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

HMH Science Dimensions™ Grade 5

An Efficacy Study

Houghton Mifflin Harcourt

Report Number 540
June 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 14 teachers from seven different schools across five states who agreed to a tryout of one unit of the Grade 5 level of the program.

The unit chosen for the study was *Unit 5: Systems in Space.* 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 8 weeks to complete the study beginning the end of February 2017 and ending the beginning of June 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 420 Grade 5 students. These students were administered both a pretest and post-test. The tests were 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 420 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 size for the low pretest scoring group was large and the effect size for the higher scoring pretest group was medium. All the effect sizes were substantively important.

The study provides reliable evidence that for the unit of study used by the teachers the increase in student performance was statistically significant and the effect sizes showed the increases to be large or medium 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: Grade 5. Fourteen teachers agreed to try out the unit. None of the teachers had previously used the program. The unit, Unit 5: Systems in Space, was used by the teachers from the end of February 2017 until the beginning of June 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: Grade 5 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: Grade 5 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: Grade 5 Program* during the second semester of the 2016-17 academic year. Fourteen teachers in seven different schools participated in the study.

Eight teachers completed a teacher survey regarding program usage. The results are presented in Table 1.

Table 1
Teacher Reported Program Usage

Teacher	Days Per Week	Minutes Per Day
One	3	Fewer than 40
Two	3	40-45
Three	4	40-45
Four	4	45-50
Five	5	45-50
Six	5	45-50
Seven	5	45 50
Eight	5	50-55
Average	4.25	43.75-48.12

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 5, includes 4 lessons listed below.

Unit 5: Systems in Space

Lesson 1: How Does Gravity Affect Matter on Earth?
Lesson 2: What Daily Patterns Can Be Observed?
Lesson 3: What Patterns Can be Observed in a Year?
Lesson 4: What is the Sun?

Description of the Assessments

The pretest and post-test 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 2 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 2
Pretest and Post-test Statistics

Test	Mean Standard Score	Standard Deviation	KR 20*	SEm**
Unit 5 Systems in Space Pretest	279	47.4	.69	26
Unit 5 Systems in Space Post-test	321	43.7	.71	24

**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 3 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 15% and ranged from 9% to 38%. The percentage of students enrolled in free/reduced lunch programs averaged 41% and ranged from 28% to 58%.

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]

Table 3
Demographic Description of the Schools Included in the Study

	State	Location	Grades	Enrollment	% non-Caucasian	% Free/Reduced Lunch
1	KY	Rural	PK-6	671	12%	58%
2	WI	Small Town	5 to 8	621	9%	38%
3	ID	Small Town	3 to 6	244	11%	42%
4	ID	Small Town	PK to 6	492	9%	41%
5	ID	Small Town	K to 6	576	10%	28%
6	CA	Urban	K to 8	547	38%	NA
7	AL	Urban	PK to 12	960	15%	NA
	Average			587	15%	41%

[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.

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. Beyond the level considered to be substantively important. Interpretations of effect sizes in this report include the following guidelines:

.20 to .49 = small

.50 to .79 = medium

.80+ = large

Table 4 shows that the average scores of the 420 students participating in the study increased at a statistically significant level. The effect size was substantively important and is classified as large.

Table 4
Paired Comparison *t*-test Results
Pretest/Post-test Standard Score Comparisons

	Number Students	*Mean Standard Score*	*SD*	*t*-test	*Significance*	*Effect Size*
Pretests	420	279	47.4	19.079	≤.0001	.92
Post-tests	420	321	43.7			

The total group of 420 students was divided into two equal sized groups based on their pretest scores. The 210 students scoring lowest on the pretests were considered to be lower achievement students while the 210-scoring highest on the pretests were considered to be higher achievement students.

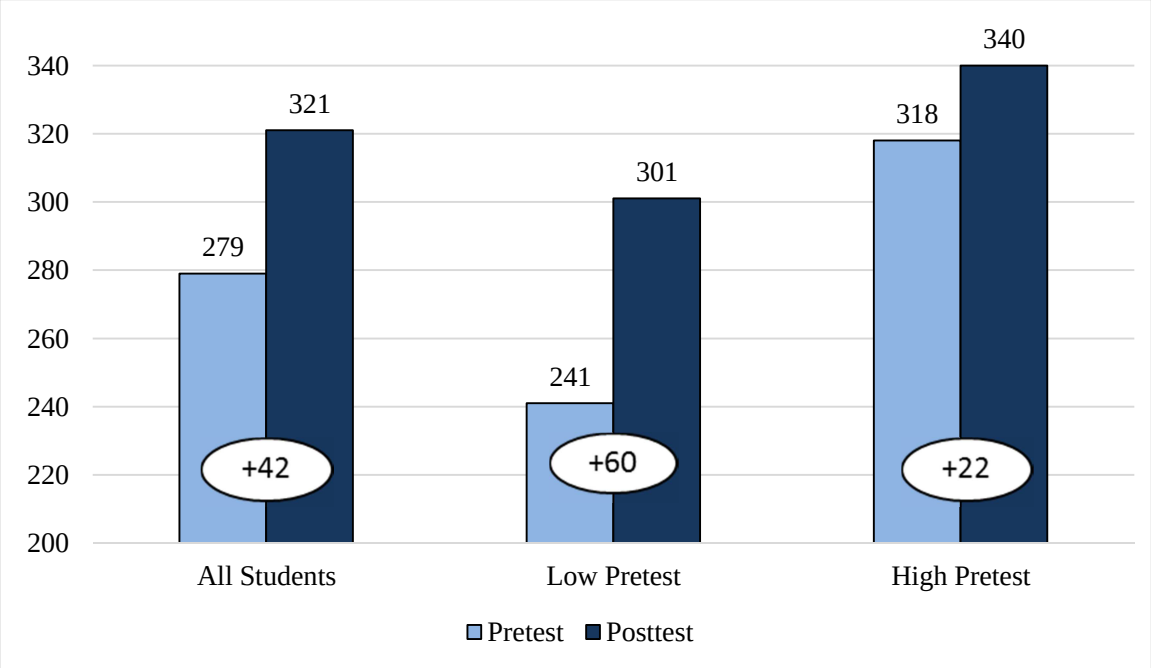
Table 5 shows that both the low pretest scoring group and the high pretest scoring group made statistically significant gains. *The effect sizes for both groups were substantively important and were classified as large for both the lower and higher pretest scoring groups.*

Table 5
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	210	241	28.2	18.805	≤.0001	1.64
Post-test	210	301	43.3			
Higher Scoring Group						
Pretest	210	318	26.9	9.909	≤.0001	.72
Post-test	210	340	34.2			

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 279 on the pretests and 321 on the post-tests, a gain of a little over one half of a standard deviations. The low pretest students scored a mean standard score of 241 on the pretests and a mean standard score of 301 on the post-tests, a gain of 60 standard score points which is somewhat higher than one full standard deviations. The high pretest students scored a mean standard score of 318 on the pretests and a mean standard score of 340 on the post-tests which is a little less than one half of a standard deviation.

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: Grade 5 Program* based on a single unit of instruction. The study took place during the second semester of the 2016-17 academic year and was carried out by 14 teachers in seven schools located in five states. The student population included a smaller average percentage of non-Caucasian students (15%) than the national average (50%). The average percentage of students eligible for free-reduced price lunch programs (41%) was approximately the same as the national average (48%).

Research Question 1

- Does the implementation of a single unit in the *Science Dimensions: Grade 5 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: Grade 5 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 the low pretest scoring groups the effect size was large and for the high pretest scoring group the effect size was medium.

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

The *Science Dimensions: Grade 5 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 or medium.