SRI International

Evaluation of Rocketship Education's Use of DreamBox Learning's Online Mathematics Program

Haiwen Wang
Katrina Woodworth

Center for Education Policy
SRI International
August 2011

333 Ravenswood Avenue • Menlo Park, California 94025-3493 • 650.859.2000 • www.sri.com

Introduction

Rocketship Education is a charter management organization at the forefront of the small but growing movement to expand the use of blended and hybrid learning in K–12 schools. Distinct from distance learning, blended and hybrid systems have a combination of online and offline learning in which students engage in adult-supervised online instruction for a part of their school day (Horn & Staker, 2011; U.S. Department of Education, Office of Educational Technology, 2010). Rocketship seeks to transform public education by developing an instructional model that supplements traditional face-to-face instruction with instruction provided via computer-based programs and tutoring. At Rocketship schools, the online instruction happens in the Learning Labs and focuses on developing students' reading and mathematics skills.

Enthusiasm for blended and hybrid learning stems from its potential to increase personalization and boost productivity. As noted in the U.S. Department of Education's National Educational Technology Plan (2010, p. 4), "Contemporary technology offers unprecedented performance, adaptability, and cost-effectiveness." While blended and hybrid learning systems are still in the early stages of development, significant growth is expected over the next decade (Horn & Staker, 2011). For its part, Rocketship has ambitious expansion plans. Opening its first school in 2007, Rocketship was operating three schools in San Jose, California, in 2010–11. By 2030, it intends "to expand into 50 different cities across the U.S., bringing the unique Rocketship Hybrid Model to millions of students" (Rocketship Education, 2011).

To help inform the ongoing development of Rocketship's hybrid model, SRI International researchers conducted an independent evaluation of the impact of supplemental online instruction on student learning. We applied a randomized controlled trial (RCT) to examine the short-term effects of online mathematics curricula on elementary school students. This report focuses on the DreamBox program, as currently implemented in Rocketship's Learning Lab with kindergarten and first-grade students.

The primary research questions were as follows:

1. What impact does supplemental online mathematics instruction (DreamBox Learning) have on students' mathematics learning by the end of one semester?
2. Do effects differ for students with different characteristics (i.e., English learner status, grade level, pretest scores, participation in Response to Intervention [RtI])?

We begin with a summary of the research literature on the effects of online instruction in K–12 schools, then describe our methods, and finally present our findings. We conclude with a discussion of the implications of this research.

Literature Review

Although online learning is becoming increasingly popular in U.S schools, few rigorous studies have been conducted on the effect of online learning programs, including blended learning systems, on student outcomes in K–12 education. In a meta-analysis of research on online learning, Means et al. (2009) found only five experimental or quasi-experimental studies that compared online and blended programs with face-to-face instruction and met the criteria for inclusion in the meta-

analysis (all five compared blended learning with face-to-face instruction).[1] Four of the five studies found positive effects of blended programs on student achievement on researcher-developed assessments in algebra, history, and science (Long & Jennings, 2005; O'Dwyer, Carey, & Kleiman, 2007; Sun, Lin, & Yu, 2008). These findings, however, should be interpreted with caution because researcher-developed assessments tend to overalign with the interventions of interest and therefore may overestimate their effects. The review for the meta-analysis did not uncover any studies of online learning programs in K–12 education that relied on standardized external outcome measures.

Experimental studies of other computer-based programs that were not delivered online (i.e., not web based) but were designed to support instruction using technology failed to detect positive effects on standardized tests. Rouse and Krueger (2004) found a small positive effect for the Fast ForWord reading program on a computer-based measure of language skills but no effect on reading achievement on the Clinical Evaluation of Language Fundamentals (CELF-3) or on state standardized reading assessments. Similarly, Borman, Benson, and Overman (2009) found that Fast ForWord did not have an effect on eighth-grade students' language and reading comprehension on the Comprehensive Test of Basic Skills (CTBS/5). Likewise, Dynarski et al. (2007) and Campuzano et al. (2009) evaluated the effects of multiple reading and mathematics software programs and did not find significant effects of these programs on Stanford Achievement Test (SAT-9 and SAT-10) scores.

None of these studies included kindergarten or first-grade students. This highlights the lack of knowledge about the effect of technology-supported learning in the early grades—the focus of this study. There have been no prior experimental or quasi-experimental studies on the effects of DreamBox Learning.

Research Design

We conducted an RCT involving all students in kindergarten and first grade in each of the three Rocketship schools in operation in 2010–11. Students were randomly assigned to one of two conditions: (1) online mathematics instruction supplementing face-to-face mathematics instruction (treatment) or (2) face-to-face mathematics instruction only (control). We randomly assigned individual students, separately within and by grade level (K and 1), at a 4 to 1 ratio to the treatment and control groups.

The experiment spanned 4 months (mid-October through mid-February), including 70 days of instruction. Students in treatment and control groups were scheduled to receive 100 to 110 minutes per day of face-to-face mathematics instruction in their classrooms. Students in the treatment group were scheduled to receive an additional 20 to 40 minutes per day of online mathematics instruction, with most sessions lasting 40 minutes, while the control students from the same class received online literacy instruction. In all three schools, some low-achieving students, regardless of their treatment assignment, participated in an RtI program in which they were scheduled to receive literacy tutoring as well as about 45 minutes of DreamBox each day. (See Exhibit 1 for an overview of a typical daily schedule for a Rocketship student.)

[1] The criteria included applying an experimental or quasi-experimental study and providing sufficient information to support computation of an effect size.

Exhibit 1
Sample Daily Schedule for a Second-Grade Student, Fall 2010

Time	Activity
7:30 AM	Breakfast
8:00 AM	Literacy, science, and social studies
11:20 AM	Lunch/recess
12:00 PM	Mathematics
1:40 PM	Learning Lab (online instruction)
3:20 PM	PE/outside play
4:00 PM	Dismissal or afterschool program for students in RtI (online instruction and small group tutoring)
6:00 PM	Dismissal for students in RtI

With this design, the evaluation essentially estimated the effect of supplemental online mathematics instruction versus the online literacy program on students' mathematics outcomes. A result of this design is that the estimated DreamBox effect is confounded with the effect of receiving additional mathematics instruction. In other words, because we are not comparing DreamBox instruction with another form of mathematics instruction, we cannot isolate the effect of DreamBox from the effect of additional instructional time.

Rocketship administered the Northwest Evaluation Association's (NWEA) mathematics tests in September 2010 (pretest) and January/February 2011 (posttest) to students included in the experiment. In the primary grades, NWEA's Measures of Academic Progress (MAP) assessment in mathematics is aligned with national mathematics standards (e.g., those developed by the National Council of Teachers of Mathematics). Our analysis included both the general NWEA mathematics scores and subtest scores for problem solving, number sense, computation, measurement and geometry, and statistics and probability. All the scores are in the RIT scale,[2] which is scaled using the Item Response Theory (IRT) and has the same meaning regardless of the grade of the student.

The Intervention

Here, we describe the DreamBox Learning program and its alignment with the NWEA assessment and provide information about its implementation at Rocketship schools.

DreamBox Learning provides an adaptive learning environment that tailors instruction to student needs and provides feedback to teachers to facilitate student learning. DreamBox generates information on program use (e.g., notifications of students who are struggling with a concept or unit or working inefficiently in the program) and student progress (proficiency and growth), but does not prescribe a specific role for teachers. DreamBox Learning recommends students spend a minimum of 90 minutes per week on the program.

The DreamBox Learning curriculum is based on the National Council of Teachers of Mathematics standards and has been aligned with Common Core State Standards. It focuses on learning numbers

[2] The RIT Scale is a curriculum scale that uses individual item difficulty values to estimate student achievement. For more information, see http://www.nwea.org/support/article/532/rit-scale

and operations, place value, and number sense. The number-related activities often make use of the open number line, thereby touching upon measurement and geometry. Exhibit 2 lists the NWEA subtest strands and indicates where DreamBox instruction is aligned with them. Because at Rocketship so much instruction is provided face to face with teachers, the alignment between the face-to-face instruction provided over the course of the experiment and the NWEA subtests is also indicated.

Exhibit 2
Alignment of DreamBox and Face-to-Face Instruction with NWEA Subtest Strands

NWEA Subtests	DreamBox Instruction		Face-to-Face Instruction	
	Kindergarten	First Grade	Kindergarten	First Grade
Problem solving			Partial	Partial
Number sense	✓	✓	✓	✓
Computation	✓	✓	✓	✓
Measurement and geometry	Partial	Partial	✓	✓
Statistics and probability			✓	✓

Over the course of this experiment, treatment students (kindergarteners and first-graders) accessed DreamBox in the Rocketship schools' Learning Labs, and control students from the same homeroom accessed an online literacy program in the same lab. The labs are run by lab coordinators, who are noncredentialed hourly staff and play a minimal role in instruction. Finally, while the DreamBox Learning program does generate information for teachers, it was not used by Rocketship's classroom teachers to modify instruction for students in either the treatment or control group.

Data Collection

Rocketship provided student demographic information, pre- and posttest scores on the NWEA mathematics test, and program usage data, including the actual hours students spent on the program during the experiment. In addition, we collected school calendars and computer lab schedules for each school, which we used to calculate scheduled participation time.

The Sample: Student Characteristics and Achievement

A total of 583 students were in the study sample—all students in grades K–1 in the three schools. Among students included in the experiment, 87% were Hispanic students, 81% were English learners, 88% were eligible for the FRPM program, and 4% had been identified for special education (Exhibit 3). Of these students, 10% participated in RtI during the experiment. The treatment and control groups were balanced in terms of these background characteristics; almost all differences were less than 5% and none were statistically significant at a .05 significance level.

Exhibit 3
Student Characteristics by Treatment and Control Condition

	Overall	Treatment	Control
N	583	466	117
Female (%)	53.3	52.4	57.3
Hispanic (%)	87.3	86.7	89.7
English learner (%)	80.6	82.4	73.5
FRPM (%)	87.7	87.8	87.2
Special education (%)	4.1	4.7	1.7
RtI participation (%)	9.6	9.7	9.4

Exhibit 4 presents the means and standard deviations of the pre- and posttest scores (NWEA mathematics test scores in September 2010 and in January/February 2011) for the treatment and control students. The differences in pretest scores were in general less than 3 points, all within .2 standard deviations of the scores for the entire sample, and none of the differences were statistically significant at a .05 significance level, meeting the What Works Clearinghouse (WWC) standards for a balanced sample.

Exhibit 4
Pre and Post NWEA Math Test Scores by Treatment and Control Condition

	Treatment					Control				
	Pretest			Posttest		Pretest			Posttest	
	N	Mean	*SD*	Mean	*SD*	*N*	Mean	*SD*	Mean	*SD*
Math overall	446	146.0	18.0	159.0	16.6	111	144.7	15.0	156.2	15.1
Problem solving	444	147.0	19.3	161.4	16.3	109	144.7	17.1	159.8	15.2
Number sense	444	146.9	20.0	159.6	18.9	109	143.4	16.6	157.0	17.2
Computation	438	147.5	22.4	163.0	20.7	108	147.0	19.8	158.8	19.5
Measurement and geometry	441	144.5	18.9	155.5	18.3	109	144.8	18.4	151.8	18.1
Statistics and probability	443	145.5	19.3	156.3	18.9	109	145.1	15.6	154.1	17.6

Fewer students are reported in Exhibit 4 than Exhibit 3 because it includes only those students for whom we had both pre- and posttest data. As discussed below, 26 students were excluded from the impact analysis because of missing pretest and/or posttest scores.

Data Analysis

To understand the DreamBox usage patterns among treatment and control students, we conducted initial descriptive analyses. We then identified the student characteristics associated with greater usage time using ordinary least squares (OLS) regression to predict usage hours for students assigned to the treatment group.

We conducted two types of analysis to examine the effects of DreamBox. One was an intent-to-treat (ITT) analysis in which we studied the effect of being assigned to the treatment group regardless of

each student's actual time spent on DreamBox. We estimated the ITT effect on posttest achievement adjusting for students' demographic background, pretest scores, RtI status, grade level, and school fixed effects.[3] We also examined the interaction between treatment and pretest score, gender, eligibility for the FRPM program, RtI status, grade level, and school fixed effects to examine whether DreamBox has differential effects on student subgroups.

The ITT analysis offers an unbiased estimate of the effect between the treatment and control groups, but it may underestimate the effect of the treatment because some control students received the treatment while some treatment students did not. Therefore, we also conducted a treatment-on-the-treated (TOT) analysis to study the effect of usage hours on student outcomes. The most straightforward approach to the TOT analysis is to use the usage hours to predict the outcomes and therefore estimate the effect of usage hours on these outcomes. However, because students who spent more time accessing DreamBox may be more motivated to learn mathematics than those who had fewer usage hours, their outcomes might have improved more even if they had not used the programs more (because they may also learn more through other sources). Therefore, the estimated effect of actual usage hours on student achievement may be biased since it may be confounded with the effect of unmeasured motivation factors.

To address this selection bias issue, we used an instrumental variable (IV) approach, where we applied a two-stage least squares regression, using treatment assignment as the instrument to model the actual hours a student participated in the program and then estimating the effect of the predicted program hours from this model on the outcomes. The effect of predicted participation hours, unlike actual hours students spent on the program, is not subject to selection bias; thus, we could obtain an unbiased estimate of the effect of participation.

Summary of Findings

To summarize the findings, we first present information about students' DreamBox usage and factors related to usage. We then turn to the ITT and TOT results for the effects of using DreamBox on student performance on NWEA mathematics test scores.

Program Usage

The usage data revealed considerable treatment crossover (control students using DreamBox) and significant variation in dosage among treatment students. On average, students in the treatment group logged 21 hours on DreamBox over the 4-month experiment (Exhibit 5); with approximately 16 instructional weeks, this translates to just under 80 minutes a week.

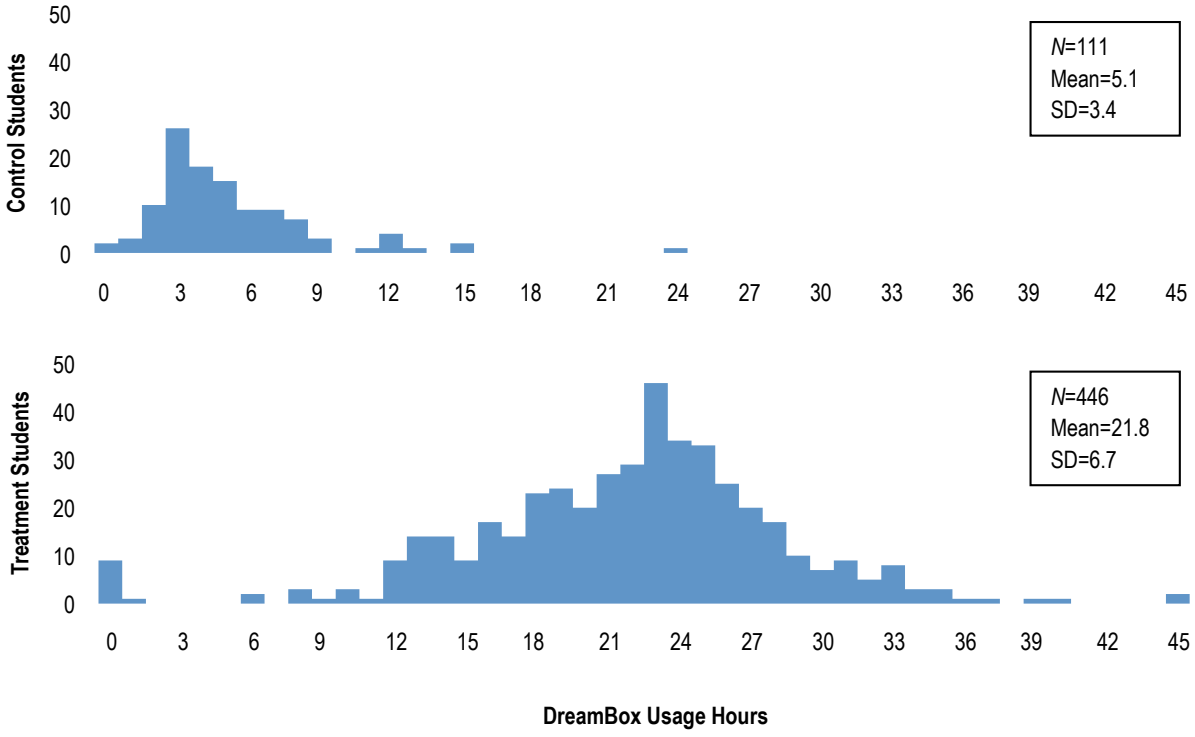
[3] We also posited a hierarchical model with classroom and student levels, with treatment condition at the student level. The results are very similar to those from the OLS regression and are not presented in this summary.

Exhibit 5
Participation Hours by Treatment Condition and RtI Status

		N	Mean	*SD*
Treatment				
	Overall	446	21.8	6.7
	Non-RtI	404	21.3	6.3
	RtI	42	26.5	8.2
Control				
	Overall	111	5.1	3.4
	Non-RtI	100	4.8	3.2
	RtI	11	8.0	4.1

Some control students obtained a significant number of DreamBox usage hours, and only two control group students had zero hours of usage (Exhibit 6). The primary reason for this is that the control students who participated in RtI had access to DreamBox. The effect of RtI is evident in the higher average usage hours for RtI students in both the treatment and control groups (Exhibit 5). The participation of non-RtI control students in DreamBox, however, was not anticipated. Students may have accessed the treatment by logging in to the program while the lab coordinator was out of the Learning Lab (the substitute supporting staff may not have been as vigilant as the coordinators in ensuring students logged in to the correct program). In one school, a lab coordinator was on leave for about a month during the experiment, and the average DreamBox usage hours for control students in that school was higher than in the other schools.

Exhibit 6
Distribution of Usage Hours by Treatment Condition



The variation in usage hours within the treatment group was also substantial. Again, some of the variation can be attributed to some students accessing DreamBox also through RtI. Another reason is that students receiving special education services were pulled out from the Learning Lab. While we do not know all the reasons for the variation in usage or the discrepancy between scheduled and actual hours, it is not unexpected: These are young children traveling to a computer lab, and time is undoubtedly spent on transitioning to a different learning environment and logging in for instruction. Yet of some concern is the fact that a small number of treatment students (9 of 446) had zero usage hours. This could have been some kind of mistake on the lab coordinator's part, or it may be that some students just did not log in to the program in the Learning Lab.

Predictors of Program Usage

In light of the substantial variation in usage hours for treatment students, we posited a regression model to predict usage hours using student characteristics, achievement on the pretest, and scheduled online program hours. We found that among students assigned to the treatment group, being assigned to RtI, being eligible for FRPM, being a first-grader versus a kindergartener, and being in one school versus another were related to more usage hours. This indicates variations in usage hours among student subgroups and between schools. (For detailed regression results, please see the appendix.)

Results for ITT and TOT Analysis

A total of 583 students were in the DreamBox experiment. As mentioned, 26 were excluded from the impact analysis because of missing pretest and/or posttest scores. We compared demographic, background, and achievement information between students who were included in the analysis and students who were not. Students in the two groups were similar on all characteristics except that those who were in the analytic sample were more likely to be eligible for FRPM. We did not look at differential attrition between treatment and control groups because very few control students were excluded from the analytic sample. For both the ITT and TOT analyses, we used Cook's D distance statistic to identify outliers for each outcome. Depending on the outcome, two to five students were excluded from the models. We present the results without the outliers in the models.

We found statistically significant ITT effects of DreamBox on NWEA's overall mathematics test score as well as on the measurement and geometry subtest but not on the problem solving, number sense, computation, and statistics and probability subtests (Exhibit 7). (For detailed regression results, please see the appendix.)

Exhibit 7
Summary of Regression Results for the ITT Effects on NWEA Mathematics Scores

	Math Overall	Problem Solving	Number Sense	Computation	Measurement and Geometry	Statistics and Probability
Effect on RIT scale score	2.30**	1.02	1.53	2.68	2.91*	2.20
SE.	(0.83)	(1.11)	(1.23)	(1.41)	(1.23)	(1.36)
Effect size	0.14	0.06	0.08	0.13	0.16	0.12

*$p < .05$

Students in the DreamBox treatment group scored an average of 2.3 points higher on the NWEA overall mathematics test than similar students in the control group; this difference translates into an effect size of .14. This difference also translates to an improvement index of 5.5 percentile points, which suggests that being assigned to the treatment group would have led to a 5.5 point increase in the percentile rank for the average (50th percentile) student in the control group. Students in the treatment group scored an average of 2.9 points higher on the measurement and geometry subtest than their peers in the control group; this difference translates into an effect size of .16. This difference also translates to an improvement index of 6.4 percentile points, suggesting that being assigned to the treatment group would have led to a 6.4 point increase in the percentile rank for the average (50th percentile) student in the control group. Although we found no statistically significant effects on the problem solving, number sense, computation, or statistics and probability subtests, the effects all have a positive sign, suggesting that DreamBox improved student math achievement in a comprehensive way. We did not find statistically significant differential effects for student subgroups (see the appendix for the model interacting DreamBox effects with student characteristics).

Consistent with the ITT findings, we found significant TOT effects of DreamBox usage hours on the NWEA overall mathematics test score as well as on the measurement and geometry subtest. (For detailed regression results, please see the appendix.) The results are robust with the different methods used (ordinary least squares regression; hierarchical modeling with student and classroom levels, with fixed or random treatment effects).

Discussion

Given the expected growth of blended and hybrid learning systems, rigorous research on both the efficacy and effectiveness of technology-based instruction is essential. To date, the research has been limited, especially when it comes to the use of technology with our youngest students.

This study's positive findings about the effects of DreamBox instruction are likely to fuel the sense of optimism about the promise of online learning, especially in light of the relatively modest treatment. In interpreting these findings, we urge educators and policymakers to keep in mind a basic principle of scientific research—that research findings contribute to the ongoing refinement of hypotheses but do not represent a conclusion. Positive results merit continued and even expanded use, but ongoing evaluation is needed to build a body of evidence, especially as interventions are implemented in varied ways in diverse settings.

Moreover, this study examined the effects of using the DreamBox program for only a short period of time. Using the program for a longer time may have different effects. Further, because we only examined the short-term effects of the program, we do not know how long the estimated positive DreamBox effect would persist. Follow-up of the experiment would be needed to address these questions.

References

Borman, G. D., Benson, J., & Overman, L. (2009). A randomized field trial of the Fast ForWord Language computer-based training program. *Educational Evaluation and Policy Analysis*, *31*: 82–106.

Campuzano, L., Dynarski, M., Agodini, R., & Rall, K. (2009). *Effectiveness of reading and mathematics software products: Findings from two student cohorts* (NCEE 2009–4041). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Dynarski, M., Agodini, R., Heaviside, S., Novak, T., Carey, N., Campuzano, L., Means, B., Murphy, R., Penuel, W., Javitz, H., Emery, D., & Sussex, W. (2007). *Effectiveness of reading and mathematics software products: Findings from the first student cohort.* Washington, DC: U.S. Department of Education.

Horn, M. B., & Staker, H. (2011). *The rise of K-12 blended learning.* Mountain View, CA: Innosight Institute.

Long, M., & Jennings, H. (2005). *"Does it work?": The impact of technology and professional development on student achievement.* Calverton, MD.: Macro International.

Means, B., Toyama, Y., Murphy, R., Bakia, M., & Jones, K. (2009). *Evaluation of evidence-based practices in online learning: A meta-analysis and review of online-learning studies.* Washington, DC: U.S. Department of Education.

O'Dwyer, L. M., Carey, R., & Kleiman, K. (2007). A study of the effectiveness of the Louisiana Algebra I online course. *Journal of Research on Technology in Education*, *39*(3): 289–306.

Rocketship Education. (2011). Mission statement. Retrieved from http://rsed.org/index.php?page=mission

Rockman et al. (2007). *ED PACE final report*. Submitted to the West Virginia Department of Education. San Francisco, CA: Author. Retrieved from www.rockman.com/projects/146.ies.edpace/finalreport

Rouse, C. E. , & Krueger, A. B. (2004). Putting computerized instruction to the test: A randomized evaluation of a "scientifically based" reading program. *Economics of Education Review*, *23*(4): 323–338.

Sun, K., Lin, Y., & Yu, C (2008). A study on learning effect among different learning styles in a Web-based lab of science for elementary school students. *Computers & Education*, *50* (4):1411–22.

U.S. Department of Education, Office of Educational Technology (2010). *Transforming American education: Learning powered by technology. National education technology plan 2010.* Washington, DC: Author.

Appendix

Model Estimation for Usage Hours Prediction and ITT and TOT Effects

Exhibit A-1

Regression Results for Factors Related to Usage Hours for Students in the Treatment Condition

	Hours on DreamBox (*N* = 456)	
Intercept	16.69	
	(4.51)	
Pretest math overall	-0.02	
	(0.03)	
Female	0.43	
	(0.66)	
Hispanic	-0.78	
	(1.03)	
English learner	-0.51	
	(0.89)	
FRPM	3.19	**
	(1.02)	
Special education	-2.46	
	(1.52)	
RtI	2.45	*
	(1.18)	
Grade 1	3.30	**
	(1.01)	
School A	0.23	
	(0.85)	
School B	2.71	***
	(0.81)	
Scheduled lab hours	0.08	
	(0.06)	
R^2	0.13	

*p < .05, **p < .01, ***p < .001

Exhibit A-2
Regression Results for the ITT Effects of DreamBox on NWEA Math Scores

	Math Overall (*N* = 552)		Problem Solving (*N* = 549)		Number Sense (*N* = 549)		Computation (*N* = 543)		Measurement and Geometry (*N* = 546)		Statistics and Probability (*N* = 550)	
Intercept	59.07		81.98		64.78		72.81		48.60		49.94	
	(4.26)		(5.63)		(6.25)		(7.25)		(6.24)		(6.92)	
Pretest math overall	0.68	***	0.54	***	0.73	***	0.61	***	0.71	***	0.60	***
	(0.03)		(0.06)		(0.07)		(0.07)		(0.07)		(0.07)	
Pretest to outcome			0.01		-0.06		-0.04		0.04		0.11	
			(0.04)		(0.05)		(0.06)		(0.06)		(0.06)	
Treatment	**2.30**	**	**1.02**		**1.53**		**2.68**		**2.91**	*	**2.20**	
	(0.83)		**(1.11)**		**(1.23)**		**(1.41)**		**(1.23)**		**(1.36)**	
Female	0.34		1.29		-0.44		0.64		0.54		-0.09	
	(0.67)		(0.89)		(0.98)		(1.14)		(0.99)		(1.09)	
Hispanic	-2.58	*	-3.83	**	-4.22	**	-5.59	**	-2.52		1.79	
	(1.06)		(1.41)		(1.55)		(1.80)		(1.57)		(1.73)	
English learner	0.59		-2.22		1.56		2.43		0.13		0.10	
	(0.89)		(1.18)		(1.30)		(1.50)		(1.31)		(1.43)	
FRPM	-2.52	*	0.15		-2.91		-3.52		-3.38	*	-1.91	
	(1.14)		(1.51)		(1.65)		(1.92)		(1.67)		(1.83)	
Special education	-2.24		-0.82		-1.52		-2.73		-2.24		-3.05	
	(1.64)		(2.17)		(2.44)		(2.82)		(2.40)		(2.66)	
RtI	-4.51	***	-2.82		-6.33	****	-3.00		-5.20	**	-5.42	**
	(1.23)		(1.62)		(1.79)		(2.08)		(1.79)		(1.99)	
Grade 1	6.12	***	6.50	***	7.71	***	14.16	***	1.35		3.14	
	(1.03)		(1.36)		(1.52)		(1.84)		(1.53)		(1.68)	
School A	-0.16		-0.70		-2.21		3.72	*	0.00		-1.17	
	(0.87)		(1.15)		(1.27)		(1.48)		(1.28)		(1.41)	
School B	-0.73		-0.78		-3.14	**	4.74	***	-2.33	*	-0.78	
	(0.80)		(1.06)		(1.17)		(1.36)		(1.18)		(1.29)	
R^2	0.78		0.60		0.63		0.60		0.61		0.55	

*p < .05, **p < .01, ***p < .001

Exhibit A-3

Regression Results for the ITT Effects of DreamBox on NWEA Math Scores with Subgroup Interactions

	Math Overall (*N* = 552)		Problem Solving (*N* = 551)		Number Sense (*N* = 550)		Computation (*N* = 543)		Measurement and Geometry (*N* = 549)		Statistics and Probability (*N* = 549)	
Intercept	57.40	***	78.65	***	80.34	***	57.70	**	67.75	***	57.24	***
	(10.77)		(14.32)		(15.68)		(18.08)		(15.92)		(17.10)	
Pretest math overall	0.70	***	0.56	***	0.62	***	0.71	***	0.56	***	0.58	***
	(0.07)		(0.10)		(0.13)		(0.14)		(0.13)		(0.13)	
Pretest to outcome			0.00		-0.06		-0.04		0.07		0.10	
			(0.05)		(0.05)		(0.06)		(0.06)		(0.06)	
Treatment	3.82		5.13		-15.50		20.09		-19.97		-7.67	
	(11.60)		(15.47)		(16.91)		(19.53)		(17.22)		(18.47)	
Female	-0.28		-1.21		-1.74		-1.21		0.48		-1.78	
	(1.52)		(2.03)		(2.23)		(2.54)		(2.28)		(2.45)	
Hispanic	-2.60	*	-3.74	**	-3.94	**	-5.44	**	-2.99		1.97	
	(1.07)		(1.45)		(1.58)		(1.80)		(1.62)		(1.74)	
English learner	-1.88		-3.41		-2.87		-2.95		1.25		-1.03	
	(1.95)		(2.59)		(2.83)		(3.28)		(2.93)		(3.14)	
FRPM	-2.42		1.87		-0.71		2.22		-7.51		-4.64	
	(2.61)		(3.49)		(3.81)		(4.35)		(3.91)		(4.21)	
Special education	-2.30		-1.03		-1.65		-2.90		-2.25		-3.29	
	(1.64)		(2.21)		(2.46)		(2.80)		(2.47)		(2.66)	
RtI	-0.94		-0.07		-2.35		1.69		-3.33		-7.64	
	(2.81)		(3.77)		(4.15)		(4.72)		(4.22)		(4.52)	
Grade 1	6.50	**	8.33	**	10.90	**	14.80	***	4.81		4.00	
	(2.37)		(3.14)		(3.50)		(3.93)		(3.51)		(3.75)	
School A	2.01		0.98		0.88		4.60		1.36		1.86	
	(1.96)		(2.63)		(2.87)		(3.30)		(2.96)		(3.17)	
School B	0.23		-0.03		-2.48		5.73		-5.46	*	1.58	
	(1.82)		(2.46)		(2.72)		(3.06)		(2.77)		(2.95)	
Treatment*Pretest math overall	-0.02		-0.01		0.12		-0.11		0.13		0.05	
	(0.08)		(0.11)		(0.12)		(0.14)		(0.12)		(0.13)	
Treatment*Female	0.77		3.43		1.39		2.15		-0.45		2.31	
	(1.69)		(2.27)		(2.49)		(2.83)		(2.54)		(2.74)	
Treatment*Eng lrnr	3.19		0.87		5.16		6.89		(0.56)		1.46	
	(2.19)		(2.92)		(3.18)		(3.67)		(3.29)		(3.53)	
Treatment*FRPM	-0.09		-1.99		-2.57		-7.03		5.75		3.51	
	(2.89)		(3.89)		(4.23)		(4.83)		(4.34)		(4.66)	
Treatment*RtI	-4.51		-3.24		-5.28		-5.96		-2.76		2.85	
	(3.12)		(4.19)		(4.61)		(5.25)		(4.69)		(5.02)	
Treatment*Grade 1	-0.57		-2.48		-3.83		-1.20		-3.25		-1.43	
	(2.61)		(3.50)		(3.85)		(4.31)		(3.86)		(4.16)	
Treatment*Sch A	-2.71		-2.45		-3.67		-1.09		-1.28		-4.09	
	(2.19)		(2.94)		(3.21)		(3.69)		(3.29)		(3.54)	
Treatment*Sch B	-1.09		-1.36		-0.76		-0.64		4.01		-3.17	
	(2.03)		(2.74)		(3.03)		(3.42)		(3.07)		(3.30)	
R^2	0.781		0.594		0.633		0.612		0.606		0.558	

*p < .05, **p < .01, ***p < .001

Exhibit A-4

Results for Two-Stage Least Squares Regression of the TOT Effects of DreamBox Participation on NWEA Math Scores

	Math Overall (*N* = 552)		Problem Solving (*N* =549)		Number Sense (*N* =549)		Computation (*N* = 543)		Measurement and Geometry (*N* =546)		Statistics and Probability (*N* = 550)	
Intercept	57.96		81.49		64.11		71.59		47.22		49.06	
	(4.27)		(5.69)		(6.34)		(7.29)		(6.30)		(6.99)	
DreamBox participation hours	**0.14**	**	**0.06**		**0.09**		**0.16**		**0.17**	*	**0.13**	
	(0.05)		**(0.07)**		**(0.07)**		**(0.08)**		**(0.07)**		**(0.08)**	
Pretest math overall	0.69	***	0.55	***	0.73	***	0.61		0.72	***	0.62	***
	(0.05)		(0.06)		(0.07)		(0.07)		(0.07)		(0.07)	
Pretest of outcome			0.01		-0.06		-0.04		0.03		0.10	
			(0.04)		(0.05)		(0.06)		(0.06)		(0.06)	
Female	0.33		1.29		-0.43		0.65		0.55		-0.09	
	(0.66)		(0.89)		(0.98)		(1.13)		(0.98)		(1.08)	
Hispanic	-2.62	*	-3.85	**	-4.25	**	-5.64	**	-2.59		1.73	
	(1.05)		(1.40)		(1.55)		(1.77)		(1.56)		(1.72)	
English learner	0.62		-2.20		1.58		2.46		0.18		0.12	
	(0.87)		(1.17)		(1.29)		(1.49)		(1.30)		(1.43)	
FRPM	-2.47	*	0.16		-2.87		-3.44		-3.26	*	-1.83	
	(1.13)		(1.51)		(1.65)		(1.91)		(1.66)		(1.83)	
Special education	-1.91		-0.67		-1.29		-2.43		-1.82		-2.76	
	(1.61)		(2.15)		(2.43)		(2.78)		(2.38)		(2.65)	
RtI	-4.95	***	-3.02		-6.62	***	-3.55		-5.75	***	-5.85	**
	(1.22)		(1.62)		(1.80)		(2.08)		(1.80)		(1.99)	
Grade 1	5.62	***	6.29	***	7.40	***	13.56	***	0.74		2.68	
	(1.03)		(1.38)		(1.54)		(1.85)		(1.54)		(1.69)	
School A	-0.23		-0.74		-2.26		3.66	*	-0.08		-1.26	
	(0.85)		(1.15)		(1.27)		(1.47)		(1.27)		(1.41)	
School B	-1.12		-0.96		-3.40	**	4.27	**	-2.84	**	-1.15	
	(0.79)		(1.07)		(1.18)		(1.36)		(1.19)		(1.31)	
R^2	0.78		0.59		0.63		0.60		0.61		0.54	

*p < .05, **p < .01, ***p < .001