Educational Research
Institute of America
www.eriaonline.com • email: info@eriaonline.com
An affiliate of Beck Evaluation & Testing Associates, Inc.

HMH Science Dimensions™

Middle Grades, An Efficacy Study

Houghton Mifflin Harcourt

Report Number 538
April 2017

Table of Contents

Abstract

The focus of this study was the effectiveness of the *Science Dimensions* Program™ © *2018–2019.* Science Dimensions is a new K-12 program for kindergarten to grade 12 students, published by Houghton Mifflin Harcourt. The study included five teachers from four different schools across three states who agreed to a tryout of one unit of the Middle Grades level of the program. The 357 students were enrolled in either grade 7 or grade 8.

The unit chosen for the study was *Unit 1: The Dynamic Earth.* The length of the study was based on the time it would take each teacher to complete the instruction for the unit. Teachers took approximately eight weeks to complete the study beginning the end of November 2016 and ending the beginning of February 2017. Pretest and post-test assessments were developed for the unit and were administered to students before the program began and after instruction was completed.

The study was conducted with a total of 357 grade 7 and 8 students. These students were administered both a pretest and post-test designed to cover the unit of study selected for tryout analysis.

In addition to analyzing the gain scores for the total group of students, analyses were conducted separately for higher and lower scoring students. Higher and lower scoring students were identified based on the students' pretest scores. Those scoring highest on the pretests were designated as the high scoring students and those scoring lowest on the pretests were designated as the lower scoring students.

The average gain scores for the total group of 357 students were statistically significant. The effect size for the total group was large. The scores for the low and high pretest scoring groups also increased statistically significantly. The effect sizes for both high and low pretest scoring groups were above the .25 level and considered to be substantively important. For the total group as well as both the low and high pretest scoring groups, the effect sizes were large.

The study provides reliable evidence that for the unit of study used by the teachers in this study the increase in student performance was statistically significant and the effect sizes showed the increases to be large and substantively important.

Overview of the Study

Houghton Mifflin Harcourt Publishing Company contracted with Educational Research Institute of America (ERIA) to conduct a study to evaluate the effectiveness of a single instructional unit of the *Science Dimensions™ © 2018–2019* Program: Middle Grades. Five teachers agreed to try out the unit. None of the teachers had previously used the program. The unit, Unit 1: The Dynamic Earth, was used by the teachers from the end of November 2016 until the beginning of February 2017.

A different pretest/post-test pair of assessments was developed for the unit of study. The pretests were administered prior to the time the teacher began using the chosen unit and post-tests were administered after instruction for the unit was completed. Teachers took about eight weeks to complete the unit of instruction.

Research Questions

The following research questions guided the design of the study and the data analyses:

- Does the implementation of a single unit in the *Science Dimensions: Middle Grade Program* lead to improved student knowledge and understanding of the objectives of the unit of study?
- Does the implementation of a single unit in the *Science Dimensions: Middle Grade Program* lead to improved student knowledge and understanding of the objectives of the of the unit of study for higher pretest scoring students as well as for lower pretest scoring students?

Design of the Study

The design of the program called for the implementation of a single unit of the *Science Dimensions: Middle School Program* during the 2016-17 academic year. Five teachers in four different schools participated in the study.

Two of the five teachers completed a teacher survey regarding program usage. Both teachers reported using the program 5 days a week for 45 to 50 minutes per day.

Program Overview

The Science Dimensions 2018-19 program is described by the publisher as follows:

> *HMH Science Dimensions™ © 2018–2019 is a brand-new, K–12 science program built specifically to address the Three Dimensions of Science Learning outlined in the Framework for K-12 Science Education and the Performance Expectations of NGSS. Built with a digital-first mentality, this program provides an authentic approach to increasing student achievement in science and preparing teachers for engineering instruction.* HMH Science Dimensions: Engineered for the Next Generation!
>
> *HMH Science Dimensions' curriculum materials align with the National Research Council's Three Dimensions of Learning: Disciplinary Core Ideas, Crosscutting Concepts, and Science and Engineering Practices. These intertwined curriculum strands are expertly woven together into each lesson in order to meet the Performance Expectations (PEs).*

The focus of the study, Unit 1, includes 4 lessons listed below.

Unit 1: The Dynamic Earth

Lesson 1: Weathering, Erosion, and Deposition
Lesson 2: The Rock Cycle
Lesson 3: Earth's Plates
Lesson 4: Earth's Changing Surface

Description of the Assessments

The pretest and posttest used in the study were developed by ERIA curriculum experts. Tests were developed to match the content of the unit used in the study.

Table 1 provides a summary of the pretest and post-test statistics. The table shows that the reliabilities of the tests are high and provide adequate stability to assess achievement of the content of the unit.

Table 1
Pretest and Post-test Statistics

Unit 1 The Dynamic Earth	Mean Standard Score	Standard Deviation	KR 20*	SEm**
Pretest	276	44.3	.75	22
Post-test	324	43.7	.80	19

**KR 20 stands for Kuder-Richardson 20 measure of internal-test reliability*
***SEm stands for Standard Error of Measurement.*

Description of the Study Sample

Table 2 provides the demographic characteristics of the schools included in the study. It is important to note that the school data does not provide a description of the make-up of the classes that participated in the study. However, the data does provide a general description of the schools and an estimate of the make-up of the classes included in the study.

The percentage of students classified as minority students (non-Caucasian) averaged 49% and ranged from 9% to 94%. The percentage of students enrolled in free/reduced lunch programs averaged 55% and ranged from 12% to 86%.

By comparison, the National Center for Educational Statistics reports that approximately 50% of the students enrolled in U.S. public schools are classified as non-Caucasian, and the reported national average for students enrolled in free/reduced lunch programs in public schools is reported as approximately 48%.[1]

[1] *The National Center for Educational Statistics* (NCES) reported that for the 2011–2012 school year, 48.1% of public school students were enrolled in free/reduced lunch programs. No free/reduced lunch data were available for the 2012–2013 school year. Also, the NCES reported that for the 2012–2013 school year, 49.8% of public school students were classified as minority (non-Caucasian) students.

Table 2
Demographic Description of the Schools Included in the Study

	State	Location	Grades	Enrollment	% non-Caucasian	% Free/Reduced Lunch
1	CT	Suburban	6 to 8	629	12%	12%
2	WI	Urban	5 to 8	621	9%	38%
3	IL	Urban	7 to 8	726	80%	86%
4	CT	Urban	PK to 8	662	94%	85%
Averages				660	49%	55%

Data Analyses and Results

Standard scores were used for the data analyses. Raw scores were converted to standard scores with a mean of 300 and a standard deviation of 50. Data analyses and descriptive statistics were computed for the students' standard scores.

Paired comparison *t*-tests were used to determine if differences in pretest and post-test scores were significantly different. The ≤.05 level of significance was used as the level at which differences would be considered statistically significant.

In addition, effect size (Cohen's *d*) was computed for each of the comparisons. This statistic provides an indication of the strength of the effect of the treatment regardless of the statistical significance. Interpretations of effect sizes in this report include the following guidelines:

.20 to .49 = small

.50 to .79 = medium

.80+ = large

Table 3 shows that the average scores of the 357 students participating in the study increased at a statistically significant level. The effect size was substantively important and is classified as large.

Table 3
Paired Comparison *t*-test Results
Pretest/Post-test Standard Score Comparisons

	Number Students	*Mean Standard Score*	*SD*	*t*-test	*Significance*	*Effect Size*
Pretests	357	276	44.3	22.775	≤.0001	1.08
Post-tests	357	324	43.7			

The total group of 357 students was divided into two equal sized groups based on their pretest scores. The 179 students scoring lowest on the pretests were considered to be lower achievement students while the 178 scoring highest on the pretests were considered to be higher achievement students.

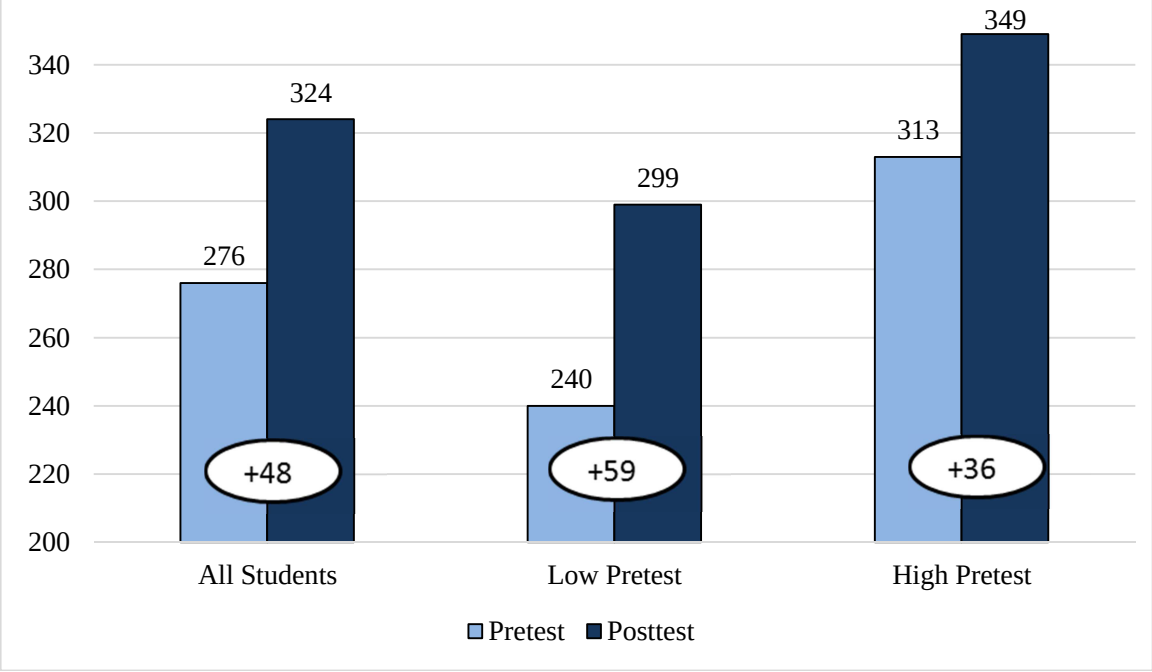
Table 4 shows that both the low pretest scoring group and the high pretest scoring group made statistically significant gains. *The effect sizes for both groups were substantively important and were classified as large for both the lower and higher pretest scoring groups.*

Table 4
Paired Comparison *t*-test Results
High- and Low-Scoring Pretest Groups

	Number of Students	*Mean Standard Score*	*SD*	*t-test*	*Significance*	*Effect Size*
Lower Scoring Group						
Pretest	179	240	25.7	17.167	≤.0001	1.61
Post-test	179	299	44.9			
Higher Scoring Group						
Pretest	178	313	24.5	17.986	≤.0001	1.48
Post-test	178	349	24.3			

Figure 1 provides a graphic representation of the gains achieved by the students. In an eight- week period, using assessments focused on just one unit of instruction, the total group of students scored 276 on the pretests and 324 on the post-tests, a gain of a little over one standard deviations. The low pretest students scored a standard score of 240 on the pretests and an average standard score of 299 on the post-tests, a gain of 59 standard score points which is about one and a half standard deviations. The high pretest students scored an average standard score of 313 on the pretests and an average standard score of 349 on the post-tests which is a gain of about one and a half standard deviations.

Figure 1
Pretest/Post-test Gain Comparison
All Students, Low Pretest Students, High Pretest Students



Conclusions

This study sought to determine the effectiveness of the *Science Dimensions: Middle Grade Program* based on a single unit of instruction. The study took place during the 2016-17 academic year and was carried out by five teachers in four schools located in three states. The student population included a similar average percentage of non-Caucasian students (49%) than the national average (50%). The average percentage of students eligible for free-reduced price lunch programs (55%) was a bit higher than the national average (48%).

Research Question 1

- Does the implementation of a single unit in the *Science Dimensions: Middle Grade Program* lead to improved student knowledge and understanding of the objectives of the unit of study?

Student achievement growth from pretesting to post-testing increased statistically significantly. The effect size was large and above a substantively important level.

Research Question 2

- Does the implementation of a single unit in the *Science Dimensions: Middle Grade Program* lead to improved student knowledge and understanding of the objectives of the unit of study for higher pretest scoring students as well as for lower pretest scoring students?

Student achievement growth for the high achieving and low achieving students increased statistically significantly. Both the high and low pretest scoring students effect sizes were above a substantively important level. For both the low and high pretest scoring groups the effect sizes were large.

For this tryout study, both research questions can be answered positively:

The *Science Dimensions: Middle Grade Program* produced statistically significant increases based on pretest/post-test scores designed to assess the students' knowledge and understanding of the program. For all comparisons, the effect sizes were large.