Science Dimensions

Primary Grades

An Efficacy Study

Houghton Mifflin Harcourt

Table of Contents

Abstract

The focus of this study was the effectiveness of the primary grade levels of the Science Dimensions © 2018, a science program for kindergarten to grade 12 students published by Houghton Mifflin Harcourt. The study included 54 teachers and students from 11 different schools. The percentage of students in the study enrolled in the National School Lunch Programs (NSLP) is about 20% higher than the national percentage of students enrolled in those programs. The percentage of non-Caucasian students included in the study is about 15 percent lower than the national average.

Four-hundred ninety-five grade 1 students and 461 grade two students were included in the study. Only those students who took both a pretest and posttest were included in the data analyses. Teachers used the program for their science instruction for 4 or 5 days per week and an average of approximately 25 to 40 minutes per day. The program was used by the teachers for the first time. All the teachers had at least five years of teaching experience.

The study allowed for as many weeks as teachers needed to complete 2 instructional units out of the Science Dimensions grade 1 and grade 2 programs. Since science was taught for fewer days per week than subjects like reading and math, the study lasted for approximately 4 months. Each instructional unit covered one broad science topic. Pretests and posttests were developed by science/language arts curriculum specialists and were based on the program standards, unit content, and the Next Generation Science Standards (NGSS). In addition to analyzing the gain scores for the total group of students at each grade, analyses were conducted separately for higher and lower pretest scoring students.

The average gain scores for the total group of students at both grade 1 and grade 2 were statistically significant and the effect sizes were large. The average gain scores for the low and high scoring groups at each grade level were also statistically significant. The effect size for the grade 1 low scoring group was large and for the high scoring group the effect size was medium. The effect sizes for the grade 2 high and low scoring pretest groups were both large.

All of the effect sizes at both grades exceeded by a large margin the effect sizes needed to determine a substantively important increase.

Overview of the Study

Houghton Mifflin Harcourt School Publishers contracted with the Educational Research Institute of America (ERIA) to conduct a one semester study to evaluate the effectiveness of the Science Dimensions program for students at the primary grade level. This study included students in grades 1 and 2 and compared assessments administered to students mid-January 2018 to assessments administered mid-May 2018.

Research Questions

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

- Does the implementation of the *Science Dimensions Program* in grades 1 and 2 lead to improved student science achievement?

- Does the implementation of the *Science Dimensions Program* in grades 1 and 2 lead to different increases of student science achievement as a function of student ability level?

Design of the Study

The design of the program called for the implementation of the Science Dimensions program to grade 1 and grade 2 students during the second semester of the 2017–2018 academic year. The teachers reported that they had not used the program in a prior academic year.

At grade 1 the study included a total of 28 teachers from 11 schools. At grade 2 there were 26 teachers from 10 schools. On average teachers reported using the program 4 or 5 times a week with an average usage time of about 25 minutes per class. Grade 1 and 2 teachers generally taught at the same schools.

Each grade 1 and each grade 2 teacher selected two units from the program at their grade level. The two units were chosen from a total of 6 units at grade 1 and five units at grade 2. The two selected units of study were used as the content to be taught during the spring semester.

Instructional Program Overview

The instructional program is described by the publisher as follows:

> *Designed for an inspiring, high-impact K–12 learning experience, HMH Science Dimensions creates a supportive instructional path for teachers and a dynamic learning environment for students. Teachers guide students to learn through exploration, analysis, application, and explanation—in short, to think like scientists.*

The science Dimensions program for the elementary grades was further described as follows:

> *Single-grade elementary editions feature complete print and digital curricula: write-in textbooks and Interactive Online Student and Teacher Editions with built-in professional development materials.*

> *Students are natural scientists—born with a curiosity that fuels learning. Educators have the critical role of channeling this innate learning power to help students think like real scientists. With new science standards in place, this role has taken on greater importance—and more complexity.*
>
> *Designed from the ground up to address the Next Generation Science Standards NGSS), HMH Science Dimensions™ puts students in charge of their learning and enables teachers to seamlessly guide their students on this new instructional path. This unique design means better engagement, deeper understanding, and greater student achievement.*

Teachers using the Science Dimensions programs in the elementary grades generally do not follow a sequential use of the units which make up each program. Rather, the teachers choose which units they will use and follow an order that makes sense to them. For this study, teachers were asked to choose two instructional units which would serve as their second semester instructional program. Teachers may have used more than two units during the second semester, but two chosen units were the focus of the study.

Table 1 lists the instructional units for grades 1 and 2 and the number of teachers who selected the unit as one of the two units to be taught. The two units were chosen from a total of 6 units at grade 1 and 5 units at grade 2. The selected units of study were used as the content to be taught during the spring semester. The number of teachers choosing to teach chapters are listed below for the grade 1 and grade 2 programs.

Table 1
Program Units and Number of Teachers Teaching Each Unit

Unit Number	Grade 1: Units	Teachers Selected	Unit Number	Grade 2: Units	Teachers Selected
1	Engineering and Technology	0	1	Engineering and Design Process	1
2	Sound	8	2	Matter	13
3	Light	10	3	Environments for Living Things	19
4	Plant and Animal Structures	19	4	Earth's Surface	12
5	Living Things and Their Young	15	5	Changes to Earth's Surface	7
6	Objects and Patterns in the Sky	6			
Total		56	Total		52

Description of the Assessments

The pretests and posttests used in the study were developed by ERIA curriculum experts. Tests were developed to match the performance expectations and the core ideas of the program. The HMH Next Generation Science Standards (NGSS) were also used to guide the assessment development.

Table 2 provides the test statistics. The table shows that the reliabilities of the posttests provide adequate stability to assess achievement. Of importance is the fact that the test reliabilities are higher for the posttests than for the pretests. This is almost certainly the result of instruction which would result in less random guessing on the posttests than on the pretests. The reliabilities for all 4 of the tests are appropriate for assessing gain scores.

Table 2
Pretest and Posttest Statistics Grades 1 and 2

Test	*Mean Standard Scores*	*Standard Deviation*	*KR 20*	*SEm**
Grade 1 Pretest	277	53.0	.47	38.5
Grade 1 Posttest	323	33.1	.65	19.6
Grade 2 Pretest	268	46.9	.54	31.8
Grade 2 Posttest	332	27.2	.68	15.4

*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, thereby, an estimate of the make-up of the classes included in the study.

Table 3 shows that the percentage of students classified as non-Caucasian ranged from 6% to 84% with an average of 34%. By comparison, approximately 48% of the students enrolled in U.S. public schools are classified as non-Caucasian.

The percentage of students enrolled in National School Lunch Programs ranged from 39% to 100% and averaged 73% across the sample of schools. By comparison, the reported national average for students enrolled in National School Lunch Programs in public schools was reported as approximately 48%.[1]

---

1 The National Center for Educational Statistics (NCES) reported that for the 2014-2015 school year, 51.8% of public school students were enrolled in free/reduced lunch programs. Also, the NCES reported that for the 2014-2015 school year, 48% of public school students were classified as minority (non-Caucasian) students.

Table 3
Demographic Description of the Schools Included in the Study

	Location	Grades	Enrollment	Percent Enrolled	
				Non-Caucasian	NSLP*
1	Town: Distant	PK-5	502	48%	77%
2	Town: Remote	PK-2	386	84%	100%
3	Rural: Remote	PK-5	39	74%	100%
4	Rural: Fringe	PK-5	365	21%	87%
5	Town: Distant	PK-5	447	34%	65%
6	Town: Distant	PK-5	395	29%	74%
7	Town: Distant	PK-5	575	31%	71%
8	Town: Distant	PK-5	435	22%	61%
9	Town: Distant	PK-5	570	11%	39%
10	Rural: Fringe	PK-5	482	6%	60%
11	Town: Remote	K-4	285	14%	68%
	AVERAGES		407	34%	73%

*National School Lunch Program

Data Analyses and Results

Standard scores were used for all 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 using the students' standard scores.

For most of the comparisons, paired comparison *t*-tests were used to determine if differences in pretest and posttest scores were significantly different. The comparisons were conducted for differences between the *Science Dimensions January 2018* (pretest) and the *Science Dimensions May 2018* (posttest). 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

Grade 1 Results

Table 4 shows that the average scores of the 495 grade 1 students participating in the study increased at a statistically significant level. The effect size was substantively important and is classified as large.

Table 4
Grade 1 Total Group Paired Comparison *t*-test Results
Pretest/Posttest Standard Score Comparisons

	Number Students	*Mean Standard Score*	*SD*	*t*-test	*Significance*	*Effect Size*
Pretests	495	277	53.0	19.508	≤.0001	1.041
Posttests	495	323	33.1			

The total group of 495 grade 1 students was divided into two equal sized groups based on their pretest scores. The 247 students scoring lowest on the pretest were considered lower achievement students while the 248 scoring highest on the pretest scores were considered higher achievement students.

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

Table 5
Grade 1 Paired Comparison *t*-test Results
High- and Low-Scoring Pretest Groups

Test	Number of Students	Mean Standard Score	SD	t-test	Significance	Effect Size
Lower Scoring Group						
Pretest	247	234	38.8	26.706	≤.0001	2.117
Posttest	247	315	37.7			
Higher Scoring Group						
Pretest	248	319	22.7	6.074	≤.0001	0.498
Posttest	248	331	25.4			

Figure 1 provides a graphic representation of the gains achieved by the grade 1 students. In one semester the grade 1 students increased their average standard scores by 46 points. The low achieving science students increased their average standard scores by 81 points which is many times larger than the increase of the high pretest scores. However, the high pretest score group was hampered by very high pretest raw scores which limited their increase.

Figure 1
Grade 1 Pretest/Posttest Gain Comparison
All Students, Low Pretest Students, High Pretest Students
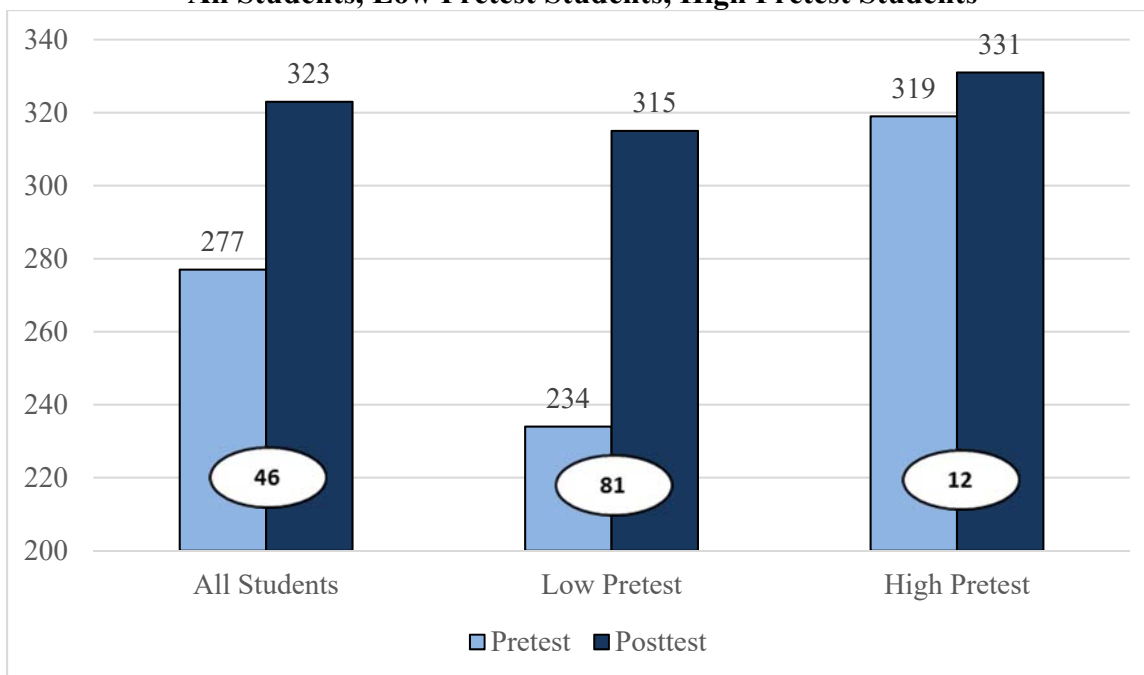
Grade 2 Results

Table 6 shows that the average scores of the 461 grade 2 students participating in the study increased scores at a statistically significant level. The effect size was substantively important and is classified as medium.

Table 6
Grade 2 Total Group Paired Comparison *t*-test Results
Pretest/Posttest Standard Score Comparisons

	Number Students	*Mean Standard Score*	*SD*	*t*-test	*Significance*	*Effect Size*
Pretests	461	268	46.9	29.346	≤.0001	1.669
Posttests	461	332	27.2			

Based on their pretest scores, the total group of 461 grade 2 students was divided into two approximately equal sized groups of 230 students for the low scoring group and 231 for the high scoring students. The students scoring lowest on the pretest were considered lower science achievement students while the students scoring highest on the pretest scores were considered higher science achievement students.

Table 7 shows that both groups made statistically significant gains. The effect sizes for both groups were substantively important, and the effect sizes for both the low group and high groups were large.

Table 7
Grade 2 Paired Comparison *t*-test Results
High- and Low-Scoring Pretest Groups

Test	*Number of Students*	*Mean Standard Score*	*SD*	*t-test*	*Significance*	*Effect Size*
Lower Scoring Group						
Pretest	230	229	29.4	35.106	≤.0001	3.151
Posttest	230	325	31.4			
Higher Scoring Group						
Pretest	231	307	21.9	19.333	≤.0001	1.519
Posttest	231	339	20.2			

Figure 2 provides a graphic representation of the gains achieved by the grade 2 students. In one semester, the grade 2 students increased their average scores by 64 standard score points. The low achieving science students increased their average scores by 96 standard score points while the high achieving science students increased their average scores 32 standard score points. As with the grade 1 students, the higher pretest group was limited in achieving a large gain because their pretest scores were quite high.

Figure 2
Grade 2 Pretest Posttest Gain Comparison
All Students, Low Pretest Students, High Pretest Students
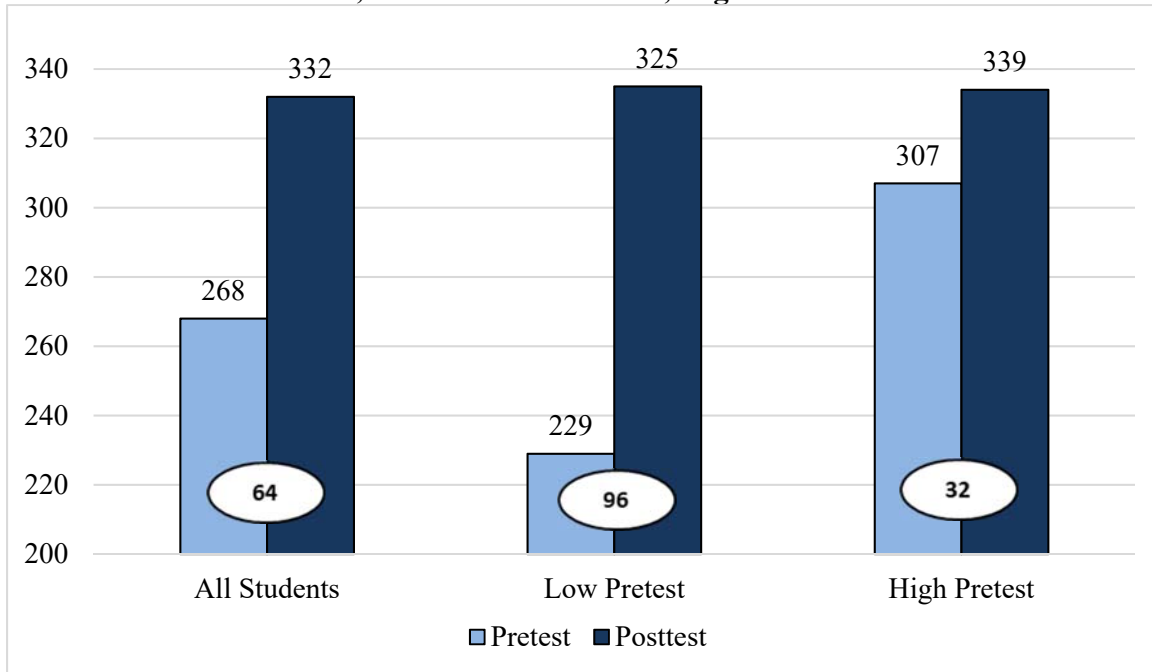
Conclusions

This study sought to determine the effectiveness of the Science Dimensions program at grades 1 and 2 by comparing growth on reliable and valid pretests and posttests. The study took place during the second semester of the 2018 academic year and was carried out in 11 different schools and 54 teachers. The student population included an average of 23% more students eligible for free-reduced price lunch programs than the national average. The percentage of non-Caucasian student was about 15% lower than the national average.

Two research questions guided the study and the conclusions for each are reported below.

Research Question 1

- Does the implementation of *Science Dimensions* in grades 1 and 2 lead to improved student science achievement?

For both grades included in the study science achievement growth from pretesting to post-testing was statistically significant. The effect sizes at both grades were above a substantively important level and were large at both grades 1 and 2.

Research Question 2

- Does the implementation of *Science Dimensions* in grades 1 and 2 lead to improved student science achievement as a function of student ability level prior to using the program?

Low pretest scoring grade 1 students increased their achievement at a statistically significant level and the effect size was large. High pretest scoring grade 1 students increased their achievement at a statistically significant level and the effect size was medium.

Low pretest scoring grade 2 students increased their achievement at a statistically significant level and the effect size was large. High pretest scoring grade 2 students increased their achievement at a statistically significant level and the effect size was medium.

Based on this study, both research questions can be answered positively:

The Science Dimensions produced statistically significant increases for students at grades 1 and 2. The effect sizes for the total group at grade 1 and grade 2 were large.

The Science Dimensions produced statistically significant growth for both higher ability and lower ability students in grades 1 and 2. For grade 1 students, the effect sizes for the higher pretest scoring student was medium, and the effect size for the lower pretest scoring students was large. For grade 2 students, the effect sizes for both the higher and lower pretest scoring students was large.

The effect sizes for both grades and both levels were above a substantively important level.