Adaptive Math and Student Achievement: Evidence from a Randomized Controlled Trial of DreamBox Learning

Matthew A. Lenard* Anisa Rhea†

September 29, 2018

Prepared for the Society for Research on Educational Effectiveness (SREE) Spring 2019 Conference

Abstract

This paper reports on the implementation and impacts of DreamBox Learning on elementary math outcomes in a large, southeastern school district. DreamBox is an adaptive, online solution that delivers individualized lessons to nearly 3 million students in more than 100,000 classrooms nationwide. In spring 2018, the program was randomly assigned to 12 schools serving roughly 8,000 students in grades K-5. Implementation results show that most students successfully completed the recommended number of weekly lessons but failed to spend sufficient time on the program to meet the vendor’s usage guidelines. Intent-to-treat effects show that students gained 0.12 standard deviations on an early elementary skills test but did not outperform the control group on the state end-of-grade test.

Keywords: math, education technology, randomized controlled trial

*Harvard Graduate School of Education, mlenard@g.harvard.edu (corresponding author)
†Wake County Public School System, arhea@wcpss.net
1 Background

School districts have increasingly adopted technology-based resources in their attempts to improve student learning (Voogt and Voogt-Knezev, 2008). To this end, districts spend an estimated $8 billion annually on educational software (Richards and Stebbins, 2014). In the midst of this spending, a growing body of research has attempted to measure the causal impact of technology resources on various student achievement outcomes (Campuzano et al., 2009; Given et al., 2008; Wijekumar et al., 2012). More recently, a series of meta-analyses have found that the education technology space includes a complex system of products and services that has generated mixed results across a range of subjects, grade levels, subgroups and geographies (Bulman and Fairlie, 2016; Escueta et al., 2017; Cheung and Slavin, 2012). The evaluation herein builds on two previous studies of the program (What Works Clearinghouse, 2013; Wang and Woodworth, 2011; CEPR, 2016) and provides results for the largest randomized trial of the product to date.

2 Setting, Participants & Research Design

To recruit schools for our analytic sample, we notified school leaders from the district’s elementary schools that DreamBox would be offered as a resource for schools willing to commit to full implementation in grades 4 and 5. We received opt-in responses from 24 schools and ranked them by their 2016-17 school-level composite mathematics score. This process resulted in 12 matched pairs and DreamBox was randomly assigned to one school within each pair. The intervention took place during the second semester of the 2017-18 school year.

To estimate the impact of DreamBox on students who are nested within schools and where schools are the unit of assignment, we fit our data to a cluster two-level model with random effects. We estimate impacts on two sets of outcome variables: The Number Knowledge Test (NKT) (grades K-2) and the state’s end-of-grade (EOG) test (grades 4-5), both of which are standardized with a mean of zero and a standard deviation of one. In doing so, we estimate intent-to-treat (ITT) and treatment-on-treated (TOT) effects. The ITT effect represents the impact of the exogenous offer of DreamBox to the 12-school sample of treatment sites. This provides the empirical impact of the offer irrespective of the rate of takeup and tells us what would happen if all elementary schools were offered this program. This model is specified as:

\[
Outcome_{ij} = \beta_0 + \beta_1(Offer) + Prior_{ij} + Z_{ij} + \lambda_{ij} + \epsilon_j
\]  

Where \(Outcome_{ij}\) represents our dependent variable of interest—the standardized score for each math measure—for student \(i\) nested in school \(j\); \(\beta_0\) represents the constant; \(\beta_1\) represents the coefficient of our \(Offer\) predictor (the offer of DreamBox) for student \(i\) nested in school \(j\); \(Prior\) represents standardized prior achievement for the same outcome of interest; \(Z\) represents a vector of student- and school-level control variables; residual \(\lambda_{ij}\) represents the random effect of student \(i\) in school \(j\); and residual \(\epsilon_j\) represents the random effect of school \(j\).

To estimate the treatment-on-treated impacts of using DreamBox, we employ a two-stage least squares (2SLS) regression where the exogenous offer of DreamBox is our instrument for use.

3 Data

To develop our analytic sample, we use the district’s administrative data and test scores, as well as usage data from the vendor. The district’s student information system file includes unique
student identifiers, demographic information (e.g., race/ethnicity and sex) and status indicators (e.g., disability, English language learner, and gifted/talented).

To measure usage and estimate treatment-on-treated effects, we use implementation data provided by the vendor. These data contain two primary variables: the number of minutes students spend on the program and the number of lessons they complete. In its promotional literature, DreamBox offers implementation guidelines for each of these two indicators, suggesting that students aim to use the program for 60-90 minutes each week and complete 5-8 lessons during this time. Thus, in reporting descriptive usage data, we group minutes into discrete categories to reflect these goals.

4 Results

Table 1 shows pre-treatment balance between the treatment and control groups. No differences between the treatment and control groups are statistically different, though a slight higher percentage of control group students attend Title I elementary schools.

Figures 1-2 show school-level usage data in terms of the percentage of students meeting DreamBox guidelines for minutes used and lessons completed. At all but two schools, most students did not meet the 60-minute weekly threshold. However, at all schools, roughly one-half to two-thirds of students met the 5-lesson goal, suggesting that students typically complete lessons faster than the vendor’s minutes guidelines suggest.

Figures 5-8 display coefficient plots indicating the direction, magnitude, and statistical significance of a range of impacts. The randomized offer of DreamBox appears to have caused an increase of 0.12 on the Number Knowledge Test (NKT), an effect that is consistent across four separate models and statistically significant ($p < 0.05$) in the fully-specified model that includes controls for prior achievement (NKT MOY test), demographics, and matched-pair fixed effects (Figure 5).

5 Conclusion

The rapid adoption of education technology solutions has led to a comparable level of scrutiny from stakeholders about the ability of such resources to impact student achievement. In this paper, we expand the limited evidence base on DreamBox to show that its moderate impacts are concentrated in the early elementary grades. Students in grades K-2 gained 0.12 on the NKT but these results did not translate into the later elementary grades. One explanation for this inconsistency may be that that NKT primarily measures skill development, a prominent feature of DreamBox, while the EOG measures content mastery. The magnitude of the NKT results are consistent with those reported by both Wang and Woodworth (2011) and CEPR (2016) on the NWEA assessment. On implementation, we find that across schools and grade levels, students were more successful at meeting benchmarks for lessons more so than minutes. This suggests that the vendor should revisit its guidelines and further study the relationships among minutes, lessons and achievement.
References


### Tables

<table>
<thead>
<tr>
<th>Variable</th>
<th>(1) Control Mean/SE</th>
<th>(2) Treatment Mean/SE</th>
<th>T-test Difference (1)-(2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>0.518 (0.006)</td>
<td>0.523 (0.005)</td>
<td>-0.005</td>
</tr>
<tr>
<td>Asian</td>
<td>0.083 (0.031)</td>
<td>0.107 (0.058)</td>
<td>-0.023</td>
</tr>
<tr>
<td>Hispanic</td>
<td>0.173 (0.028)</td>
<td>0.189 (0.028)</td>
<td>-0.016</td>
</tr>
<tr>
<td>Black</td>
<td>0.223 (0.036)</td>
<td>0.218 (0.053)</td>
<td>0.004</td>
</tr>
<tr>
<td>White</td>
<td>0.479 (0.051)</td>
<td>0.450 (0.065)</td>
<td>0.029</td>
</tr>
<tr>
<td>Other</td>
<td>0.122 (0.033)</td>
<td>0.141 (0.058)</td>
<td>-0.018</td>
</tr>
<tr>
<td>LEP</td>
<td>0.107 (0.020)</td>
<td>0.126 (0.018)</td>
<td>-0.020</td>
</tr>
<tr>
<td>SWD</td>
<td>0.115 (0.008)</td>
<td>0.129 (0.007)</td>
<td>-0.015</td>
</tr>
<tr>
<td>Gifted in Math</td>
<td>0.021 (0.004)</td>
<td>0.020 (0.004)</td>
<td>0.001</td>
</tr>
<tr>
<td>Year-Round Calendar</td>
<td>0.406 (0.161)</td>
<td>0.410 (0.161)</td>
<td>-0.004</td>
</tr>
<tr>
<td>Title I</td>
<td>0.370 (0.143)</td>
<td>0.272 (0.130)</td>
<td>0.098</td>
</tr>
<tr>
<td>School FRPL Percentage</td>
<td>0.238 (0.034)</td>
<td>0.257 (0.044)</td>
<td>-0.020</td>
</tr>
<tr>
<td>N</td>
<td>8438</td>
<td>7530</td>
<td></td>
</tr>
<tr>
<td>Clusters</td>
<td>12</td>
<td>12</td>
<td></td>
</tr>
</tbody>
</table>

F-test of joint significance (F-stat) 2.281**
F-test, number of observations 15968

**Notes:** The value displayed for t-tests are the differences in the means across the groups. The value displayed for F-tests are the F-statistics. Standard errors are clustered at variable school_code_short. ***, **, and * indicate significance at the 1, 5, and 10 percent critical level.

Table 1: Pre-Treatment Balance between Treatment and Control Groups
Figures

Figure 1: Percentage Meeting DreamBox Weekly Minutes Benchmarks, by School
Figure 2: Percentage Meeting DreamBox Weekly Lessons Benchmarks, by School
Figure 3: Percentage Meeting DreamBox Weekly Minutes Benchmarks, by Grade
Figure 4: Percentage Meeting DreamBox Weekly Lessons Benchmarks, by Grade
Figure 5: Intent-to-Treat (ITT) Impact of DreamBox on Number Knowledge Test Score, Grades K-3

Note: This chart shows standardized effect sizes. The marker indicates the size of effect. Markers below the red 0-line indicate a negative effect and markers above indicate a positive one. The horizontal lines indicate 95% confidence intervals (CI). If the CI touches the red 0-line, the effect is not significant (p<.05).
Figure 6: Treatment-on-Treated (TOT) Impact of DreamBox on Number Knowledge Test Score, Grades K-3

Note: This chart shows standardized effect sizes. The marker indicates the size of effect. Markers below the red 0-line indicate a negative effect and markers above indicate a positive one. The horizontal lines indicate 95% confidence intervals (CI). If the CI touches the red 0-line, the effect is not significant (p<.05).
Figure 7: Intent-to-Treat (ITT) Impact of DreamBox on End-of-Grade Math Test Score, Grades 4-5

Note: This chart shows standardized effect sizes. The marker indicates the size of effect. Markers below the red 0-line indicate a negative effect and markers above indicate a positive one. The horizontal lines indicate 95% confidence intervals (CI). If the CI touches the red 0-line, the effect is not significant (p<.05).
Figure 8: Treatment-on-Treated (TOT) Impact of DreamBox on End-of-Grade Math Test Score, Grades 4-5

Note: This chart shows standardized effect sizes. The marker indicates the size of effect. Markers below the red 0-line indicate a negative effect and markers above indicate a positive one. The horizontal lines indicate 95% confidence intervals (CI). If the CI touches the red 0-line, the effect is not significant (p<.05).