A Quasi-experimental Study of NGSS Curriculum and PD on Three-dimensional Learning Outcomes

Susan M. Kowalski, Jefferey Snowden, Lisa Carey, Betty Stennett

Lindsey Mohan and Mark Bloom

BSCS Science Learning

Background/Context

A Framework for K-12 Science Education (NRC, 2012) and the *Next Generation Science Standards* (2013) call for new curriculum materials to support student engagement with three-dimensional learning, professional development experiences to support teachers in using the materials, and assessments to measure students' mastery of these new and ambitious standards.

Curriculum designed for NGSS. Reiser (2013) articulated how phenomenon-based units designed for NGSS are fundamentally different from curriculum materials in the past. The EQuIP review (Achieve, 2015) formally defined the characteristics of curricula designed for the NGSS.

Curriculum-based PD for NGSS. A small number of organizations have designed curriculum-based PD (NGSX, OpenSciEd, American Museum of Natural History, BSCS) to support teachers as they implement curricula designed for NGSS.

Assessment. DeBarger, Penuel, Harris, and Kennedy (2015) and Harris, Krajcik, Pellegrino, and McElhaney (2016) provided guidance on NGSS assessment development, but research documents the challenges associated with developing NGSS-aligned assessments (Gorin & Mislevy, 2013; Wertheim, Osborne, Quinn, Pecheone, Schultz, Holthuis, & Martin, 2016).

Curriculum and PD work together to support teachers as they guide student three-dimensional learning. Formative and summative assessment provides the teacher feedback and allows researchers to document the efficacy of a curriculum/PD model.

To date, **there has been little or no research into the efficacy of curricula designed for the NGSS, supported by curriculum-based PD, and assessed using instruments designed to measure students' three-dimension learning outcomes**. The project and research we describe begins to fill that void.

We describe the results of a four-year project that

- developed an NGSS digital middle school body systems unit;
- developed an NGSS assessment;
- developed on online curriculum-based PD model to support curriculum enactment;
- conducted a quasi-experiment of the intervention; and
- measured pre/post changes in teacher knowledge and practice.

Purpose and Research Questions

The research goal (with parallel research questions) for this project was to determine the extent to which the package of PD and curriculum materials achieved two important outcomes:

1. Support teacher implementation of NGSS-aligned science instructional materials.
2. Enhance student three-dimensional science learning.

We hypothesized that the package of PD and curriculum materials would enhance teachers' content knowledge and practice, ultimately enhancing three-dimensional student learning.

Setting

We provided teachers with digital curriculum materials. PD was provided in an online environment with both asynchronous and synchronous components.

Participants

Teachers and their students from 14 states in urban, suburban, and rural areas participated. Table 1 provides the demographic characteristics of the students in the study.

Intervention

Curriculum design. The curriculum included four chapters related to human body systems. The anchoring phenomenon was the story of a girl who suddenly began losing weight. Students study the doctor's notes about the girl and begin an 8-week learning experience to investigate the interaction of the respiratory, nervous, muscular, and digestive systems with the circulatory system, including the development of understanding of how damage to one part of the body can lead to symptoms throughout the body. Students use scientific argumentation and modeling as they develop an understanding of the body as system of interacting sub-systems, composed of groups of cells. Achieve rated the curriculum highly.

PD design. The online PD used a video-based analysis-of-practice model known as *Science Teachers Learning from Lesson Analysis* [STeLLA] (Roth et al., 2011; Taylor et al., 2017). STeLLA has strong evidence of efficacy in enhancing teacher and student outcomes. We translated STeLLA to an online environment for this project.

As part of the online STeLLA PD modules, teachers experienced the curriculum as learners, studied teaching strategies that are part of STeLLA, reflected on the use of teaching strategies within the curriculum, and analyzed classroom video.

Teachers participated in 11 PD modules with synchronous and asynchronous components during the summer and school year. The total PD duration was 50 hours that extended over a period of 16 weeks.

Research Design

We conducted a quasi-experiment to investigate the efficacy of the intervention in supporting students' three-dimensional learning outcomes. In this two-year cohort control study, teachers used their extant materials the first year, participated in online PD in the summer between school years and into the fall of the second year, and used the intervention materials in the fall of the second year. Comparison students were taught using extant materials, and treatment students used intervention materials. The design allowed us to control for individual

teacher effects. In intervention research, teacher effects can contribute as much as 21% of the variance in student outcomes (Nye, Hedges, & Konstantopoulos, 2004).

Data Collection and Analysis

The student assessment included multiple choice and open-ended response items. Comparison and treatment students completed a pretest and posttest in the Fall of 2017 and Fall of 2018, respectively. Teachers completed a content pretest and posttest and filmed themselves teaching a lesson before and after the summer PD. We scored teacher video using a rubric designed to assess teachers' use of the STeLLA strategies (Roth & Kowalski, 2015).

We analyzed student data using three-level HLM, with students nested within classes, and classes nested within teacher. Each teacher had multiple comparison classes (class periods) in the 2017-2018 school year, and multiple treatment classes in the 2018-2019 school year. We used a random slopes model to examine variation in treatment effect by teacher.

We analyzed teacher data using matched-pairs t-test and investigated the relationship between teacher pretest and posttest using OLS regression with teacher pretest and years of experience as explanatory variables.

Findings/Results

Teachers demonstrated significant increases in content knowledge ($p < .001$, $d = .88$) and classroom instructional practice ($p < .001$, $d = 1.19$). Teacher posttest scores did not depend on teacher years of experience (Table 2).

Treatment students outperformed comparison students (Table 3). The difference was statistically significant ($p = .007$; $d = .280$). The 95% confidence interval of the effect size is large [-0.216, 0.696] indicating that the effects were variable across teachers.

Conclusions

We find that when teachers use curriculum designed for the NGSS and are supported by extensive PD in an online environment, students perform better on assessments designed to measure ambitious three-dimensional learning goals.

References

DeBarger, A. H., Penuel, W. R., Harris, C. J., & Kennedy, C. A. (2016). Building an assessment argument to design and use next generation science assessments in efficacy studies of curriculum interventions. *American Journal of Evaluation*, *37*(2), 174-192.

Gorin, J. S., & Mislevy, R. J. (2013, September). Inherent measurement challenges in the next generation science standards for both formative and summative assessment. In *Invitational research symposium on science assessment*. Educational Testing Service.

Harris, C. J., Krajcik, J. S., Pellegrino, J. W., & McElhaney, K.W. (2016). *Constructing assessment tasks that blend disciplinary core Ideas, crosscutting concepts, and science practices for classroom formative applications.* Menlo Park, CA: SRI International.

National Research Council. (2012). *A framework for K-12 science education: Practices, crosscutting concepts, and core ideas. Committee on a Conceptual Framework for New K-12 Science Education Standards.* Board on Science Education, Division of Behavioral and Social Sciences and Education. Washington, DC: National Academy Press.

NGSS Lead States. (2013). *Next Generation Science Standards: For states, by states*. Washington, DC: The National Academies Press.

Nye, B., Hedges, L.V. & Konstantopoulos, S. (2004). How large are teacher effects? *Educational Evaluation and Policy Analysis, 26*(3), 237–57.

Reiser, B. (2013). *What Professional Development Strategies Are Needed for Successful Implementation of the Next Generation Science Standards?* Paper presented at the Invitational Research Symposium on Science Assessment.

Roth, K. J., Garnier, H., Chen, C., Lemmens, M., Schwille, K., & Wickler, N. I. Z. (2011). Videobased lesson analysis: Effective science PD for teacher and student learning. Journal of Research in Science Teaching, 48(2), 117-148.

Roth, K. J., & Kowalski, S.M. (2015). *STeLLA II Lesson video coding manual*. Colorado Springs, CO: Biological Sciences Curriculum Study (BSCS).

Taylor, J. A., Getty, S. R., Kowalski, S. M., Wilson, C. D., Carlson, J., & Van Scotter, P. (in press). *An Efficacy Trial of Research-Based Curriculum Materials with Curriculum-Based Professional Development*. Manuscript submitted for publication.s

Wertheim, J., Osborne, J., Quinn, H., Pecheone, R., Schultz, S., Holthuis, N., & Martin, P. (2016). An analysis of existing science assessments and the implications for developing assessment tasks for the NGSS.

Table 1.

Descriptive Statistics of the Sample: Means and Sample Sizes by Treatment Group.

Variables	Total sample (N = 3,244)	Treatment condition (N = 1,592)	Comparison condition (N = 1,652)
Students			
Mean pretest score person measure (SD)	-1.722 (1.020)	-1.742 (0.985)	-1.686 (1.040)
Mean posttest score person measure (SD)	-1.363 (1.081)	-1.365 (1.105)	-1.351 (1.045)
FRL (%)	37	39	36
Female (%)	48	47	49
Underrepresented minority (%)	48	48	48
Asian (%)	13	14	11
White (%)	75	73	77
African American (%)	14	14	14
Native Hawaiian or Pacific Islander (%)	7	8	6
American Indian or Alaska Native (%)	12	13	11
Hispanic or Latino/a (%)	21	22	20
Other Race or Ethnicity (%)	17	16	18
Grade 6 (%)	11	4	18
Grade 7 (%)	64	70	58
Grade 8 (%)	23	23	22
Grade 9 (%)	2	3	2
English learner (%)	5	7	3

Table 2.

Pretest-posttest Changes for Teacher Outcomes.

Outcome	N	Pre mean (SD)	Post mean (SD)	Effect size (d)	95% CI of the effect size	
					Lower	Upper
CK: Overall (person measure)	18	12.11 (2.74)	14.5 (2.68)	0.88	0.54	1.22
Classroom Practice (person measure)	18	35.44 (20.06)	63.94 (27.14)	1.19	0.86	1.52

Note: CI = confidence interval; CK = content knowledge; effect sizes are standardized mean difference effect sizes.

Table 3.

Test of Main Effect of Treatment on Student Achievement.

Variable	Coefficient	Standard error	t-ratio	d.f.	p-value
Level 3 (teacher)					
Intercept	16.800	0.427	39.375	28	< .001
MnPre	2.035	0.805	2.529	28	0.017
Level 2 (class)					
γ_{010} (avg. treatment effect across teachers)	1.454	0.502	2.899	29	0.007
Level 1 (student)					
Pre	0.653	0.036	18.090	2,033	< .001