106–250 total activities (moderate use) in Mathletics had higher end-of-year NWEA MAP® achievement compared to students who completed between 1–105 total activities (low use; $p < .001$).
- ✓ Grades 3–5 students who used Mathletics had higher scores on NWEA MAP® at the end of the year than students who did not use the program (effect size = 0.11; $p < .001$). This means that a grade 3–5 student at the 50th percentile, using Mathletics would result in them moving to the 54 percentile on average.

Conclusions

This study provides results to satisfy ESSA evidence requirements for Level II (Moderate Evidence) given the study design and positive, statistically significant findings for Grade 3–5 students.

TABLE OF CONTENTS

Introduction	3
Methods	4
Program Implementation Findings	6
Student Findings	7
Conclusions and Recommendations	9
Appendix A. Mathletics Logic Model	10
Appendix B. Additional Information on Study Design and Methods	11
Appendix C. Additional Information on Outcome Findings	13

Introduction

3P Learning recognizes that K–8 math programs do not provide a wide variety of standards-aligned practice and fluency questions that are tailored to students' specific learning needs and delivered to students in an engaging way. Educators who want to differentiate instruction often have to source or create and score their own activities which can be time-consuming. The Mathletics program provides a gamified online math program for students with practice activities that foster a positive problem-solving attitude towards math.

As part of their ongoing efforts to demonstrate the efficacy of Mathletics, 3P Learning contracted with LearnPlatform by Instructure, a third-party edtech research company, to examine the relationship between usage of Mathletics and student achievement. After developing a logic model for Mathletics (Shah & Lee, 2024; Appendix A), LearnPlatform by Instructure designed the study with the following research questions to satisfy Level II requirements (Moderate Evidence) according to ESSA:

Program Implementation Research Questions

1. Overall, how many Mathletics activities were completed by students during the 2022–23 school year?
2. Among Mathletics users, what were the usage patterns?

Effectiveness Research Questions

After controlling for students' prior math achievement and demographics,

3. How were different Mathletics usage patterns related to students' spring 2023 math achievement?
 - a. Which usage pattern(s) of Mathletics had the greatest impact on students' spring 2023 math achievement?
4. What was the overall impact of Mathletics on students' spring 2023 math achievement?

Methods

This section of the report briefly describes the setting, participants, measures, and analysis methods.

Setting

The study included one school district in the Midwestern US and an analysis sample of grades 3–5 students across eight elementary schools.

Participants

There were 1,885 students in the final analytic sample. According to demographic data provided by the district, the racial breakdown of students in the sample was as follows: White (57%), Hispanic (19%), African American (12%), and multi-racial (10%). Females made up 49% of the group, while males accounted for 51%. In addition, the percentages of students enrolled in each of the grades was as follows: third grade (32%), fourth grade (34%), and fifth grade (35%; See Appendix B for details).

Measures

This study included the following measures to provide insights into Mathletics implementation and evidence about the potential impacts of Mathletics on student achievement.

Mathletics Usage Metrics. Researchers utilized 2022–23 student-level usage (i.e., total activities completed). These usage data informed the extent to which students used Mathletics during the school year and whether students' use of Mathletics related to learning outcomes on NWEA MAP® Mathematics.

NWEA MAP® Mathematics Scores. NWEA MAP® RIT scales are stable, equal interval scales that use individual item difficulty values to measure student achievement independent of grade level. The scores are vertically scaled so that student scores can be compared over time and across grade levels. The RIT scale ranges from 100–350. Researchers used the RIT scale score as an overall measure of math achievement at two time points: pretest (i.e., fall 2022) and posttest (i.e., spring 2023).

Data Analysis

Researchers used a variety of quantitative analytic approaches to answer the research questions. First, researchers used descriptive statistics to examine participant characteristics and support implementation analyses. Researchers then used two-level multilevel modeling analysis (i.e., students nested in schools) to examine how Mathletics use related to student math end-of-year outcomes controlling for their beginning-of-year scores. The analyses included student-level covariates to control for potential selection bias (e.g., grade-level, gender). Researchers also used two-level multilevel modeling analysis to examine whether there were any differences between students who used Mathletics during the 2022–23 school year and a statistically similar sample of students who did not

use the program. In addition, researchers calculated standardized effect sizes (Hedges' g) to determine the magnitude of changes in student outcomes.

Baseline Equivalence

To ensure the validity of the study's findings and to adhere to What Works Clearinghouse's quasi-experimental study standards, the researchers assessed the equivalence of students' beginning-of-year assessment scores and demographic characteristics (i.e., grade, gender, race, IEP status, SES, ELL status, and gifted status), between student usage groups (i.e., students who used Mathletics and students who did not use Mathletics). Baseline differences for beginning-of-year assessment scores, were significant (effect size = 0.09). Baseline differences with an effect size between 0.05 and 0.25 must include acceptable statistical adjustments in analyses according to What Works Clearinghouse (WWC) criteria (WWC, 2022). Therefore, beginning-of-year assessment scores were statistically controlled for in the final model. See Appendix B for more details regarding baseline equivalence.

Program Implementation Findings

The charts below highlight Mathletics use during the 2022–23 school year based on 3P Learning’s internal usage data (Table 1). Overall, grades 3–5 users completed an average of 114 Mathletics activities (SD = 94) in total.

Table 1: Average Mathletics Usage by Grade

		Gr 3	Gr 4	Gr 5
	Number of Mathletics users	504	519	538
	Average Mathletics activities completed	100	118	122

Researchers conducted a *k*-means cluster analysis to group students by similar levels of Mathletics usage based on the number of total activities completed. For total activities completed, grades 3–5 students fell into three usage categories ranging from low usage (mean = 53, range = 1–105 total activities), to moderate usage (mean = 158, range = 106–250 total activities), and high usage (mean = 343, range = 251–724 total activities; Figure 1).

Forty-three percent of grades 3–5 students completed 106 total activities or more in Mathletics.

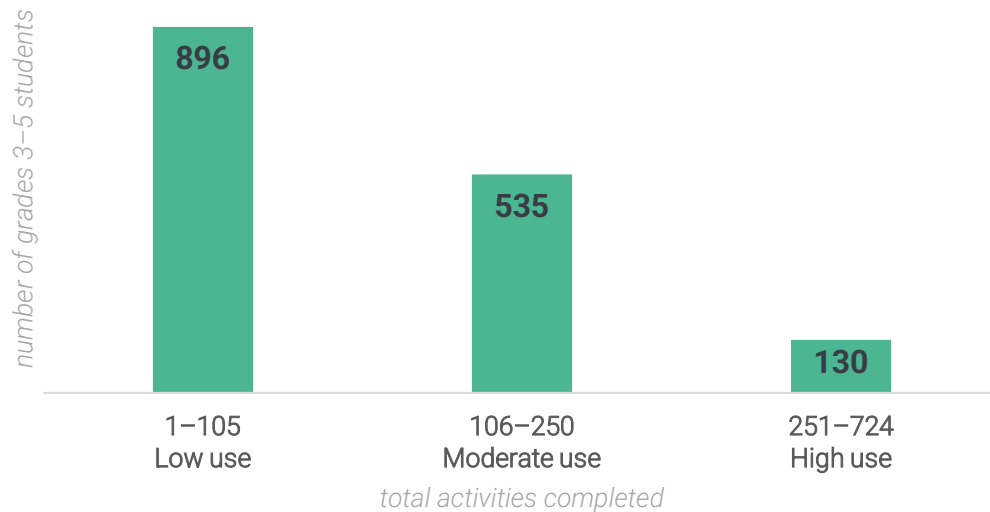


Figure 1. Overall distribution of total activities completed in Mathletics by grades 3–5 students.

Student Findings

Researchers examined whether greater usage of Mathletics related to higher end-of-year NWEA MAP® mathematics achievement using two-level multilevel modeling analysis (i.e., students nested in schools) that included beginning-of-year NWEA MAP® achievement, grade, gender, IEP status, ELL status, gifted status, and 504 status as covariates. To allow for better interpretability of results, marginal means charts are presented below. The orange vertical lines at the top of each bar represent a 95% confidence interval (see Appendix C for more details about the model and the corresponding Hedges' *g* effect sizes).

Overall Relationship Between Total Activities Completed and Grades 3–5 Students' Math Achievement on NWEA MAP®

Key Finding 1. Overall, grades 3–5 students who completed more than 106 total activities (moderate and high use) in Mathletics had higher end-of-year NWEA MAP® achievement compared to students who completed between 1–105 total activities (low use; effect size 0.15 and 0.03, respectively) (Figure 2).

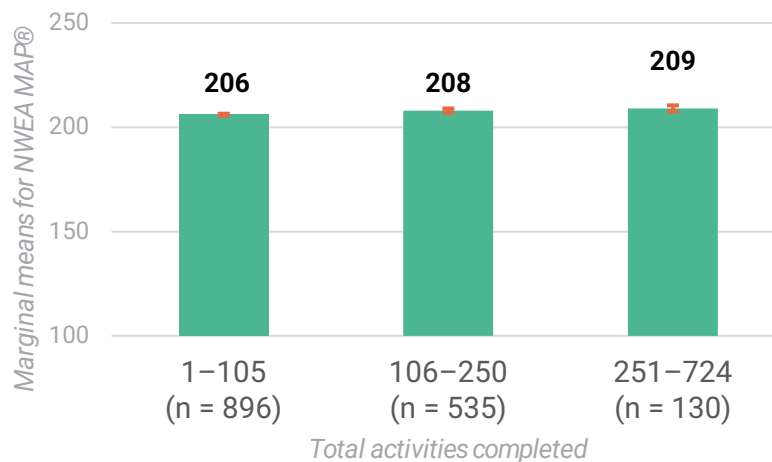


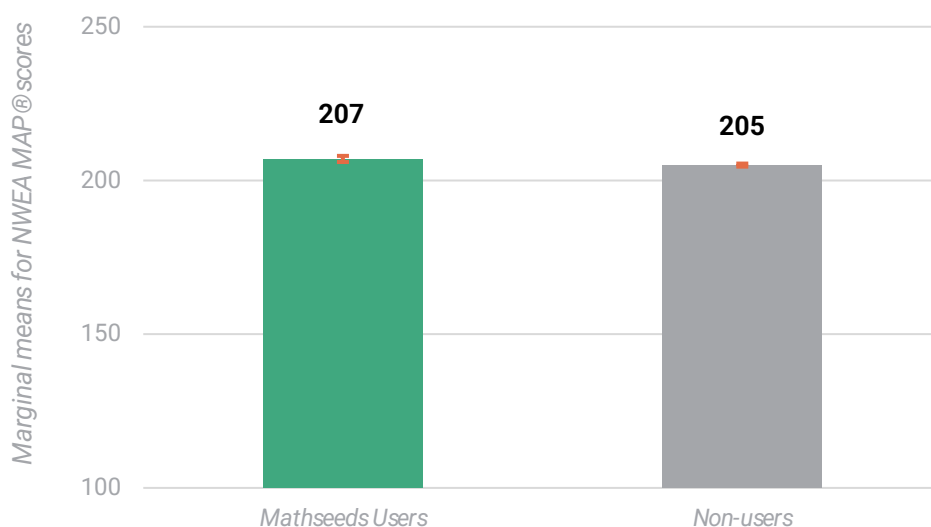
Figure 2. Marginal means for NWEA MAP® for Grades 3–5 Mathletics users ($n = 1,561$) by usage group.

Differences Between Grades 3–5 Students Who Used Mathletics and Students Who Did Not Use the Program at the End of the Year

Researchers also conducted two-level multilevel modeling analysis to examine whether there were any differences between students who used Mathletics during the 2022–23 school year and a statistically similar sample of students who did not use the program. Results at the end-of-the-year show that students who used Mathletics scored statistically significantly higher on end-of-the-year NWEA MAP® than students who did not use the program (see Key Finding 2).

Key Finding 2. Grades 3–5 students who used Mathletics had higher scores on NWEA MAP® at the end of the year than grades 3–5 students who did not use the program (effect size = 0.11). The RIT scale score difference of 2 equated to a 4 percentile point improvement. Such that, for a grade 3–5 student at the 50th percentile, using Mathletics would result in them moving to the 54th percentile on average (i.e., a 4 percentile point improvement; $p < .001$).

If a comparison student, at the 50th percentile, had used Mathletics, they would have been expected to perform at the 54th percentile



Note: The orange vertical lines at the top of each bar represent a 95% confidence interval.

Conclusions and Recommendations

In sum, the findings support a relationship between Mathletics usage and improved math skills for grades 3–5 students. Furthermore, students who used Mathletics scored statistically significantly higher on the end-of-the-year NWEA MAP® assessment than students who did not use the program. Given the set of positive outcome findings, this study provides results to satisfy ESSA evidence requirements for Level II (Moderate Evidence). Specifically, this study met the following criteria for Level II:

- ✓ Comparative design
- ✓ Proper design and implementation
- ✓ Statistical controls through covariates
- ✓ At least one statistically significant, positive finding

Researchers recommend the following next steps:

- Grades 3–5 students who completed at least 105 total activities had higher scores on the end-of-year NWEA MAP® assessment. 3P Learning should continue to explore ideal implementation at other sites using this baseline information.
- Given the positive findings from the comparative study, 3P Learning should consider recruiting a district that has enough Mathletics users enrolled in other grades to better understand how math achievement for these students compares to that of students using other math programs in those grades.

Acknowledgements

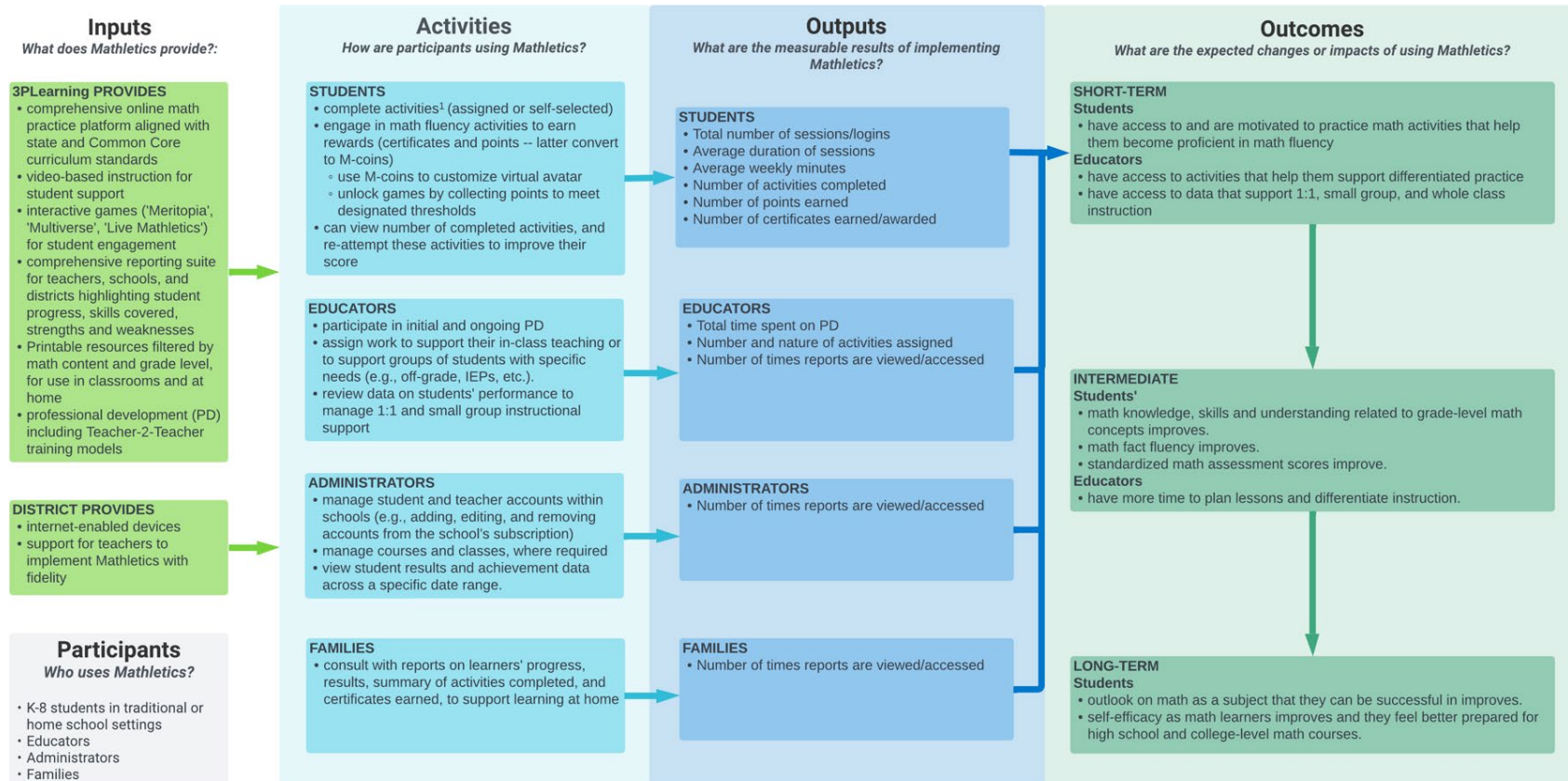
The authors would like to extend their deepest thanks to Avery Wall who supported the preparation of data for this report.

Appendix A. Mathletics Logic Model



Mathletics Logic Model

Problem Statement: K-8 math programs do not provide a wide variety of standards-aligned practice and fluency questions that are tailored to students specific learning needs and delivered to students in an engaging way. Educators who want to differentiate instruction often have to source or create and score their own activities which can be time consuming. Mathletics offers educators and students an easy-to-use platform that hosts engaging, standards-aligned individualized practice and reinforcement activities of essential mathematics concepts from K-8.



Appendix B. Additional Information on Study Design and Methods

Baseline Equivalence

Researchers conducted baseline equivalence analyses to determine whether there were baseline differences in characteristics between students who used Mathletics and students who did not use the program during the 2022–23 school year. Specifically, researchers used chi-square analyses on student-level demographics and a two-level multilevel model analysis on beginning-of-year scores. As noted in Table B1, there were statistically significant demographic differences between groups regarding gender, race, individualized education program (IEP) status, socioeconomic status (SES), and gifted status.

Table B1. Baseline Equivalence Analysis of Student-Level Demographics by User Group

Characteristics	Users (<i>n</i> = 1,561)		Non-users (<i>n</i> =324)		Chi-squared	<i>p</i> -Value	Effect Size
	Percent	<i>N</i>	Percent	<i>N</i>			
<i>Grade</i>							
3	32	504	29	93	1.90	.387	0.03
4	33	519	36	118			
5	34	538	35	113			
<i>Gender</i>							
Male	50	779	57	184	5.09	.024	-0.05
Female	50	782	43	140			
<i>Race</i>							
Black	12	182	15	47	17.73	.007	0.10
Hispanic	19	301	18	59			
Multi-race	9	144	15	49			
White	58	911	50	163			
<i>Individualized Education Program (IEP)</i>							
IEP Status	20	317	25	82	4.02	.045	-0.05
No IEP Status	80	1,244	75	242			
<i>Socioeconomic Status (SES)</i>							
SES Status	5	84	13	41	22.92	<.001	-0.11
No SES Status	95	1,477	87	283			
<i>English Language Learner (ELL)</i>							
ELL Status	8	126	8	26	0.00	.977	0.00

Characteristics	Users (n = 1,561)		Non-users (n =324)		Chi-squared	p-Value	Effect Size
	Percent	N	Percent	N			
No ELL Status	92	1,435	92	298			
Gifted							
Gifted Status	5	78	2	5	7.60	.006	0.06
No Gifted Status	95	1,483	98	319			

As noted in Table B2, results from the two-level multilevel modeling analysis reveal that there was a statistically significant difference between student groups in terms of math achievement at baseline, indicating that researchers need to include MAP NWEA® beginning-of-year math achievement in the outcome model.

Table B2. Baseline Equivalence Analysis of MAP NWEA® Beginning-of-Year Math Scores

Predictor	Unstd. Beta Coefficient	Standard Error	Test statistic	p-value
MAP NWEA® beginning-of-year math achievement (effect size = 0.09)	1.54	0.74	2.07	.038
Grade	9.41	0.34	28.03	<.001
Gender	2.96	0.55	5.38	<.001
IEP	-15.41	0.68	-22.6	<.001
SES	-3.89	1.10	-3.53	<.001
ELL	-9.01	1.05	-8.54	<.001
Gifted	18.07	1.34	13.47	<.001
504 Status	-3.15	1.22	-2.59	.010
School-level random effects	5.49	3.06	50.63	<.001

Appendix C. Additional Information on Outcome Findings

Table C1. NWEA MAP® unadjusted, raw means by Grade for Fall 2022 and Spring 2023

Sample	N	Fall 2022 Mean	SD	Spring 2023 Mean	SD
Overall Grade 3–5	1,561	193.9	16.4	207.1	15.3
Grade 3	504	183.7	13.1	200.2	12.9
Grade 4	519	194.7	13.9	208.0	13.8
Grade 5	538	202.7	15.9	212.7	16.2

Overall Relationship Between Total Activities and Grades 3–5 Students' Math Achievement on NWEA MAP®

Table C2. Grades 3–5 Students' Math Achievement on NWEA MAP® by Total Activities on Mathematics

Predictor	Unstd. Beta Coefficient	Standard Error	Test statistic	p-value
Moderate Use vs. Low Use (Hedges' $g = 0.15$)	2.20	0.42	5.17	<.001
High Use vs. Low Use (Hedges' $g = 0.03$)	2.69	0.72	3.72	<.001
High Use vs. Moderate Use (Hedges' $g = 0.03$)	0.49	0.72	0.68	.495
Fall 2022 NWEA MAP® percentile scores	0.78	0.02	49.14	<.001
Grade	-1.28	0.27	-4.74	<.001
Gender	1.39	0.37	3.73	<.001
IEP	-2.46	0.52	-4.73	<.001
ELL	-2.33	0.72	-3.25	.001
Gifted	3.39	0.90	3.78	<.001
504 Status	-2.00	0.83	-2.4	.016
School-level random effects	0.55	0.42	6.76	.005

Difference Between Grades 3–5 Students who used Mathletics and Students Who Did Not Use the Program at the End of the Year

Researchers conducted a two-level multilevel modeling analysis to examine whether there were any differences between students who used Mathletics during the 2022–23 school year and a statistically similar sample of students who did not use the program. In addition to the outcome of interest (i.e., end-of-year NWEA MAP® scores), the model included the student group, beginning-of-year NWEA MAP® scores, and student demographic characteristics (Table C3).

Table C3. Differences between NWEA MAP® End-of-the-Year Scores by Student Group

Predictor	Unstd. Beta Coefficient	Standard Error	Test statistic	p-value
Students who used Mathletics vs. Students who did not use the program (Hedges' $g = 0.11$)	1.79	0.47	3.83	<.001
Fall 2022 NWEA MAP® percentile scores	0.82	0.01	56.11	<.001
Grade	-1.52	0.25	-6.01	<.001
Gender	1.24	0.35	3.54	<.001
IEP	-2.72	0.48	-5.63	<.001
ELL	-2.11	0.68	-3.11	.002
Gifted	2.81	0.89	3.15	.002
School-level random effects	0.84	0.55	14.36	<.001