



Technical Appendix • May 2016

DREAMBOX LEARNING ACHIEVEMENT GROWTH in the Howard County Public School System and Rocketship Education

Abstract

In this technical appendix, we present analyses of the relationship between usage of the DreamBox mathematics program and student achievement in Grades 3–5 in the Howard County Public School System (HCPSS) and the Rocketship Education (Rocketship) charter school network for the 2013–2014 and 2014–2015 school years. Our analyses of nearly 3,000 students per year include all classrooms in which at least 75% of students had some DreamBox usage during the school year (approximately 100 total classrooms each year across both sites). We consider achievement on the Northwest Evaluation Association (NWEA) Measures of Academic Progress (MAP) mathematics assessment as well as on state assessments in Maryland and California in 2013–2014 and the PARCC assessment in Howard County in 2014–2015.

Individual student usage measures based on time spent on DreamBox lessons and the number of DreamBox lessons completed fell below DreamBox recommendations in both sites. Regression analyses controlling for students' prior test scores generally suggest time spent in the DreamBox program (particularly time in lessons recommended by DreamBox) was positively and significantly related to achievement gains on MAP and state tests. The regression coefficients imply that a student at the median (or 50th percentile) in Howard County would gain about 1.4 percentile points on the MAP for 7.1 hours of DreamBox usage. A student at the median in Rocketship would gain about 3.9 percentile points on the MAP with 8.1 hours of DreamBox usage.

One of the weaknesses of the multiple regression analysis is that DreamBox usage may partially reflect students' motivation levels if students who were willing to spend more time on the software were likewise inclined to spend more time studying. In light of this, we also matched HCPSS students in DreamBox classrooms to similar students in HCPSS schools without DreamBox and found that, in 2014–2015, an average student in a DreamBox classroom performed about 1.9 percentile points better on the MAP and 1.7 percentile points better on the PARCC after an average DreamBox usage of 7.1 hours. We separately examined the relationship between DreamBox usage and achievement gains for the same students at different points in time and found similar positive and statistically significant relationships in HCPSS and Rocketship.

Context and Overview

In 2014, the Center for Education Policy Research at Harvard University piloted a new type of research in partnership with HCPSS and Rocketship. Our goal was to explore the impact of DreamBox Learning on student achievement and to determine if there were any patterns of usage related to improved achievement outcomes for students using only data normally collected by HCPSS, Rocketship, and DreamBox.

Using data from the 2013–2014 and 2014–2015 school years, the guiding research questions were:

1. How did HCPSS and Rocketship elementary schools implement DreamBox in their classrooms over these two years?
2. What was the relationship between DreamBox usage and achievement gains on interim and end-of-year assessments for students in these schools?

3. Was DreamBox adoption causally related to changes in students' achievement?

This document is divided into four sections. In the first section, we describe our sample and our definition of “using-classrooms.” In the second section, we describe the variation in usage at the classroom and student levels in HCPSS and Rocketship. In the third section, we describe the relationship between DreamBox usage and measured achievement gains for individual students on interim assessments (i.e., MAP) and state tests. In the final section, we describe our analyses of the impact of DreamBox adoption on student achievement in both sites.

Section 1: Our Sample

To understand the impact of DreamBox usage and to facilitate our ability to identify students receiving the DreamBox treatment, we created a construct we call “using-classrooms.” We define using-classrooms as those where 75% or more of the students used DreamBox during a particular school year. Within these using-classrooms, we did not impose any artificial minimum on the amount of usage for a student to be included in the sample, so we could capture the full range of usage patterns in classrooms in which 75% or more of students had some DreamBox usage.

We limited our sample to students in Grades 3 through 5 in all analyses so that we could include end-of-year state test performance as an outcome measure. Students who were introduced to DreamBox outside of one of our using-classrooms were excluded from the analysis.

Following these definitions, in HCPSS we identified 82 using-classrooms in seven elementary schools in 2013–2014 and 68 using-classrooms in five elementary schools in 2014–2015. In Rocketship schools in California, we identified 18 using-classrooms in eight elementary schools in 2013–2014 and 28 using-classrooms in eight elementary schools in 2014–2015.¹

In HCPSS, the using-classrooms included 1,363 students in 2013–2014 and 1,116 in 2014–2015. In Rocketship, they included 1,556 students in 2013–2014 and 1,870 in 2014–2015.²

The availability of assessment scores varied across sites. For most students, we had end-of-year scores on the California or Maryland state math tests for the spring of 2013 and 2014. We also had end-of-year scores on the PARCC mathematics assessment in Maryland for the spring of 2015. (We did not receive state test scores from California in the spring of 2015.) In addition, for all students in Rocketship schools and a subset of HCPSS schools, we received scores on the MAP test from the fall, winter, and spring of the 2013–2014 school year. We have full MAP coverage for fall, winter, and spring in 2014–2015 for Rocketship and HCPSS. (See Table 1.)

¹ “Classrooms” in Rocketship are defined as groups of 50–110 students in the same school and grade level that receive group instruction from a given math teacher. They may participate in lab time with different teachers and tutors. See <http://www.rsed.org/Blended-Learning.cfm>.

² Because students in the usage data had IDs in multiple formats, we matched students with non-standard IDs by full name, school, and grade level. Our definition of using-classrooms thus captures students who both (a) appear in the usage data and (b) successfully merge with class records by agency ID or name. The upper limit on the number of students in Grades 3–5 in schools with concentrated usage who appear not to have had usage but might have just merged unsuccessfully due to unmatched IDs was 69 in HCPSS in 2013–2014 and 263 in 2014–2015, and 211 in Rocketship in 2013–2014 and 44 in 2014–2015. The 82 using-classrooms in HCPSS in 2013–2014 were taught by 65 unique teachers, and the 68 using-classrooms in 2014–2015 were taught by 57 unique teachers.

Table 1. Inclusion of Interim and State Assessments, by Site and by Year

Site	Interim assessments						State assessments		
	2013–2014			2014–2015			2013	2014	2015
	Fall	Winter	Spring	Fall	Winter	Spring	Spring	Spring	Spring
HCPSS	Some	Some	Some	Yes	Yes	Yes	Yes	Yes	Yes
Rocketship	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No

Section 2: Student Usage of DreamBox

FINDING 1: Average usage time and lesson completion were similar in using-classrooms between the two sites, but usage in both sites fell short of DreamBox recommendations (see inset).

In HCPSS, students averaged 17 weeks with any usage of DreamBox, and spent an average of 38 minutes per week with some usage in 2013–2014. The following year, students averaged 16.5 weeks with any usage and spent an average of 35 minutes per week with some usage. They spent an average of 16 minutes per session in 2013–2014 and in 2014–2015.

In Rocketship, students averaged 13 weeks with positive usage of DreamBox, and spent an average of about 42 minutes per week with some usage in 2013–2014. In 2014–2015, students averaged 17 weeks with usage and spent an average of 44 minutes per week with positive usage. They averaged 17 minutes per session in 2013–2014 and 19 minutes per session in 2014–2015.

Table 2 describes students' usage by site and year. Figure 1 illustrates usage time per week, time per session, and lessons completed per week in both sites.

DreamBox Usage Recommendations:

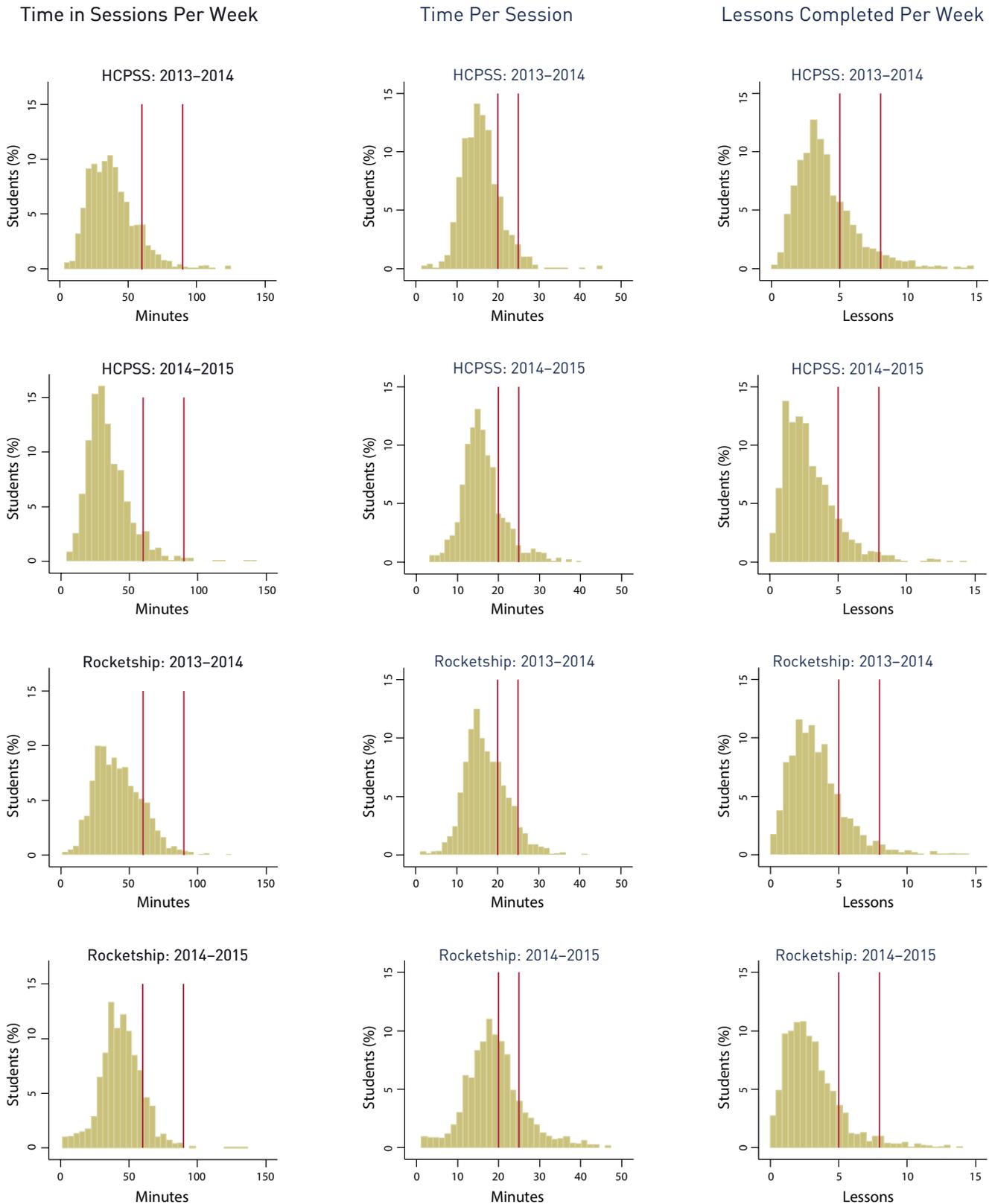
- 60–90 minutes of usage per student per week
- 20–25 minutes per session
- 5–8 lessons completed per week

Table 2. Usage of DreamBox, by Site and by Year

Usage measure	HCPSS		Rocketship	
	2013–2014	2014–2015	2013–2014	2014–2015
Number of classrooms using DreamBox	82	68	18	28
Number of students in using-classrooms	1,363	1,116	1,556	1,870
Share of schools' students in using-classrooms	76.4%	62.3%	90.6%	100%
Students' average total usage time	739.9 minutes	651.7 minutes	611.4 minutes	814.3 minutes
Students' average usage time per week	38.4 minutes	34.7 minutes	42.0 minutes	44.4 minutes
Students' average weeks with positive usage	17.2 weeks	16.5 weeks	13.3 weeks	17.4 weeks
Students' average lessons completed per week	4.0 lessons	2.9 lessons	3.4 lessons	3.1 lessons
Students' average usage time per session	15.9 minutes	16.2 minutes	17.2 minutes	19.1 minutes

Note. Shares of students within HCPSS are calculated out of the seven schools with usage in 2013–2014, not the district as a whole. Shares are calculated out of all eight Rocketship schools. This share is the number of students in using-classrooms divided by the total number of students in these schools. Average total usage time, average usage time per week, average weeks with positive usage, average lessons completed per week, and average usage time per session are representative of students in using-classrooms.

Figure 1. Time in Sessions Per Week, Time Per Session, and Lessons Completed Per Week in HCPSS and Rocketship, 2013–2014 and 2014–2015



Note. Red lines indicate the recommended ranges of DreamBox usage.

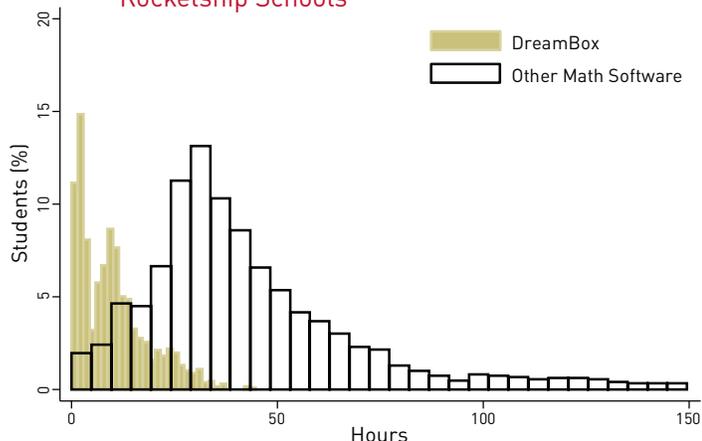
FINDING 2: While patterns were similar, DreamBox usage differed between the two sites in at least three ways:

- 1) In HCPSS, DreamBox tended to be used for instruction with students who were behind in math.
- 2) In HCPSS, out-of-school use seemed to be driven by after-school programs; any out-of-school use was less common in Rocketship.
- 3) In Rocketship schools, DreamBox was one of several math programs students used, while it was the primary math software product in use in HCPSS.

First, teachers in HCPSS seemed to use DreamBox to help students who were behind in math to catch up. In assessing the relationship between the time that students spent in sessions and their prior-year test scores, we found a negative and statistically significant relationship in both years in HCPSS. Specifically, the correlation between time in sessions and prior-year state test scores was -0.13 and statistically significant in 2013–2014 and -0.18 and significant in 2014–2015. Similarly, the correlation between the number of weeks with usage and prior-year math scores was -0.20 and significant in 2013–2014, and -0.14 and significant in 2014–2015. These negative correlations suggest that students with lower test scores tended to have higher usage. In Rocketship, by contrast, both correlations were small and insignificant for 2013–2014 and -0.05 and -0.07 , respectively, and significant for 2014–2015. In other words, a Rocketship student’s usage was not heavily associated with his or her achievement on the prior year’s test.

Second, usage outside of school hours differed across the two locations. Usage time outside school for HCPSS generally appears to have been spent in before-school or after-school intervention programs or extra tutoring sessions for students who were selected for additional DreamBox time. Time outside school for Rocketship students was generally voluntary and also less common. The average HCPSS student spent 132.0 minutes in out-of-school usage in 2013–2014 and 117.4 minutes in 2014–2015; the average Rocketship student spent 47.0 minutes in 2013–2014 and 67.2 minutes in 2014–2015.

Figure 2. Total Hours of Math Software Usage in Rocketship Schools

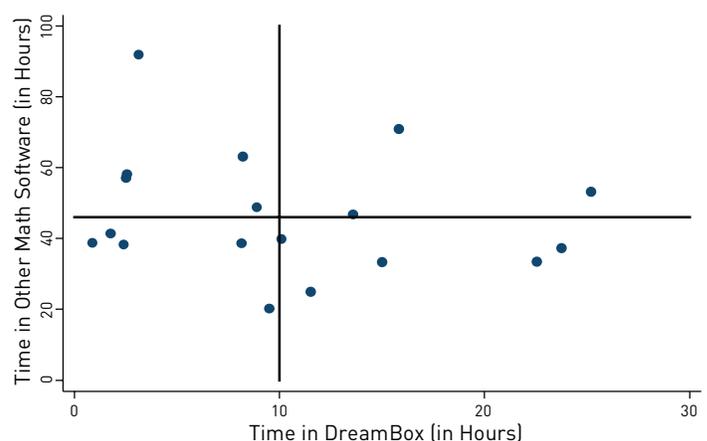


Note. Not represented are 45 students (2.9% of the usage sample) who had over 150 hours of usage on other software.

Third, in HCPSS schools, students were not directed to use other math software, and we have no record of any other usage. Conversely, students in using-classrooms in Rocketship schools averaged 2,843 minutes (or about 47 hours) of usage in other math software in 2013–2014. Moreover, many students used both DreamBox and the other software for math. Figure 2 portrays the average amount of DreamBox usage and other math software usage by students within Rocketship schools.

It is important to note that while students attending Rocketship schools utilized other software, our analyses indicate there was little relationship between the amount of DreamBox usage in a classroom and usage of other math software. At the classroom level, the correlation between DreamBox and other math software usage was -0.20 ; however, this correlation was not statistically distinguishable from zero. (See Figure 3 for an illustration of DreamBox usage and other software usage by classroom in Rocketship schools.) Therefore, although we do have a measure of the amount of time students spent with the other software, the estimated relationship between DreamBox usage and student achievement gains is largely unaffected by whether or not we control for the other usage.³

Figure 3. Time Using DreamBox and Other Math Software



Note. Each dot represents a using-classroom in Rocketship in 2013–2014.

³ Our estimates would be affected only if there were a statistically significant correlation between DreamBox usage and other software usage. If the correlation were positive, then we would overstate the relationship between DreamBox usage and achievement by not controlling for usage of the other software. Alternatively, if the correlation were negative, the coefficient on DreamBox usage would have been somewhat understated. Because the correlation was small, however, we do not believe it had a large effect on our findings.

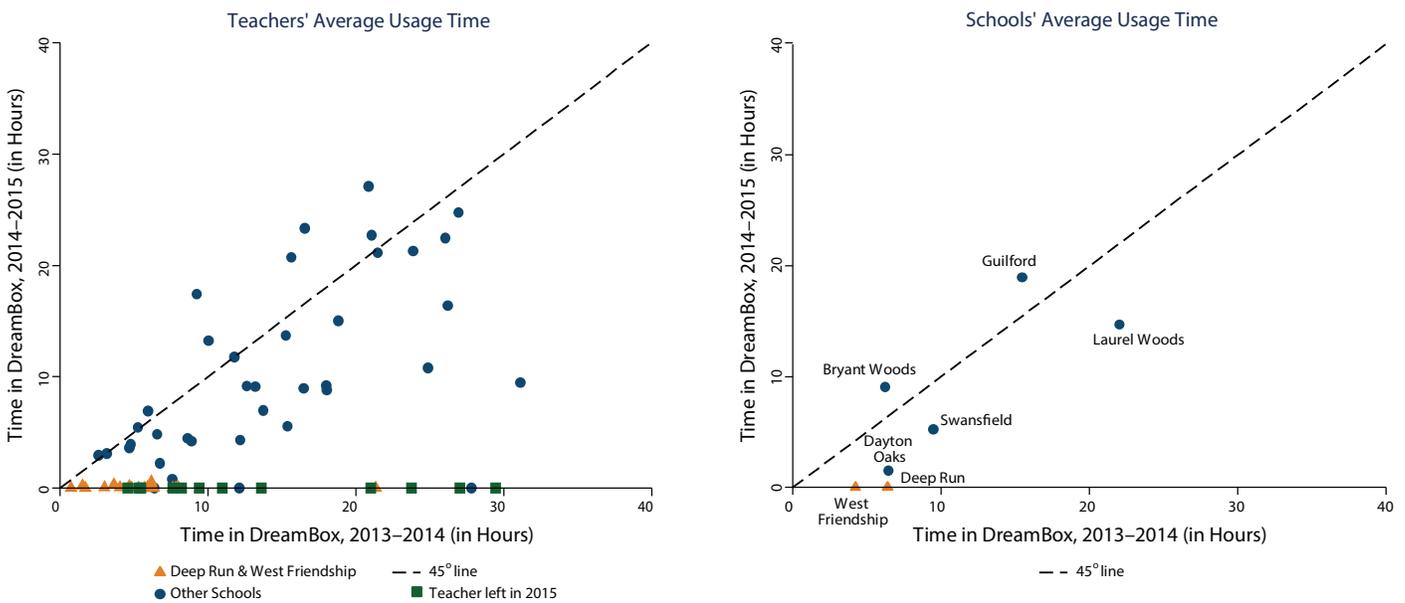
FINDING 3: Variation in software usage was driven largely by teacher- and school-level practices, rather than by student choices.

Teacher and school implementation were important drivers of the differing levels of DreamBox usage by students. Within both sites, more than half of the variation in student-level usage was associated with the school attended or the teacher to whom the student was assigned. For instance, in 2013–2014 in HCPSS, 23% of the variance in students’ DreamBox usage was due to differences in teacher assignment, and 34% of the variance was associated with the school attended.⁴ The relationships were similar in 2014–2015. In Rocketship in 2013–2014, 18% of the variance in students’ DreamBox usage was related to the teacher and 56% was related to the school. In 2014–2015, a large proportion of the variance (36%) was due to teacher assignment and 27% was due to schools. In short, more than half of the

variance in student usage in our sample was due to between-school and between-teacher, within-school differences.

We examined the relationship between HCPSS teachers’ usage of DreamBox in 2013–2014 with their usage of DreamBox in 2014–2015, and found that the extent to which a teacher used the software in his/her classroom in the prior year affected classroom usage the subsequent year; teachers and schools that used DreamBox heavily one year tended to do so the next year. In Figure 4, we compare the average amount of time students in a class used DreamBox in 2014–2015 relative to a similar period in 2013–2014. Among those with non-zero usage in both years, the correlation is 0.78.⁵

Figure 4. HCPSS Teachers’ and Schools’ Average DreamBox Usage Time: 2013–2014 versus 2014–2015



⁴ Because the teacher effects were measured by differences in classroom usage within school, the percentages are additive and sum to 100% of the total variance.

⁵ In HCPSS, teachers in two schools—Deep Run and West Friendship—dropped DreamBox for Grades 3–5 between 2013–2014 and 2014–2015 and were, therefore, excluded from the analysis.

Section 3: The Relationship Between DreamBox Usage and Student Achievement Gains

FINDING 4: When pooled across the two sites, individual student usage was positively and significantly correlated with achievement gains.

When pooled across the two sites, the time that students spent in lessons was positively correlated with student achievement gains. The correlation between students' time in lessons and state test gains (0.07) was statistically significant for 2013–2014. The correlations were somewhat higher with student gains on the MAP assessment: 0.15 and significant for spring-to-spring MAP in 2013–2014, 0.13 and significant for spring-to-spring MAP in 2014–2015, and 0.11 and significant for fall-to-spring MAP in 2014–2015.⁶ When we analyzed the results separately for HCPSS and Rocketship, the estimated correlations were similar for MAP. (See Table 3.)

However, while we found an insignificant relationship with the state tests in HCPSS in 2013–2014, that relationship was 0.20 and significant for the PARCC assessment in 2014–2015. Although we did not receive the 2015 state test scores for California, in 2013–2014 the correlation between minutes of individual student usage and state test gains was 0.13 and significant in Rocketship. (See Table 3.)

Table 3. Correlations Between Time in Lessons and Achievement Gains

Measure	Pooled		HCPSS		Rocketship	
	2013–2014	2014–2015	2013–2014	2014–2015	2013–2014	2014–2015
State tests	0.07*	N/A	-0.01	0.20*	0.13*	N/A
Spring-to-spring MAP	0.15*	0.13*	0.17*	0.16*	0.16*	0.12*
Fall-to-spring MAP	0.12*	0.11*	0.09	0.09*	0.15*	0.12*

⁶ For the correlations, we represent achievement gains on state tests as the residual from a regression of current-year test score on prior-year test score, and we represent achievement gains on the MAP as residuals from regressions of current-year spring MAP score on prior-year spring MAP score or current-year fall MAP score.

FINDING 5: Among DreamBox users, time spent on lessons recommended by DreamBox was more strongly associated with achievement gains than time spent on non-recommended lessons.

We examined time in lessons more closely through the lens of DreamBox’s recommended lessons. We did not have a direct measure of time on recommended versus non-recommended lessons. As such, we created an approximation of time spent on lessons that were recommended by multiplying total time in lessons by the

share of lessons completed that were recommended. This calculation assumes that students spent the same amount of time per lesson on recommended and non-recommended lessons, and that students left the same share of recommended and non-recommended lessons incomplete. (See Table 4.)

Table 4. Time in Lessons and Student Achievement on Interim Assessments, by Site and by Year

	HCPSS Mean (Std. Dev.)		Rocketship Mean (Std. Dev.)	
	2013–2014	2014–2015	2013–2014	2014–2015
Time in lessons (in hours)	8.6 (7.3)	7.1 (7.0)	6.3 (5.5)	8.1 (5.6)
Time in recommended lessons (estimated)	5.2 (5.7)	4.7 (5.2)	3.1 (3.5)	6.0 (5.0)
Time in non-recommended lessons (estimated)	3.1 (3.1)	2.0 (2.4)	2.8 (2.7)	1.5 (1.6)
Percentage of variance in time in lessons Associated with teacher assignment	23	23	18	36
Grade 3 raw fall MAP score	198.9 (14.2)	190.7 (14.6)	191.4 (13.6)	191.8 (15.1)
Grade 3 raw fall-spring MAP difference	10.4 (6.9)	11.2 (7.7)	14.9 (8.7)	16.5 (8.5)
Grade 4 raw fall MAP score	204.1 (11.9)	203.6 (14.0)	201.6 (14.1)	202.6 (14.9)
Grade 4 raw fall-spring MAP difference	8.9 (5.7)	9.4 (7.4)	13.5 (8.4)	16.3 (9.1)
Grade 5 raw fall MAP score	213.8 (15.7)	212.7 (15.9)	212.5 (16.2)	208.2 (16.3)
Grade 5 raw fall-spring MAP difference	6.8 (6.3)	8.2 (7.5)	12.9 (9.2)	14.5 (10.4)

Note. The 2013–2014 MAP scores and fall-spring differences in HCPSS come from 316 students in Guilford and Dayton Oaks where the tests were administered. The 2014–2015 MAP scores and fall-spring differences cover 990 students across all elementary schools with usage. Scores and differences in Rocketship cover all elementary schools in both years.

We separately examined relationships for students in HCPSS and Rocketship between both state tests and MAP tests and their time in DreamBox lessons. Specifically, we regressed end-of-year spring MAP scores on beginning-of-year fall MAP scores for Rocketship in 2013–2014 and for both sites in 2014–2015 (with full school coverage for HCPSS). We also regressed the end-of-year test scores in 2014 on prior scores and a measure of time in lessons for both sites and did the same in 2015 with PARCC scores from HCPSS.

As illustrated in Table 5a, for MAP scores, the coefficient on time in lessons was positive and significant in each specification. The coefficient on time in recommended lessons was positive and significant in each specification when broken out, and the difference in the coefficients for time in recommended and non-recommended lessons was statistically significant in HCPSS in 2014–2015.

In Table 5b, for state test scores, time in lessons on its own was positively and significantly related to performance for the California Standards Test (CST) in Rocketship in 2013–2014 and PARCC in HCPSS in 2014–2015, and the coefficient on time in lessons was indistinguishable from zero for Maryland School Assessment (MSA) in HCPSS in 2013–2014. Time in recommended lessons separately was significant and positive with a larger magnitude for CST and PARCC as well, and time in non-recommended lessons was insignificant for both sites in 2013–2014 and negative and significant for PARCC in HCPSS in 2014–2015. Differences between coefficients for time in recommended and non-recommended lessons were statistically significant for all specifications in Table 5b.

Table 5a. Associations Between Time in Lessons and Fall-to-Spring Changes in MAP, by Site and by Year

Fall-Spring Changes in MAP in Rocketship, 2013–2014			
	Coefficient (se)	Coefficient (se)	
Time in lessons (in hours)	0.015* (0.003)		
Time in recommended lessons (in hours)			0.019* (0.005)
Time in non-recommended lessons (in hours)			0.012* (0.006)
<i>N</i> =	1,382		1,382
<i>R</i> ² =	0.683		0.684

Fall-Spring Changes in MAP in HCPSS, 2014–2015			Fall-Spring Changes in MAP in Rocketship, 2014–2015		
	Coefficient (se)	Coefficient (se)		Coefficient (se)	Coefficient (se)
Time in lessons (in hours)	0.005* (0.002)		Time in lessons (in hours)	0.012* (0.002)	
Time in recommended lessons (in hours)		0.010* (0.003)	Time in recommended lessons (in hours)		0.013* (0.003)
Time in non-recommended lessons (in hours)		-0.003 (0.007)	Time in non-recommended lessons (in hours)		0.014 (0.008)
<i>N</i> =	953	953	<i>N</i> =	1,712	1,712
<i>R</i> ² =	0.799	0.799	<i>R</i> ² =	0.684	0.685

Note. All regressions also include a control for score on the fall test. The sample includes students in Grades 3–5 in both sites. Samples are limited to students in using-classrooms for the respective school years.
* *p* < .05.

Table 5b. Associations Between Time in Lessons and Year-Over-Year Changes in State Tests, by Site and by Year

Year-Over-Year Changes in MSA in HCPSS, 2013–2014			Year-Over-Year Changes in CST in Rocketship, 2013–2014		
	Coefficient (se)	Coefficient (se)		Coefficient (se)	Coefficient (se)
Time in lessons (in hours)	-0.001 (0.002)		Time in lessons (in hours)	0.016* (0.003)	
Time in recommended lessons (in hours)		0.004 (0.003)	Time in recommended lessons (in hours)		0.023* (0.006)
Time in non-recommended lessons (in hours)		-0.016 (0.008)	Time in non-recommended lessons (in hours)		0.008 (0.008)
<i>N</i> =	705	705	<i>N</i> =	1,231	1,231
<i>R</i> ² =	0.742	0.743	<i>R</i> ² =	0.554	0.555
Year-Over-Year Changes from MSA to PARCC in HCPSS, 2014–2015					
	Coefficient (se)	Coefficient (se)			
Time in lessons (in hours)	0.016* (0.003)				
Time in recommended lessons (in hours)		0.035* (0.005)			
Time in non-recommended lessons (in hours)		-0.034* (0.011)			
<i>N</i> =	586	586			
<i>R</i> ² =	0.700	0.707			

Note. All regressions also include a control for score on the prior-year state test. The HCPSS samples include students in Grades 4–5, and the Rocketship sample includes students in Grades 3–5. Samples are limited to students in using-classrooms for the respective school years.
* *p* < .05.

Combining our understanding of average student usage and the effects of student usage on achievement gains, we were able to approximate the percentile-point gains that students could expect to achieve based on their usage. On average, students in HCPSS had 7.1 hours of usage in 2014–2015 and students in Rocketship had 8.1 hours of usage in the same year. The regression coefficients in Table 5a imply an achievement gain of 0.005 standard deviations per hour * 7.1 hours = 0.036 standard deviations in HCPSS and 0.012 standard deviations per hour * 8.1 hours = 0.097 standard deviations in Rocketship for the MAP assessment. In other words, a student at the median (or 50th percentile) in HCPSS would gain about 1.4 percentile points with 7.1 hours of usage and about 2.9 percentile points with 14.2 hours of usage. A student at

the median in Rocketship would gain about 3.9 percentile points with 8.1 hours of usage and 7.7 percentile points with 16.2 hours of usage.⁷

We replicated these regressions specifically for students in Rocketship to check the sensitivity of the relationships to usage of other mathematics software. In the first column of each panel in Table 6, we look only at the effect of time spent in DreamBox. In the second column, we add a control for other software usage. In neither case did controlling for other software usage change the coefficient on DreamBox usage. As illustrated in Table 6, in each case, coefficients on measures of DreamBox usage time that were significant remained positive and significant when we controlled for usage time in other mathematics software.

⁷ Separate analyses (not shown) suggest this relationship is linear and that achievement gains continue to rise at the same rate as DreamBox usage increases.

Table 6. Association Between Time in Lessons and Student Achievement in Rocketship Controlling for Other Math Software Usage, 2013–2014

State Test in Rocketship Schools, 2013–2014				MAP Test in Rocketship Schools (Fall to Spring), 2013–2014			
	Coefficient (se)	Coefficient (se)	Coefficient (se)		Coefficient (se)	Coefficient (se)	Coefficient (se)
Time in lessons (in hours)	0.017* (0.004)	0.017* (0.004)		Time in lessons (in hours)	0.016* (0.003)	0.016* (0.003)	
Time in recommended lessons (in hours)			0.025* (0.006)	Time in recommended lessons (in hours)			0.021* (0.005)
Time in non-recommended lessons (in hours)			0.009 (0.008)	Time in non-recommended lessons (in hours)			0.013* (0.006)
Other math software usage (in hours)		0.000 (0.000)	0.000 (0.000)	Other math software usage (in hours)		0.000 (0.000)	0.000 (0.000)
<i>N</i> =	1,216	1,216	1,216	<i>N</i> =	1,364	1,364	1,364
<i>R</i> ² =	0.560	0.560	0.562	<i>R</i> ² =	0.684	0.685	0.686

Note. All regressions also include a control for score on the prior-year state test. The HCPSS samples include students in Grades 4–5, and the Rocketship sample includes students in Grades 3–5. Samples are limited to students in using-classrooms for the respective school years.
* *p* < .05.

FINDING 6: Progress at grade level through the DreamBox curriculum appears to be positively and significantly associated with achievement gains when controlling for baseline test scores.

We examined achievement gains as related to the amount of progress students made through the DreamBox curriculum. We defined total progress at grade level as the percentage of the DreamBox curriculum completed in a student’s own grade level, as reported by DreamBox for the latest week the student appears in the data from 2014–2015. This is separate, as defined by DreamBox, from content in a grade level skipped or passed over as a result of a placement exam.

On average in HCPSS, students in using-classrooms in Grades 3 through 5 completed 8.5% of the curriculum at grade level in 2013–2014 and 10.2% in 2014–2015. In Rocketship, students completed 2.0% in 2013–2014 and 12.5% in 2014–2015, on average. We defined total progress below grade level as the year-end amount of curricular progress for the grade level one below the student’s, and we defined total progress above grade level as the amount of progress for the grade level one above the student’s.

We separately regressed MAP test scores from the spring of 2015 on fall MAP scores from the same year and measures of progress at, above, and below each student’s grade level. We also regressed PARCC test scores from

the spring of 2015 on prior-year MSA scores and the same measures of progress relative to grade level in HCPSS. For both PARCC and MAP in both sites, total progress at grade level was positively and significantly related to achievement. In HCPSS, progress below grade level was positively and significantly related to achievement on PARCC and MAP, and progress above grade level was positively and marginally significantly related to MAP achievement.

The regression coefficients in Table 7 imply an achievement gain of 0.004 standard deviations per percentage point of completion * 10.2% = 0.041 standard deviations in HCPSS and 0.006 standard deviations per percentage point of completion * 12.6% = 0.076 standard deviations in Rocketship for the MAP assessment. In other words, a student at the median (or 50th percentile) in HCPSS would gain about 1.6 percentile points with 10.2% progress at grade level and about 3.3 percentile points for 20.4% progress at grade level. A student at the median in Rocketship would gain about 3 percentile points with 12.6% progress at grade level and 6 percentile points with 25.2% progress at grade level.

Table 7. Associations Between Grade-Level Progress and Student Achievement, by Site and by Year

Fall-Spring Changes in MAP in HCPSS, 2014–2015		Year-Over-Year Changes for PARCC in HCPSS, 2014–2015	
	Coefficient (se)		Coefficient (se)
Total progress at grade level (in percentage points)	0.004* (0.002)	Total progress at grade level (in percentage points)	0.009* (0.003)
Total progress above grade level (in percentage points)	0.001 (0.003)	Total progress above grade level (in percentage points)	0.002 (0.006)
Total progress below grade level (in percentage points)	0.001* (0.001)	Total progress below grade level (in percentage points)	0.005* (0.002)
	<i>N</i> = 953		<i>N</i> = 586
	<i>R</i> ² = 0.802		<i>R</i> ² = 0.724
Fall-Spring Changes in MAP in Rocketship, 2014–2015			
	Coefficient (se)		
Total progress at grade level (in percentage points)	0.006* (0.001)		
Total progress above grade level (in percentage points)	0.003 (0.002)		
Total progress below grade level (in percentage points)	0.000 (0.001)		
	<i>N</i> = 1,712		
	<i>R</i> ² = 0.692		

Note. All regressions also include controls for baseline test score and time in DreamBox lessons during the 2014–2015 school year. The PARCC sample includes students in Grades 4 and 5 in HCPSS, and the MAP test samples include students in Grades 3–5 in both sites. All samples are limited to students in using-classrooms in 2014–2015.
* *p* < .05.

Two Approaches to Understanding DreamBox’s Effect in HCPSS and Rocketship

One of the weaknesses of the above analysis is that DreamBox usage may partially reflect students’ motivation levels—students who are willing to spend more time on the software may also be inclined to spend more time studying. As a result, the estimated impact of software usage may reflect differences in student motivation, not differences in software usage. To the extent that more motivated students simply have higher baseline achievement, we have controlled for students’ prior achievement levels. However, it could also be that more motivated students show faster gains in achievement.

Therefore, we explored two additional strategies for trying to isolate the causal impact of DreamBox usage. First, in HCPSS, we compared the gains of students in DreamBox using-classrooms to the gains of similar students in schools without DreamBox. Second, we compared changes from one semester to the next in the amount of usage against changes in achievement for the same students.

Matching Students in Using-Classrooms to Similar Students in Non-Using-Classrooms

For each student in a using-classroom, we found the closest available matching student in a non-using-classroom in the same grade elsewhere in HCPSS (matched on the basis of a baseline math score, an indicator of whether the student was Black or Hispanic, and the mean score of other students in the class on the baseline assessment). We focused on HCPSS because not all schools in HCPSS used DreamBox, and it was the district’s decision not to deploy DreamBox in all schools. Because it was out of the control of the students, it provided us with a “natural experiment” with which to test the impact of DreamBox. (In Rocketship, by contrast, students in non-using classrooms may have chosen to use other software in the learning lab.)

Table 8 displays average characteristics of students in using-classrooms and matched students in non-using-classrooms. The students in the matched classrooms had similar prior achievement, classmates with similar prior achievement, and similar demographic characteristics.

Table 8. Average Characteristics of Matched Students

	Average Characteristics of Matched Students					
	Spring 2015 MAP as Outcome			2014–2015 PARCC as Outcome		
	Prior Score	Peers' Prior Score	% Black or Hispanic	Prior Score	Peers' Prior Score	% Black or Hispanic
Using-classrooms	-0.482	-0.482	56.4%	-0.475	-0.480	56.3%
Non-using-classrooms	-0.473	-0.471	56.4%	-0.469	-0.471	56.3%

Table 9 displays end-of-year differences in achievement between the two groups on the spring 2015 MAP assessment and 2014–2015 PARCC assessment. Students in using-classrooms in Grades 3–5 scored on average 0.048 standard deviations higher on the spring 2015 MAP and .042 standard deviations higher on the spring 2015 PARCC exam. These differences are roughly consistent with what we would have predicted based on the relationships between student usage and achievement gains we reported in the prior section.

For instance, if we multiplied average usage time in HCPSS from Table 4 by the coefficient on time in lessons for each assessment from Tables 5a and 5b, we would expect differences on the order of $7.1 * 0.005 = 0.036$ standard deviations for spring 2015 MAP and $7.1 * 0.016 = 0.114$ standard deviations for PARCC. These results suggest that the relationships between usage and achievement gains in the initial regressions were not simply driven by correlations with unmeasured student or teacher traits.

We find similar effects when comparing using-classrooms to similar non-using-classrooms on end-of-year achievement. The effect size of 0.048 would move a student at the median for the MAP assessment in HCPSS up by about 1.9 percentile points for 7.1 hours of usage and 3.8 percentile points for 14.2 hours of usage. The effect size of 0.042 would move a student at the median for the PARCC assessment in HCPSS up by about 1.7 percentile points for 7.1 hours of usage and about 3.4 percentile points for 14.2 hours of usage. These effect sizes are very similar to the result from an earlier study of DreamBox in Rocketship schools that employed random assignment and found a gain of 5.5 percentile points on the MAP for average usage of 21.8 hours among treated students.⁸

Table 9. Difference Between Treated Students' and Matched Control Students' Achievement in HCPSS, 2014–2015, Grades 3–5

Outcome	Prior	Sample Size	Difference
Spring 2015 MAP	Fall 2014 MAP	990	0.048**
2015 PARCC (Std.)	Fall 2014 MAP	1,000	0.042*

* p < .05. ** p < .01

Changes in Usage and Changes in Achievement Gains for Students Over Time

Another way to try to reduce the influence of unobserved student traits such as motivation is to examine the relationship between usage time in the software and achievement gains at different points in time for the same students. If student motivation is fixed, but usage varies by semester, we should see faster gains during semesters with greater levels of usage. Further, if students have faster gains in semesters with more usage relative to semesters with lower usage when they have the same teachers, we can be more confident these gains are not arising from differences in teacher characteristics.

For this analysis, we exploited the fact that the MAP assessment is administered three times per school year—in the fall, at mid-year, and in the spring. As a result, we could measure student achievement gains during the fall and spring semesters of both school years (2013–2014 and 2014–2015).⁹ We represent the change in MAP score from the fall to winter or from the winter as the outcome for all models in Table 10. The models in Columns A, D, and G only include controls for time in lessons and semester. The models in Columns B, E, and H add teacher-year fixed effects, and the models in Columns C, F, and I add student fixed effects to generate the within-teacher-year and within-student comparisons that are of primary interest for this analysis.

In Table 10, the coefficient on time in lessons is positive and statistically significant in all three specifications in the pooled sample. Overall, within-teacher and within-student, greater levels of Dreambox usage were associated with faster gains on the MAP test. The same is true within Howard County and Rocketship, although the coefficient is only marginally statistically significant in Howard County.

⁸ Wang, H., & Woodworth, K. (2011). Evaluation of Rocketship Education's use of DreamBox Learning's online mathematics program. (Menlo Park, CA: SRI International).

⁹ The period between the fall and winter assessments in 2013–2014 corresponds to October 29–December 13, 2013; the period between the winter and spring assessments in 2013–2014 corresponds to February 2–April 27, 2014; the period between the fall and winter assessments in 2014–2015 corresponds to October 12–December 21, 2014; and the period between the winter and spring assessments in 2014–2015 corresponds to February 1–April 12, 2015.

Table 10. Association Between Time in Lessons and MAP Achievement with Student-Year Fixed Effects, Pooled and by Site for 2013–2014 Through 2014–2015

	Pooled			HCPSS			Rocketship		
	(A) Coefficient (se)	(B) Coefficient (se)	(C) Coefficient (se)	(D) Coefficient (se)	(E) Coefficient (se)	(F) Coefficient (se)	(G) Coefficient (se)	(H) Coefficient (se)	(I) Coefficient (se)
Time in lessons (in hours)	0.016* (0.002)	0.016* (0.003)	0.020* (0.006)	0.010* (0.003)	0.013* (0.004)	0.025 (0.013)	0.017* (0.003)	0.015* (0.004)	0.015* (0.007)
N =	7,595	7,595	7,595	2,077	2,077	2,077	5,518	5,518	5,518
R ² =	0.005	0.035	0.320	0.025	0.054	0.349	0.005	0.034	0.317

Note. Samples include students in using-classrooms in Grades 3-5. All specifications include dummies for semester and cluster standard errors at the student level. Columns A, D, and G include no fixed effects; Columns B, E, and H include teacher-year fixed effects; and Columns C, F, and I include student fixed effects. The pooled models include an indicator for partner site.

* p < .05.

Summary and Conclusion

Our initial correlations and regressions suggest there is a positive relationship between students’ time using DreamBox and their achievement gains on state tests and MAP assessments. The one assessment for which we cannot reject the hypothesis of no relationship is the 2013–2014 MSA state test in HCPSS (the district’s first year of a DreamBox intervention). We controlled for students’ prior test scores, but if students who used the software longer were more motivated or were different in other ways we could not observe, we might have been picking up these unobserved characteristics rather than identifying the real causal effect of the software on achievement gains.

In order to address the role that characteristics like motivation might play, we employed two additional approaches to get closer to an understanding of the causal effect of DreamBox usage. In HCPSS, both approaches to identify more causal evidence of DreamBox’s effect yielded positive and statistically significant relationships between usage and achievement gains. The natural experiment afforded by HCPSS’s adoption of DreamBox in some but not all elementary schools suggested that students in DreamBox using-classrooms scored between 0.04 and 0.05 standard deviations higher on the 2014–2015 PARCC assessment and the spring 2015 MAP assessment than did students in non-using-classrooms who were similar in their fall 2014 MAP scores, classmates’ average fall 2014 MAP scores, and racial/ethnic minority status. We also found evidence in HCPSS that individual students scored better on the winter and spring MAP assessments in 2013–2014 and 2014–2015 when they spent more time using the software.

In Rocketship, we could not conduct the natural experiment because all schools employed DreamBox, but we could analyze usage time and achievement gains at different points in time for the same students. We found results that were similar to our initial regressions and similar to our findings from the same analysis in HCPSS. Along with the findings from the natural experiment in HCPSS, these results support the findings from our initial regressions (see Tables 5a and 5b).

These findings are encouraging but not definitive, especially given the observation that students in both HCPSS and Rocketship used the software for less time than DreamBox would recommend. We would expect to see a more reliable relationship between usage and achievement gains if students met or exceeded DreamBox’s recommendations for usage, on average. In addition, HCPSS’s usage in our data represents the first two years of implementation of the intervention, and teachers and school leaders should become more comfortable with the software as they gain experience administering it. In other sites that newly adopt DreamBox, as in HCPSS, additional natural experiments should be available to assess the effectiveness of DreamBox whenever school leaders introduce the software in some but not all schools, grade levels, or classrooms.

This project was made possible through a generous gift from Sheryl Sandberg and David Goldberg. The analysis included in this report is that of the authors alone and does not necessarily reflect the views of the funders.