STEM Action Center

Grant Program

Annual Evaluation Report

ADDENDUM

2014-15

Prepared for the Utah STEM Action Center at the

Governor's Office of Economic Development

Presented by

Sarah Brasiel, Ph.D., Utah State University

Taylor Martin, Ph.D., Utah State University

Department of Instructional Technology and Learning Sciences

OCTOBER 2015

Note: Dr. Taylor Martin is currently serving as Program Director for one year at the National Science Foundation.

The opinions, findings, and conclusions expressed in this publication are those of the authors and not necessarily those of the Utah STEM Action Center, the Governor's Office of Economic Development, Utah State University, or the National Science Foundation.

For additional information about this addendum or the full report, contact:

Sarah Brasiel, Senior Researcher, Associate Director for the Active Learning Lab
Department of Instructional Technology and Learning Sciences
Utah State University
2830 Old Main Hill, EDUC 215
Logan, UT 84322
sarah.brasiel@usu.edu
(435)797-9872
www.activelearninglab.org

Recommended Citation:

Brasiel, S. & Martin, T. (2015). *STEM Action Center Grant Program Annual Evaluation Report Addendum: 2014-15*. Logan, UT: Utah State University, Department of Instructional Technology and Learning Sciences.

Table of Contents

Background

In the main evaluation report we shared findings from our analysis of product usage by students and teacher feedback during their participation in a STEM Action Center Grant program. In this addendum, we provide the findings from an analysis of the impact of this participation on students' achievement. It is important to understand that we were not able to include all students participating in the grants in our analysis. First, in order to collect data on their achievement from the state, we had to have each student's State Student Identifier (SSID). In February, after districts/charters had a few months of usage, we provided the district/charter leaders with a list of students' user names and showing evidence of usage through the end of February. We asked that they add to this file the students' SSIDs and then upload the file to our secure portal. At this time, there were approximately 121,364 students with usage data.

By the end of the school year (spring 2015) there were 150, 367 students using the products; however, we did not include these additional 29,003 students in our analysis, because they were exposed to only a few months of usage. Although our request to districts and charter schools was for SSIDs for the 121,364 students, we only received SSIDS for 101,756 students. Once we had this final list of SSIDs, we removed data for students whose parents declined having their data included in the analysis (218 students). We submitted this remaining set of 101,538 SSIDs to the Utah State Office of Education.

The reason there were only 74,627 students in the SAGE data file provided to us by the state, when we had given them a list of 101,538 SSIDs, is because some of the students did not take any SAGE Assessment the year of interest (2014-15) or the prior year (2013-14). For

example, students in grades K-3 were not included in this file, because they did not have a SAGE assessment for the prior year (2013-14). Also, students in grade 11 and 12 who had finished their required assessments (prior to spring 2015) did not take the SAGE during 2014-15 were not included in the data file. Of the remaining 74,627 students we received in the state data file, a majority of the students had complete data for mathematics prior year and current year. Less complete data was available for Language Arts and Science. Complete demographic information was provided. Once we reduced the sample to the students with complete achievement data, there were only 45,815 students remaining with complete data. These students were matched to similar students in the state data file who were not participating in any STEM Action Center grant program.

Table 1. Summary of Change in Sample Size Resulting in Final Analytic Sample

Time Period	Sample Size
Usage Through February	121,364
SSIDs received from districts/charters	101,756
SSIDs for students whose parents did not decline including them in the analysis	101,538
SSIDs provided to state requesting data	101,538
SSIDs in data provided by state	74,627
SSIDs with complete data	45,815

Next we provide the impact findings in the context of usage of students participating in the grant program. We summarize this information for two groups of students. The first group is the full sample of students with any evidence of product usage. The second sample is the group of students who met the fidelity benchmark, if one was available from the product provider. The fidelity group analysis is the only analysis that should be used to compare effects across products. The analysis of the full sample, results in information about the impact of the less than

satisfactory implementation of these products, which should not be used to draw conclusions about any of the products, since the implementation was not at the fidelity level. We are working with the STEM Action Center to set expectations of usage for schools so that we will have a greater amount of students meeting the fidelity benchmark during the 2015-16 school year.

K-12 Mathematics Grants

Before considering impact on achievement, it is important to understand as context that most students in this impact analysis only used the math technology product for approximately five months. Most licenses did not get distributed to schools and begin to be used until November. Not all providers tracked usage by time, so we asked them to set their own benchmark for usage. In the next pages we provide a summary of the usage for the full analytic sample of students with any usage and a summary of the usage for students meeting the fidelity benchmark. These students all had complete data from the prior year and the current year to be included in this analysis. For five of the eleven products there was insufficient numbers of students with data to conduct an impact analysis: Cognitive Tutor, EdReady, Odyssey Math, Reflex, and SuccessMaker. Based on the usage information in Table 2, we note that only about 10 percent of students in the full analytic sample met the fidelity benchmark.

Table 2. Number of Students and Average Usage for Full and Fidelity Analytic Samples

	Full Analytics Sample		Fidelity Analytic Sample	
Product	**Number of Students**	**Average Usage**	**Number of Students**	**Average Usage**
ALEKS	27,190	835 minutes	633	2,329 minutes
Catchup Math	254	86 minutes	32	474 minutes
iReady	3,981	302 minutes	190	1,317 minutes
MathXL	318	1,670 minutes	—	—

Product	Full Analytics Sample		Fidelity Analytic Sample	
	Number of Students	Average Usage	Number of Students	Average Usage
ST Math	5,858	20 lab logins	801	76 lab logins
Think Through Math	6,896	19 lessons	2,814	70 lessons
Total	**44,497**	—	**4,470**	—

When we received the state data file, we were notified that the student scale score could only be used if we did a separate analysis by grade. The benefit of the scale score is that it has a lot of variability across students which can be helpful in best understanding the impact of a program. To compare students across grades, given the limitations of the data, would need to use a proficiency score, which came in four levels: 1,2,3,4. Using this type of outcome the interpretation would be challenging to explain simply to the broad stakeholder audience, so in consultation with several methodologists we decided to recode proficiency into a 0/1 coded variable with 1 indicating met proficiency and 0 indicating not met proficiency. Using this outcomes of math proficiency spring 2015, we used logistic regression to compare the proficiency of students in the grant program to similar students in the state. These students were matched based on their prior year state SAGE Mathematics assessment scale score, to get the closest possible match of students to be able to detect any effect of the program on student achievement.

After the students were matched, we conducted a baseline comparison of the two groups which we provide in Appendix A. For most products, the matched students were equivalent on their prior year SAGE Mathematics Scale Score, but differed somewhat in demographic characteristics. Therefore we included covariates in our logistic regression model (discussed in Appendix B) to control for differences in student demographics and prior year SAGE reading

and math proficiency ordinal score (1,2,3,4). The full output of the results are available upon request. We focus this report on the key findings which include the following: odds ratio, standard error, *p*-value, effect size, and 95 percent confidence interval of the odds ratio.

The odds ratio can be somewhat challenging to interpret it simply, anything greater than 1.0 favors the group of students in the grant program using the technology and anything less than 1.0 favors the comparison students not in the grant program. The standard error and the related *p*-value of statistical significance have been adjusted for clustering, since students are nested in schools using these products. A *p*-value of .05 or less notes a difference between the grant students and comparison students that was statistically significant. One issue with using a *p*-value is that it is influenced by sample size. A researcher could have a really large sample and get a significant *p*-value, but the actual difference is not very meaningful. Therefore, in research many are now using an effect size to better understand if there are meaningful differences, since an effect size is not influenced by sample size as *p*-value is.

Across studies of education interventions or programs at all grade levels, 0.25 has become an acceptable standard for an educationally meaningful effect size of a program that shows promise in effecting student achievement. However, for education technology, which is usually more of a supplement to a regular curriculum, prior research has shown the effect size to be more around 0.16[1]. Therefore, we use 0.16 as a benchmark for the expected effect size of an impact when students are using a product as it is intended and meeting the fidelity benchmark.

[1] Expected Effect Size Impact based on findings from Cheung, A. C., & Slavin, R. E. (2013). The effectiveness of educational technology applications for enhancing mathematics achievement in K-12 classrooms: A meta-analysis. *Educational Research Review*, *9*, 88-113.

Anything greater than 0.16 we consider a meaningful positive differences for students using the particular mathematics technology product.

In Table 3, we provide the results of the logistic regression for the products where there was sufficient sample size to evaluate. There were only two products where achievement differences for students participating in the grant achieved the statistical significance level of $p < .05$: the fidelity sample for ALEKS and the fidelity sample for iReady. However, all products had at least one sample reach or exceed the expected benchmark for having an effect on achievement, which we determined was an effect size equal or greater than 0.16. We believe that the size of the effect of the iReady fidelity sample may be confounded by a small sample size (190 students in each group); therefore, it is best to wait until we have a larger sample meeting the fidelity benchmark to draw conclusions about the impact of use of this product. CatchUp Math had too small of a sample meeting fidelity (32 students) to conduct the analysis on the fidelity sample.

Table 3. Results from the Analysis of Impact of Technology use on Achievement

Product and Sample	Exp (B) odds ratio	Standard Error[a]	Significance Level	Effect Size	95% Confidence Interval for Exp(B) odds ratio	
					Lower	Upper
ALEKS						
Full Sample	1.014	0.026	0.607	0.01	0.964	1.067
Fidelity Sample	1.354	0.144	0.032	**0.18**	0.967	1.897
CatchUp Math						
Full Sample	1.294	0.278	0.333	**0.16**	0.730	2.293
iReady						
Full Sample	0.983	0.063	0.804	-0.01	.861	1.122
Fidelity Sample	2.765	0.279	0.002	**0.62**	1.410	5.423
MathXL						
Full Sample	1.464	0.317	0.078	**0.23**	0.821	2.611
ST Math						
Full Sample	1.125	0.126	0.296	0.07	0.910	1.390
Fidelity Sample	1.483	0.435	0.179	**0.24**	0.849	2.590
Think Through Math						
Full Sample	1.191	0.177	0.239	0.11	0.891	1.593
Fidelity Sample	1.339	0.235	0.097	**0.18**	0.952	1.884

[a]The Standard Error and Significance have been adjusted for clustering.

These findings are also shown in Figure 1 as a graph for ease of comparison. If usage increases this year (2015-16), then it may be easier to detect a statistically significant difference in achievement for more products.

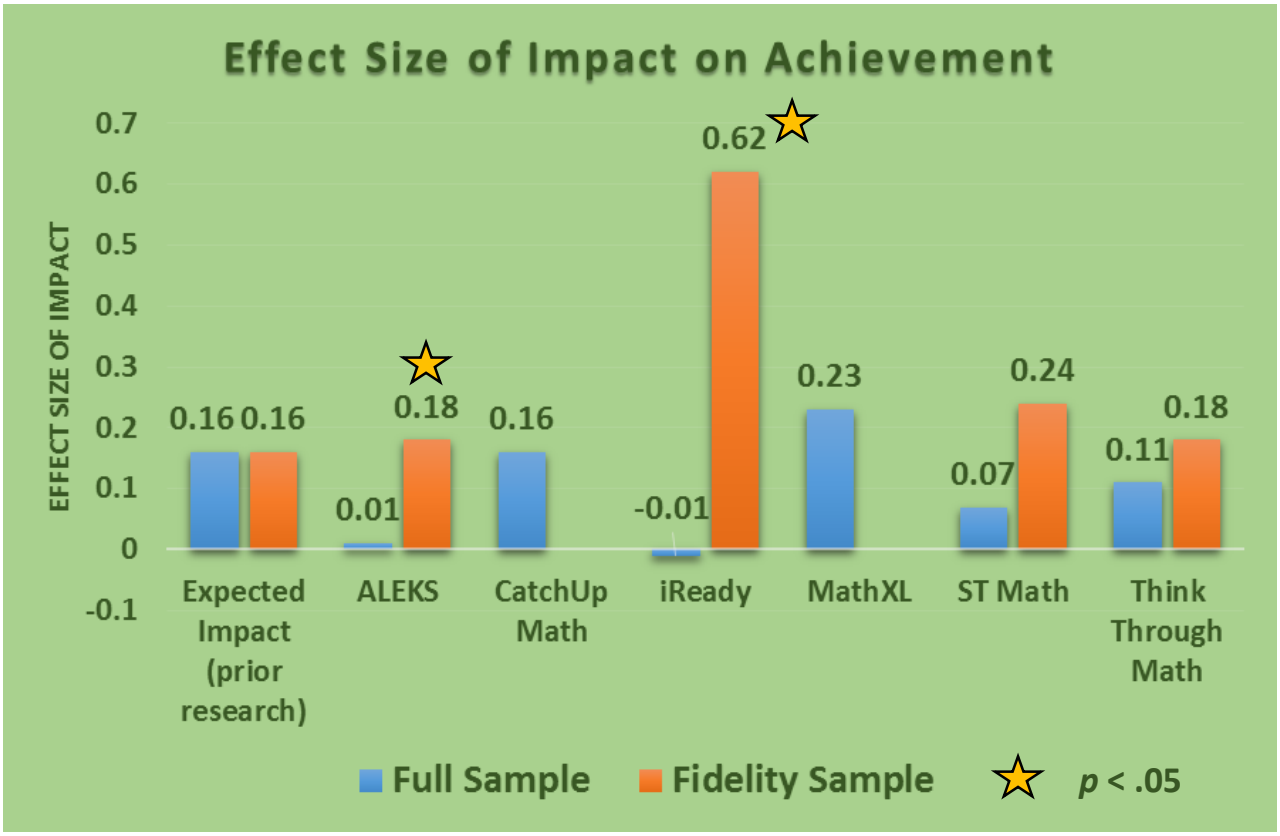

Figure 1. Graph of Results Comparing Effects Across Products

CTE Applied Science Grants

Due to implementation of the CTE grants getting off to a late start, we did not receive many SSIDs for students using the CTE products. Out of the 2,626 SSIDs we received (as shown in Table 4) only 752 of the students had complete data for mathematics to be included in the analysis and even less for science. Due to the small sample size and the minimal usage data available, we recommend that the impact analysis for this grant program be delayed until the end of the 2015-16 school year, when districts have had time to fully implement these grant programs. Any impact, positive or negative, that we might find, could not be attributed to implementation of this grant program given such limited use and such small sample size.

Table 4. Number of Students Using CTE products versus Students with SAGE Data

Product	Number of Students with SSIDs Submitted by Districts	Number of Students with SAGE Data (prior year and Current Year)
ITEEA	128	91
Pitsco	460	333
Project Lead the Way	1,656	234
STEM Academy	382	94
Total	2,626	752

Professional Learning Grants

According to the usage data provided by School Improvement Network for Edivate users, usage of the professional learning platform averaged approximately 10 minutes. For Scholastic, the usage was only reported in April to be approximately 8 activities. Given the late start of this grant program, and the limited usage, we recommend that the impact analysis for this grant program be delayed until the end of the 2015-16 school year, when districts have had time to fully implement their plan. Any impact, positive or negative, that we might find, could not be attributed to implementation of this grant program given such limited use.

High School STEM Industry Certification Grants

The High School STEM Industry Certification Grant program began spring 2015 with four partnership programs. The rest will begin fall 2015. We had relied on students completing a survey about the program to provide the SSIDs with parent permission to access their SAGE assessment data. Most of these students were 11th or 12th grade students. Unfortunately, many of these students provided an ID that could not be located I the state data file; therefore, we believe they either provided their school ID or some other ID. Also many of these students have completed all required tests and did not take the SAGE assessment. Therefore, we were unable to measure the impact of this grant program. We are working with the grantees to collect employment and college enrollment data summer 2016 for students participating in the program during the period from spring 2015-spring 2016. We can provide this information in next year's evaluation report.

Fairs, Camps, and Competitions

There were 155 students with complete SAGE assessment data who were awarded a grant to support their participation in a STEM fair, camp, or competition. This resulted in a very small number of students per grade. We did not feel comfortable evaluating a program with so few students, especially when the amount of time the student is involved varies. The STEM Action Center is using a new software tracking program for this grant during the 2015-16 school year, and we hope this new approach to project management along with coordination through districts will result in a larger number of student SSIDs submitted for our analysis next year.

Conclusions

While the results from the analysis of the impact of the K-12 mathematics technology grants demonstrate that five of the educational technology products have promise for improving

student mathematics achievement, the greatest barrier to this occurring for most students is usage. Ninety percent of students in the full analytic sample were not meeting recommended usage. When reflecting on the teacher feedback, thirty-two percent of teachers surveyed reported lack of access to computers to be a significant barrier. This 2015-16 academic year, the STEM Action Center required that school principals sign a letter of commitment to ensure students have access to technology for at least 45 minutes per week to use the software they were provided in the second year of the grant program. In addition, usage has gotten off to an earlier start this year, which means that students will have a longer opportunity to use these grants. It is important to wait another year to understand the full impact of this grant program on student achievement.

This is also true for the CTE, PD, and High School STEM Industry grants. Due to the late start of implementation, which began early spring, we recommend that impact be evaluated after the 2015-16 school year to best understand the effects of these grant programs. At this time we will also evaluate the impact of the fairs, camps, and completions grants as long as there is sufficient data.

Recommendations

Technology changes at a much faster pace than traditional curriculum models (e.g., textbooks). Therefore, it is important to not take these results as support for a decision to get into a long-term contract with any vendor. Getting locked in to a multi-year contract for a software product, could preclude a district or school from the opportunity to use newer improved products from that or another vendor. What we recommend is a 3 year R&D cycle where products are selected, with district involvement, through an RFP process, products are piloted at small scale while being evaluated for the first year, and then scaled up for 2 years of implementation to

understand impact. In addition, few students will want to use the same program for multiple years, because they may get tired of the interface or other design features. Therefore, we recommend an approach that allows local decision-making and also different product selection for different grades, every few years, to maximize the benefit from education technology.

Appendix A. Baseline Equivalence Comparison

To investigate the impact of technology use across all students using a product we used the MatchIt package in R to increase the equivalence of math pretest scores for intervention and control students through propensity score matching. Prior to matching the data set was split into 12 different files based on mathematics pretest-posttest test combinations (e.g., 3rd grade pretest/4th grade posttest, 4thgrade pretest /5th grade posttest, . . ., secondary math II pretest/secondary math III posttest). A nearest-neighbor matching algorithm with 1-to-1 matching was then conducted for each test combination using the model: intervention ~ mathematics pretest scaled scores. The matched data sets were then merged.

After matching students in the grant programs to similar students in the state, we compared the characteristics of the grant program students to the comparison students to see if there were any characteristics or achievement performance that differed significantly that we should control for in our impact analysis. Once we reviewed the baseline differences across products, we decided on the covariates to include in the statistical model to determine impact of technology use on achievement. In this appendix, we provide baseline comparison tables for each product included in the impact analysis.

What is most important is that the two groups are equivalent on the pretest achievement score, since prior research has shown that prior achievement explains greater differences in achievement than any other student characteristic. We first matched students on their scale score achievement in groups by the test they took, since scale scores are not comparable across tests. Once the students in the grant program were matched to similar students in the state by test, we

then recombined all students across tests for our analysis of baseline characteristics between the two groups.

We conducted this baseline comparison for all students with any evidence of usage or participation in the grant program, which we refer to as the full sample. Then we conduct a similar baseline comparison for only students who met the recommended level of usage according to the provider, which we refer to as the fidelity sample. In each case the comparison is with the similar students in similar schools in the state, from a group created through the propensity score matching process. For some products either no fidelity benchmark was available or there were not sufficient students who met the fidelity benchmark to include in our analysis.

ALEKS Full Sample

Table 5. Baseline student characteristics after matching for ALEKS for Full Sample

Characteristic	Intervention Students (N=27,190)	Comparison Students (N=27,190)	Difference (SE)	Test statistic	p-value	Effect Size of Difference
Mean Spring 2014 Math Scale Score (Standard Deviation)	426.81 (88.757)	426.82 (88.752)	-.003 (.761)	-.005	.996	.000
Proportion of Students who met proficiency for Spring 2014 Math SAGE Assessment						
Met Proficiency (Proficiency level 3 or 4)	.389	.389	.000	.000	1.000	.000
Proportion of Students at Proficiency Levels for Spring 2014 Language Arts SAGE Assessment						
Met Proficiency (Proficiency level 3 or 4	.414	.418	-.004	-.940	.347	-.010
Student Characteristics						
Proportion Eligible for Free/Reduced Price Lunch	.388	.352	.036	8.694	.000	.094
Proportion Female [a]	.489	.496	-.007	-1.633	.103	-.017
Proportion of Students Classified as English Language Learners	.034	.036	-.002	-1.269	.204	-.036
Proportion of Students Classified as Special Education	.113	.103	.010	3.757	.000	.063
Proportion of each racial/ethnic composition						
White	.791	.756	.035	9.750	.000	.121
Hispanic	.149	.167	-.018	-5.754	.000	-.082

Note: A t-test was used to compare continuous variables and a z-test of proportions was used for all dichotomous variables. Due to the small number of students in each subgroup who were not white and not Hispanic, only White and Hispanic are used in this comparison.

Table 6. Baseline student characteristics after matching for ALEKS for Fidelity Sample

Characteristic	Intervention Students (N=633)	Comparison Students (N=633)	Difference (SE)	Test statistic	p-value	Effect Size of Difference
Mean Spring 2014 Math Scale Score (Standard Deviation)	450.83 (81.722)	450.84 (81.719)	.005 (4.594)	.001	.999	.000
Proportion of Students who met proficiency for Spring 2014 Math SAGE Assessment						
Met Proficiency (Proficiency level 3 or 4)	.409	.409	.000	.000	1.000	.000
Proportion of Students at Proficiency Levels for Spring 2014 Language Arts SAGE Assessment						
Met Proficiency (Proficiency level 3 or 4	.501	.462	.039	1.380	.168	.095
Student Characteristics						
Proportion Eligible for Free/Reduced Price Lunch	.313	.333	-.020	-.761	.447	-.055
Proportion Female [a]	.586	.520	.066	2.362	.018	.162
Proportion of Students Classified as English Language Learners	.035	.038	-.003	-.285	.779	-.052
Proportion of Students Classified as Special Education	.068	.081	-.013	-.881	.379	-.114
Proportion of each racial/ethnic composition						
White	.825	.761	.064	2.810	.005	.238
Hispanic	.114	.163	-.049	-2.524	.012	-.251

Note: A t-test was used to compare continuous variables and a z-test of proportions was used for all dichotomous variables. Due to the small number of students in each subgroup who were not white and not Hispanic, only White and Hispanic are used in this comparison.

CatchUp Math Full Sample

Table 7. Baseline student characteristics after matching for CatchUp Math for Full Sample

Characteristic	Intervention Students (N=254)	Comparison Students (N=254)	Difference (SE)	Test statistic	p-value	Effect Size of Difference
Mean Spring 2014 Math Scale Score (Standard Deviation)	432.62 (51.52)	432.62 (51.52)	.00 (4.57)	.000	1.000	.00
Percent of Students who met proficiency for Spring 2014 Math SAGE Assessment						
Met Proficiency (Proficiency level 3 or 4)	.465	.465	.000	.000	1.000	.00
Percent of Students at Proficiency Levels for Spring 2014 Language Arts SAGE Assessment						
Met Proficiency (Proficiency level 3 or 4	.673	.516	.157	3.604	.000	.40
Student Characteristics						
Percent Eligible for Free/Reduced Price Lunch	.189	.346	-.157	-3.997	.000	-.50
Percent Female [a]	.469	.500	-.031	-0.699	.484	-.08
Percent of Students Classified as English Language Learners	.016	.024	-.008	-0.644	.522	-.25
Percent of Students Classified as Special Education	.012	.063	-.051	-3.025	.002	-1.04
Percent of each racial/ethnic composition						
White	.807	.799	.008	.227	.818	.03
Hispanic	.161	.134	.027	.858	.390	.13

Note: A t-test was used to compare continuous variables and a z-test of proportions was used for all dichotomous variables. Due to the small number of students in each subgroup who were not white and not Hispanic, only White and Hispanic are used in this comparison.

There were only 33 students using CatchUp Math who met the benchmark for fidelity of implementation, who also had complete SAGE Assessment data. This was too small of a sample to conduct an impact analysis for the fidelity sample. Therefore, no baseline comparison was conducted.

iReady Full Sample

Table 8. Baseline student characteristics after matching for iReady for Full Sample

Characteristic	Intervention Students (N=3,981)	Comparison Students (N=3,981)	Difference (SE)	Test statistic	p-value	Effect Size of Difference
Mean Spring 2014 Math Scale Score (Standard Deviation)	355.69 (58.74)	355.69 (58.74)	0.00 (1.32)	0.000	1.000	.00
Proportion of Students who met proficiency for Spring 2014 Math SAGE Assessment						
Met Proficiency (Proficiency level 3 or 4)	.426	.426	.000	0.000	1.000	.00
Proportion of Students at Proficiency Levels for Spring 2014 Language Arts SAGE Assessment						
Met Proficiency (Proficiency level 3 or 4	.410	.408	.002	0.181	.857	.01
Student Characteristics						
Proportion Eligible for Free/Reduced Price Lunch	.444	.376	.068	6.168	.000	.17
Proportion Female	.480	.488	-.008	-0.714	.478	-.02
Proportion of Students Classified as English Language Learners	.038	.044	-.006	-1.350	.177	-.09
Proportion of Students Classified as Special Education	.128	.128	.000	0.000	1.000	.00
Proportion of each racial/ethnic composition						
White	.792	.756	.036	3.840	.000	.12
Hispanic	.102	.168	-.066	-8.617	.000	-.35

Note: A t-test was used to compare continuous variables and a z-test of proportions was used for all dichotomous variables. Due to the small number of students in each subgroup who were not white and not Hispanic, only White and Hispanic are used in this comparison.

Table 9. Baseline student characteristics after matching for iReady for Fidelity Sample

Characteristic	Intervention Students (N=190)	Comparison Students (N=190)	Difference (SE)	Test statistic	p-value	Effect Size of Difference
Mean Spring 2014 Math Scale Score (Standard Deviation)	372.34 (50.03)	372.34 (50.04)	.005 (5.134)	.001	.999	.00
Proportion of Students who met proficiency for Spring 2014 Math SAGE Assessment						
Met Proficiency (Proficiency level 3 or 4)	.400	.400	.000	.000	1.000	.00
Proportion of Students at Proficiency Levels for Spring 2014 Language Arts SAGE Assessment						
Met Proficiency (Proficiency level 3 or 4	.421	.401	.020	.396	.689	.05
Student Characteristics						
Proportion Eligible for Free/Reduced Price Lunch	.400	.368	.032	.641	.522	.08
Proportion Female	.489	.500	-.011	-.214	.834	-.03
Proportion of Students Classified as English Language Learners	.011	.063	-.052	-2.685	.007	-1.09
Proportion of Students Classified as Special Education	.100	.121	-.021	-.653	.516	-.13
Proportion of each racial/ethnic composition						
White	.758	.747	.011	.248	.803	.04
Hispanic	.084	.168	-.084	-2.47	.014	-.48

Note: A t-test was used to compare continuous variables and a z-test of proportions was used for all dichotomous variables. Due to the small number of students in each subgroup who were not white and not Hispanic, only White and Hispanic are used in this comparison.

Math XL Full Sample

Table 10. Baseline student characteristics after matching for Math XL for Full Sample

Characteristic	Intervention Students (N=318)	Comparison Students (N=318)	Difference (SE)	Test statistic	p-value	Effect Size of Difference
Mean Spring 2014 Math Scale Score (Standard Deviation)	560.23 (74.43)	560.22 (74.42)	.003 (5.90)	-.001	1.000	.00
Percent of Students who met proficiency for Spring 2014 Math SAGE Assessment						
Met Proficiency (Proficiency level 3 or 4)	.409	.409	.000	.000	1.000	.00
Percent of Students at Proficiency Levels for Spring 2014 Language Arts SAGE Assessment						
Met Proficiency (Proficiency level 3 or 4	.616	.522	.094	2.394	.017	.23
Student Characteristics						
Percent Eligible for Free/Reduced Price Lunch	.201	.242	-.041	-1.245	.215	-.14
Percent Female [a]	.572	.569	.003	.076	.936	.01
Percent of Students Classified as English Language Learners	.003	.006	-.003	-.565	.569	-.42
Percent of Students Classified as Special Education	.028	.025	.003	.236	.810	.07
Percent of each racial/ethnic composition						
White	.855	.827	.028	.966	.332	.13
Hispanic	.082	.119	-.037	-1.552	.121	-.25

Note: A t-test was used to compare continuous variables and a z-test of proportions was used for all dichotomous variables. Due to the small number of students in each subgroup who were not white and not Hispanic, only White and Hispanic are used in this comparison.

Pearson did not provide a flag in the data file for students who met the fidelity of implementation benchmark; therefore, we were not able to do an analysis of the impact for the fidelity sample, so no baseline comparison table is included for the fidelity sample.

ST Math Full Sample

Table 11. Baseline student characteristics after matching for ST Math for Fidelity Sample

Characteristic	Intervention Students (N=5,858)	Comparison Students (N=5,858)	Difference (SE)	Test statistic	p-value	Effect Size of Difference
Mean Spring 2014 Math Scale Score (Standard Deviation)	345.68 (56.15)	345.68 (56.15)	-.001 (1.038)	-.001	.999	.00
Proportion of Students who met proficiency for Spring 2014 Math SAGE Assessment						
Met Proficiency (Proficiency level 3 or 4)	.419	.419	.000	.000	1.000	.00
Proportion of Students at Proficiency Levels for Spring 2014 Language Arts SAGE Assessment						
Met Proficiency (Proficiency level 3 or 4	.386	.406	-.020	-2.213	.027	-.05
Student Characteristics						
Proportion Eligible for Free/Reduced Price Lunch	.534	.392	.142	15.412	.000	.35
Proportion Female	.503	.491	.012	1.299	.194	.03
Proportion of Students Classified as English Language Learners	.075	.044	.031	7.0922	.000	.34
Proportion of Students Classified as Special Education	.147	.141	.006	0.925	.358	.03
Proportion of each racial/ethnic composition						
White	.569	.747	-.178	-20.307	.000	-.49
Hispanic	.292	.177	.115	14.690	.000	.39

Note: A t-test was used to compare continuous variables and a z-test of proportions was used for all dichotomous variables. Due to the small number of students in each subgroup who were not white and not Hispanic, only White and Hispanic are used in this comparison.

ST Math Fidelity Sample

Table 12. Baseline student characteristics after matching for ST Math for Fidelity Sample

Characteristic	Intervention Students (N=801)	Comparison Students (N=801)	Difference (SE)	Test statistic	p-value	Effect Size of Difference
Mean Spring 2014 Math Scale Score (Standard Deviation)	338.41 (52.35)	338.41 (52.35)	.000	.000	1.000	.00
Proportion of Students who met proficiency for Spring 2014 Math SAGE Assessment						
Met Proficiency (Proficiency level 3 or 4)	.429	.429	.000	.000	1.000	.00
Proportion of Students at Proficiency Levels for Spring 2014 Language Arts SAGE Assessment						
Met Proficiency (Proficiency level 3 or 4	.421	.411	.010	.406	.682	.02
Student Characteristics						
Proportion Eligible for Free/Reduced Price Lunch	.577	.391	.186	7.449	.000	.46
Proportion Female	.481	.501	-.020	-.801	.424	-.05
Proportion of Students Classified as English Language Learners	.096	.041	.055	4.357	.000	.55
Proportion of Students Classified as Special Education	.180	.162	.018	.957	.337	.08
Proportion of each racial/ethnic composition						
White	.527	.755	-.228	-9.512	.000	-.62
Hispanic	.325	.170	.155	7.188	.000	.52

Note: A t-test was used to compare continuous variables and a z-test of proportions was used for all dichotomous variables. Due to the small number of students in each subgroup who were not white and not Hispanic, only White and Hispanic are used in this comparison.

Think Through Math Full Sample

Table 13. Baseline student characteristics after matching for Think Through Math for Full Sample

Characteristic	Intervention Students (N=6896)	Comparison Students (N=6896)	Difference (SE)	Test statistic	p-value	Effect Size of Difference
Mean Spring 2014 Math Scale Score (Standard Deviation)	384.37 (63.556)	384.36 (63.541)	.011	.010	.992	.000
Proportion of Students who met proficiency for Spring 2014 Math SAGE Assessment						
Met Proficiency (Proficiency level 3 or 4)	.454	.454	.000	.000	1.000	.000
Proportion of Students at Proficiency Levels for Spring 2014 Language Arts SAGE Assessment						
Met Proficiency (Proficiency level 3 or 4	.458	.457	.001	.118	.904	.002
Student Characteristics						
Proportion Eligible for Free/Reduced Price Lunch	.320	.360	-.040	-4.958	.000	-.108
Proportion Female [a]	.478	.489	-.011	-1.293	.197	-.027
Proportion of Students Classified as English Language Learners	.038	.034	.004	1.261	.208	.070
Proportion of Students Classified as Special Education	.136	.118	.018	3.174	.002	.099
Proportion of each racial/ethnic composition						
White	.782	.773	.009	1.271	.204	.032
Hispanic	.139	.156	-.017	-2.815	.005	-.082

Note: A t-test was used to compare continuous variables and a z-test of proportions was used for all dichotomous variables. Due to the small number of students in each subgroup who were not white and not Hispanic, only White and Hispanic are used in this comparison.

Think Through Math Fidelity Sample

Table 14. Baseline student characteristics after matching for Think Through Math for Fidelity Sample

Characteristic	Intervention Students (N=2814)	Comparison Students (N=2814)	Difference (SE)	Test statistic	p-value	Effect Size of Difference
Mean Spring 2014 Math Scale Score (Standard Deviation)	395.44 (56.799)	395.43 (56.766)	.015 (1.514)	.010	.992	.000
Proportion of Students who met proficiency for Spring 2014 Math SAGE Assessment						
Met Proficiency (Proficiency level 3 or 4)	.598	.598	.000	.000	1.000	.000
Proportion of Students at Proficiency Levels for Spring 2014 Language Arts SAGE Assessment						
Met Proficiency (Proficiency level 3 or 4	.572	.562	.010	.757	.447	.025
Student Characteristics						
Proportion Eligible for Free/Reduced Price Lunch	.251	.307	-.056	-4.683	.000	-.169
Proportion Female [a]	.478	.495	-.017	-1.276	.201	-.041
Proportion of Students Classified as English Language Learners	.018	.019	-.001	-.278	.779	-.033
Proportion of Students Classified as Special Education	.073	.071	.002	.290	.772	.018
Proportion of each racial/ethnic composition						
White	.826	.799	.027	2.595	.010	.108
Hispanic	.099	.133	-.034	-3.983	.000	-.202

Note: A t-test was used to compare continuous variables and a z-test of proportions was used for all dichotomous variables. Due to the small number of students in each subgroup who were not white and not Hispanic, only White and Hispanic are used in this comparison.

Appendix B. Logistic Regression

After creating the matched comparison groups and conducting the baseline equivalence analysis, a follow-up logistic regression was then conducted using proficiency on the math posttest as a binary outcome variable. The predictors in the model were the following student-level variables: intervention, eligibility for free or reduced price lunch, eligibility for special education services, English Language Learner status, gender (female), proficiency level on the mathematics pretest, and proficiency level on the language pretest. Proficiency levels were treated as categorical variables, where 1 was the lowest proficiency and 4 was the highest proficiency level. Standard errors were corrected for clustering of students within schools using bootstrapping with school ID as a stratum. The outcome, an odds ratio provides information on whether any impact is in favor of the students using the technology or the students in the comparison group. The group favored is more likely to have met proficiency on the SAGE assessment. We also include an effect size calculation for this odds ratio, which can be used to determine if the difference, or odds of meeting proficiency, is at a level that is educationally meaningful.