

DreamBox Learning

Program Description¹

DreamBox Learning is a supplemental online mathematics program that provides adaptive instruction for students in grades K–5 and focuses on number and operations, place value, and number sense. The program aims to individualize instruction for each student with millions of unique paths through the curriculum intended to match each student’s level of comprehension and learning style. The curriculum is based on the National Council of Teachers of Mathematics (NCTM) standards.

Research²

The What Works Clearinghouse (WWC) identified one study of *DreamBox Learning* that both falls within the scope of the Elementary School Mathematics topic area and meets WWC evidence standards. This study meets standards without reservations and included 557 elementary school students in kindergarten and first grade in three charter schools in San Jose, California.

The WWC considers the extent of evidence for *DreamBox Learning* on the math performance of elementary school students to be small for the mathematics achievement domain, the only domain specified in the review protocol.

Effectiveness

DreamBox Learning was found to have no discernible effects on mathematics achievement for elementary school students.

Table 1. Summary of findings³

Outcome domain	Rating of effectiveness	Improvement index (percentile points)		Number of studies	Number of students	Extent of evidence
		Average	Range			
Mathematics achievement	No discernible effects	+4	na	1	557	Small

na = not applicable

Report Contents	
Overview	p. 1
Program Information	p. 2
Research Summary	p. 3
Effectiveness Summary	p. 4
References	p. 5
Research Details for Each Study	p. 6
Outcome Measures for Each Domain	p. 8
Findings Included in the Rating for Each Outcome Domain	p. 9
Supplemental Findings for Each Outcome Domain	p. 10
Endnotes	p. 11
Rating Criteria	p. 12
Glossary of Terms	p. 13

Program Information

Background

DreamBox Learning was developed and is distributed by DreamBox Learning, Inc. Address: 305 108th Avenue NE, 2nd Floor, Bellevue, WA 98004. Email: schools@dreambox.com. Web: <http://www.dreambox.com>. Telephone: (877) 451-7845.

Program details

The *DreamBox Learning* curriculum is made up of more than 720 online lessons for two grade levels: primary lessons for grades K–2 and intermediate lessons for grades 3–5. Students complete lessons by playing math games and solving puzzles using *DreamBox Learning*'s virtual manipulatives. As students interact with the program, the lesson sequence, difficulty, pace, and number of hints are adapted in real time. *DreamBox Learning* can be used in a range of instructional settings including the classroom, small groups, or students using the program independently. Reporting tools allow teachers and administrators to monitor student progress. *DreamBox Learning* recommends students spend at least 90 minutes a week on the program.

Cost

A license to use *DreamBox Learning* for individual students, classrooms, or grades costs \$20 per student per year. A license for a school, which covers an unlimited number of students, costs \$7,000 per year.

Research Summary

The WWC identified 11 studies that investigated the effects of *DreamBox Learning* on the math performance of elementary school students.

The WWC reviewed one of those studies against group design evidence standards. This study (Wang & Woodworth, 2011) is a randomized controlled trial that meets WWC evidence standards without reservations. The study is summarized in this report. The remaining 10 studies do not meet WWC eligibility screens for review in this topic area. Citations for all 11 studies are in the References section, which begins on p. 5.

Table 2. Scope of reviewed research

Grade	K–1
Delivery method	Individual
Program type	Supplement

Summary of studies meeting WWC evidence standards without reservations

Wang and Woodworth (2011) examined the effects of *DreamBox Learning* on 557 kindergarten and first-grade students in three charter elementary schools located in San Jose, California. Within each school, all students in grades K and 1 were randomly assigned to either an intervention or comparison group. The intervention group received *DreamBox Learning* supplemental mathematics instruction every day, while the comparison group received additional literacy instruction. Among the randomly assigned students, some students identified by the school as low-achieving received after-school programming known as Response to Intervention (RtI). Those students also received *DreamBox Learning* instruction in their after-school RtI programming, regardless of intervention status. For intervention group students who participated in RtI, the after-school *DreamBox Learning* instruction was in addition to the *DreamBox Learning* instruction provided during the school day.

Summary of studies meeting WWC evidence standards with reservations

No studies of *DreamBox Learning* met WWC evidence standards with reservations.

Effectiveness Summary

The WWC review of *DreamBox Learning* for the Elementary School Mathematics topic area includes student outcomes in one domain: mathematics achievement. The findings below present the authors' estimates and WWC-calculated estimates of the size and statistical significance of the effects of *DreamBox Learning* on elementary school students. For a more detailed description of the rating of effectiveness and extent of evidence criteria, see the WWC Rating Criteria on p. 12.

Summary of effectiveness for the mathematics achievement domain

One study that meets WWC standards without reservations reported findings in the mathematics achievement domain.

Wang and Woodworth (2011) reported a statistically significant positive difference between students in the *DreamBox Learning* and comparison groups on the Measures of Academic Progress (MAP) overall math score. However, this analysis excludes cases whose outcomes (posttest math scores) were classified as outliers by the authors. Therefore, the WWC separately calculated program effects using the entire sample with valid outcomes and found no statistically significant or substantively important difference between students in the *DreamBox Learning* and comparison groups on the MAP overall math score. The WWC characterizes these study findings as an indeterminate effect.

Thus, for the mathematics achievement domain, one study showed an indeterminate effect. This results in a rating of no discernible effects, with a small extent of evidence.

Table 3. Rating of effectiveness and extent of evidence for the mathematics achievement domain

Rating of effectiveness	Criteria met
No discernible effects <i>No affirmative evidence of effects.</i>	In the one study that reported findings, the estimated impact of the intervention on outcomes in the <i>mathematics achievement</i> domain was neither statistically significant nor substantively important.
Extent of evidence	Criteria met
Small	One study that included 557 students in three schools reported evidence of effectiveness in the <i>mathematics achievement</i> domain.

References

Study that meets WWC evidence standards without reservations

Wang, H., & Woodworth, K. (2011). *Evaluation of Rocketship Education's use of DreamBox Learning's online mathematics program*. Menlo Park, CA: SRI International. Retrieved from <http://www.dreambox.com>

Additional source:

Wang, H., & Woodworth, K. (2011). *A randomized controlled trial of two online mathematics curricula*. Evanston, IL: Society for Research on Educational Effectiveness.

Studies that meet WWC evidence standards with reservations

None.

Studies that do not meet WWC evidence standards

None.

Studies that are ineligible for review using the Elementary School Mathematics Evidence Review Protocol

DreamBox Learning, Inc. (2012). *Case study: Accelerating learning: Achieving a 5.5 percentile point gain in 16 weeks*. Bellevue, WA: Author. Retrieved from: <http://www.dreambox.com> The study is ineligible for review because it does not use a comparison group design or a single-case design.

DreamBox Learning, Inc. (2012). *Case study: Building number sense: Tailoring math instruction to individual student needs*. Bellevue, WA: Author. Retrieved from: <http://www.dreambox.com> The study is ineligible for review because it does not use a comparison group design or a single-case design.

DreamBox Learning, Inc. (2012). *Case study: Embracing blended learning: Providing a world-class education to low-income minority students*. Bellevue, WA: Author. Retrieved from: <http://www.dreambox.com> The study is ineligible for review because it does not use a comparison group design or a single-case design.

DreamBox Learning, Inc. (2012). *Case study: From good to great: Top performing school district boosts math scores*. Bellevue, WA: Author. Retrieved from: <http://www.dreambox.com> The study is ineligible for review because it does not use a comparison group design or a single-case design.

DreamBox Learning, Inc. (2012). *Case study: Impactful math intervention: From the 1st national percentile to the 54th in one year*. Bellevue, WA: Author. Retrieved from: <http://www.dreambox.com> The study is ineligible for review because it does not use a comparison group design or a single-case design.

DreamBox Learning, Inc. (2012). *Case study: Making AYP in math: A Title 1 school transformation*. Bellevue, WA: Author. Retrieved from: <http://www.dreambox.com> The study is ineligible for review because it does not use a comparison group design or a single-case design.

DreamBox Learning, Inc. (2012). *Case study: Managing budget cuts: Cost-effective instruction, assessment and math intervention*. Bellevue, WA: Author. Retrieved from: <http://www.dreambox.com> The study is ineligible for review because it does not use a comparison group design or a single-case design.

DreamBox Learning, Inc. (2012). *Case study: Personalizing instruction: Providing individualized learning paths for 3,300 students*. Bellevue, WA: Author. Retrieved from: <http://www.dreambox.com> The study is ineligible for review because it does not use a comparison group design or a single-case design.

Jorgensen, M. (2011). *Results from DreamBox Learning embedded assessment study: Demonstrates 50% increase in student proficiency in math*. Bellevue, WA: Author. Retrieved from: <http://www.dreambox.com> The study is ineligible for review because it does not use a comparison group design or a single-case design.

Jorgensen, M. (2011). *Results from DreamBox Learning grade 2 assessment study: Math achievement test demonstrates 19% increase*. Bellevue, WA: Author. Retrieved from: <http://www.dreambox.com> The study is ineligible for review because it does not use a comparison group design or a single-case design.

Appendix A: Research details for Wang and Woodworth, 2011

Wang, H., & Woodworth, K. (2011). *Evaluation of Rocketship Education’s use of DreamBox Learning’s online mathematics program*. Menlo Park, CA: SRI International. Retrieved from <http://www.dreambox.com>

Table A. Summary of findings **Meets WWC evidence standards without reservations**

Outcome domain	Sample size	Study findings	
		Average improvement index (percentile points)	Statistically significant
Mathematics achievement	557 students	+4	No

Setting The study was conducted in three Rocketship Education charter schools located in San Jose, California.

Study sample The study sample included all kindergarten and first-grade students at the three schools that participated in the study, a total of 557 students after attrition from a sample of 583 who were randomly assigned. The number of classrooms included in the study is not specified. Within grade levels, students were randomly assigned to either the intervention or comparison groups at a 4 to 1 ratio. In the baseline sample, 53% of students were female, 87% were Hispanic, 81% were English language learners, 88% were eligible for free or reduced-price meals, 4% were classified as special education, and 10% participated in Response to Intervention (RtI) services.

Intervention group The experiment was conducted from mid-October through mid-February during the 2010–11 school year. Intervention students were scheduled to receive 20 to 40 minutes of *DreamBox Learning* mathematics instruction per day; usage statistics show that students averaged 21.8 hours of usage over the course of the study, or approximately 16 minutes per day. Instructional sessions were conducted in a computer lab. The authors noted that the low-achieving students who were assigned to receive RtI services were scheduled to receive 45 minutes of *DreamBox Learning* instruction in their after-school RtI programming, regardless of intervention status. For the 42 intervention group students who were assigned to RtI services, this 45 minutes was in addition to the *DreamBox Learning* instruction provided during the school day, for a total of 26.5 hours of usage over the course of the study on average. Progress and use information provided by the *DreamBox Learning* software was not used to modify face-to-face mathematics instruction for either the intervention or comparison group.

Comparison group Students in the comparison condition received no additional mathematics instruction. However, they received additional literacy instruction via an online program during the time and in the same location as intervention group students using the *DreamBox Learning* software. The 11 students in the comparison condition who were assigned to RtI services were scheduled to receive 45 minutes of *DreamBox Learning* instruction in their after-school RtI programming; the authors found that these comparison condition students averaged 5.1 hours of program usage over the course of the study.

Outcomes and measurement

The study used math test scores from the MAP assessment developed by the Northwest Evaluation Association (NWEA). The study reports the overall math score, as well as five subtest scores, for problem solving, number sense, computation, measurement and geometry, and statistics and probability. Scores were scaled using the RIT scale, “which is scaled using the Item Response Theory (IRT) and has the same meaning regardless of the grade of the student” (as cited in Wang & Woodworth, 2011, p. 3). The schools administered the assessment in September 2010 (pretest) and January/February 2011 (posttest). For a more detailed description of this outcome measure, see Appendix B.

Support for implementation

DreamBox Learning “does not prescribe a specific role for teachers” (Wang & Woodworth, 2011, p. 3). The computer labs in which students received *DreamBox Learning* instruction were run by lab coordinators, noncredentialed hourly staff who played a minimal role in instruction. The authors noted that lab coordinators sometimes may have been out of the computer lab, at which times the students would be supervised by support staff.

Appendix B: Outcome measures for the mathematics achievement domain

Mathematics achievement

Northwest Evaluation Association (NWEA) Measures of Academic Progress (MAP): Overall math score

The mathematics test in the MAP assessment developed by the NWEA includes five subtests: problem solving, number sense, computation, measurement and geometry, and statistics and probability. Scores were scaled using the RIT scale, “which is scaled using the Item Response Theory (IRT) and has the same meaning regardless of the grade of the student” (as cited in Wang & Woodworth, 2011, p. 3).

Appendix C: Findings included in the rating for the mathematics achievement domain

Outcome measure	Study sample	Sample size	Mean (standard deviation)		WWC calculations			p-value
			Intervention group	Comparison group	Mean difference	Effect size	Improvement index	
Wang & Woodworth, 2011^a								
MAP: Overall math score	Grades K and 1	557 students	159.00 (16.60)	156.20 (15.10)	1.50	0.09	+4	0.39
Domain average for mathematics achievement (Wang & Woodworth, 2011)						0.09	+4	Not statistically significant
Domain average for mathematics achievement across all studies						0.09	+4	Not statistically significant

Table Notes: For mean difference, effect size, and improvement index values reported in the table, a positive number favors the intervention group and a negative number favors the comparison group. The effect size is a standardized measure of the effect of an intervention on student outcomes, representing the average change expected for all students who are given the intervention (measured in standard deviations of the outcome measure). The improvement index is an alternate presentation of the effect size, reflecting the change in an average student’s percentile rank that can be expected if the student is given the intervention. The statistical significance of the study’s domain average was determined by the WWC. MAP = Measures of Academic Progress.

^a For Wang & Woodworth (2011), the WWC calculated the intervention group mean using a difference-in-differences approach (see WWC Handbook) by adding the impact of the program (i.e., difference in mean gains between the intervention and comparison groups) to the unadjusted comparison group posttest means. Please see the WWC Procedures and Standards Handbook version 2.1 for more information. The authors’ primary analysis utilizes an Ordinary Least Squares (OLS) approach that excludes cases whose posttest scores were deemed to be outliers by the authors. Therefore, the WWC calculated program impacts using unadjusted means and standard deviations reported by the authors for the entire sample. p-values for this analysis sample were not reported by the authors. This study is characterized as having an indeterminate effect because the effect is neither statistically significant nor substantively important. For more information, please refer to the WWC Standards and Procedures Handbook, version 2.1, page 96.

Appendix D: Summary of supplemental subscale and subgroup findings for the mathematics achievement domain

Outcome measure	Study sample	Sample size	Mean (standard deviation)		WWC calculations			p-value
			Intervention group	Comparison group	Mean difference	Effect size	Improvement index	
Wang & Woodworth, 2011^a								
<i>MAP: Problem Solving subtest</i>	Grades K and 1	553 students	161.40 (16.30)	159.80 (15.20)	-0.70	-0.04	-2	0.68
<i>MAP: Number Sense subtest</i>	Grades K and 1	553 students	159.60 (18.90)	157.00 (17.20)	-0.90	-0.05	-2	0.65
<i>MAP: Computation subtest</i>	Grades K and 1	546 students	163.00 (20.70)	158.80 (19.50)	3.70	0.18	+7	0.09
<i>MAP: Measurement and Geometry subtest</i>	Grades K and 1	550 students	155.50 (18.30)	151.80 (18.10)	4.00	0.22	+9	0.04
<i>MAP: Statistics and Probability subtest</i>	Grades K and 1	552 students	156.30 (18.90)	154.10 (17.60)	1.80	0.10	+4	0.37

Table Notes: The supplemental findings presented in this table are additional findings from the studies in this report that do not factor into the determination of the intervention rating. For mean difference, effect size, and improvement index values reported in the table, a positive number favors the intervention group and a negative number favors the comparison group. The effect size is a standardized measure of the effect of an intervention on student outcomes, representing the average change expected for all students who are given the intervention (measured in standard deviations of the outcome measure). The improvement index is an alternate presentation of the effect size, reflecting the change in an average student’s percentile rank that can be expected if the student is given the intervention.

^a For Wang & Woodworth (2011), the WWC calculated the intervention group mean using a difference-in-differences approach by adding the impact of the program (i.e., difference in mean gains between the intervention and comparison groups) to the unadjusted comparison group posttest means. Please see the WWC Procedures and Standards Handbook version 2.1 for more information. The authors’ primary analysis utilizes an Ordinary Least Squares (OLS) approach that excludes cases whose posttest scores were deemed to be outliers by the authors. Therefore, the WWC calculated program impacts using unadjusted means and standard deviations reported by the authors for the entire sample. p-values for this analysis sample were not reported by the authors. A correction for multiple comparisons was needed and resulted in a WWC-computed critical p-value of 0.01 for the *MAP: Measurement and Geometry* subtest; therefore, the WWC does not find the result to be statistically significant.

Endnotes

¹ The descriptive information for this program was obtained from publicly available sources: the program's website (<http://www.dreambox.com>, downloaded December 14, 2012) and Wang and Woodworth (2011). The WWC requests developers review the program description sections for accuracy from their perspective. The program description was provided to the developer in December 2012, and the WWC incorporated feedback from the developer. Further verification of the accuracy of the descriptive information for this program is beyond the scope of this review. The literature search reflects documents publicly available by December 2012.

² The studies in this report were reviewed using the Evidence Standards from the WWC Procedures and Standards Handbook (version 2.1) along with those described in the Elementary School Mathematics review protocol (version 2.0). The evidence presented in this report is based on available research. Findings and conclusions may change as new research becomes available.

³ For criteria used in the determination of the rating of effectiveness and extent of evidence, see the WWC Rating Criteria on p. 12. These improvement index numbers show the average and range of student-level improvement indices for all findings across the studies.

Recommended Citation

U.S. Department of Education, Institute of Education Sciences, What Works Clearinghouse. (2013, December). *Elementary School Mathematics intervention report: DreamBox Learning*. Retrieved from <http://whatworks.ed.gov>

WWC Rating Criteria

Criteria used to determine the rating of a study

Study rating	Criteria
Meets WWC evidence standards without reservations	A study that provides strong evidence for an intervention's effectiveness, such as a well-implemented RCT.
Meets WWC evidence standards with reservations	A study that provides weaker evidence for an intervention's effectiveness, such as a QED or an RCT with high attrition that has established equivalence of the analytic samples.

Criteria used to determine the rating of effectiveness for an intervention

Rating of effectiveness	Criteria
Positive effects	Two or more studies show statistically significant positive effects, at least one of which met WWC evidence standards for a strong design, AND No studies show statistically significant or substantively important negative effects.
Potentially positive effects	At least one study shows a statistically significant or substantively important positive effect, AND No studies show a statistically significant or substantively important negative effect AND fewer or the same number of studies show indeterminate effects than show statistically significant or substantively important positive effects.
Mixed effects	At least one study shows a statistically significant or substantively important positive effect AND at least one study shows a statistically significant or substantively important negative effect, but no more such studies than the number showing a statistically significant or substantively important positive effect, OR At least one study shows a statistically significant or substantively important effect AND more studies show an indeterminate effect than show a statistically significant or substantively important effect.
Potentially negative effects	One study shows a statistically significant or substantively important negative effect and no studies show a statistically significant or substantively important positive effect, OR Two or more studies show statistically significant or substantively important negative effects, at least one study shows a statistically significant or substantively important positive effect, and more studies show statistically significant or substantively important negative effects than show statistically significant or substantively important positive effects.
Negative effects	Two or more studies show statistically significant negative effects, at least one of which met WWC evidence standards for a strong design, AND No studies show statistically significant or substantively important positive effects.
No discernible effects	None of the studies shows a statistically significant or substantively important effect, either positive or negative.

Criteria used to determine the extent of evidence for an intervention

Extent of evidence	Criteria
Medium to large	The domain includes more than one study, AND The domain includes more than one school, AND The domain findings are based on a total sample size of at least 350 students, OR, assuming 25 students in a class, a total of at least 14 classrooms across studies.
Small	The domain includes only one study, OR The domain includes only one school, OR The domain findings are based on a total sample size of fewer than 350 students, AND, assuming 25 students in a class, a total of fewer than 14 classrooms across studies.

Glossary of Terms

Attrition	Attrition occurs when an outcome variable is not available for all participants initially assigned to the intervention and comparison groups. The WWC considers the total attrition rate and the difference in attrition rates across groups within a study.
Clustering adjustment	If intervention assignment is made at a cluster level and the analysis is conducted at the student level, the WWC will adjust the statistical significance to account for this mismatch, if necessary.
Confounding factor	A confounding factor is a component of a study that is completely aligned with one of the study conditions, making it impossible to separate how much of the observed effect was due to the intervention and how much was due to the factor.
Design	The design of a study is the method by which intervention and comparison groups were assigned.
Domain	A domain is a group of closely related outcomes.
Effect size	The effect size is a measure of the magnitude of an effect. The WWC uses a standardized measure to facilitate comparisons across studies and outcomes.
Eligibility	A study is eligible for review and inclusion in this report if it falls within the scope of the review protocol and uses either an experimental or matched comparison group design.
Equivalence	A demonstration that the analysis sample groups are similar on observed characteristics defined in the review area protocol.
Extent of evidence	An indication of how much evidence supports the findings. The criteria for the extent of evidence levels are given in the WWC Rating Criteria on p. 12.
Improvement index	Along a percentile distribution of students, the improvement index represents the gain or loss of the average student due to the intervention. As the average student starts at the 50th percentile, the measure ranges from -50 to +50.
Multiple comparison adjustment	When a study includes multiple outcomes or comparison groups, the WWC will adjust the statistical significance to account for the multiple comparisons, if necessary.
Quasi-experimental design (QED)	A quasi-experimental design (QED) is a research design in which subjects are assigned to intervention and comparison groups through a process that is not random.
Randomized controlled trial (RCT)	A randomized controlled trial (RCT) is an experiment in which investigators randomly assign eligible participants into intervention and comparison groups.
Rating of effectiveness	The WWC rates the effects of an intervention in each domain based on the quality of the research design and the magnitude, statistical significance, and consistency in findings. The criteria for the ratings of effectiveness are given in the WWC Rating Criteria on p. 12.
Single-case design	A research approach in which an outcome variable is measured repeatedly within and across different conditions that are defined by the presence or absence of an intervention.
Standard deviation	The standard deviation of a measure shows how much variation exists across observations in the sample. A low standard deviation indicates that the observations in the sample tend to be very close to the mean; a high standard deviation indicates that the observations in the sample tend to be spread out over a large range of values.
Statistical significance	Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups. The WWC labels a finding statistically significant if the likelihood that the difference is due to chance is less than 5% ($p < 0.05$).
Substantively important	A substantively important finding is one that has an effect size of 0.25 or greater, regardless of statistical significance.

Please see the [WWC Procedures and Standards Handbook \(version 2.1\)](#) for additional details.