Spiraling" Teacher Learning: A novel PLC platform integrating multi-grade collaboration, classroom artifacts, and mobile technology.

Marlene Saint Martin (UCLA), Jose Felipe Martinez (UCLA), Matthew J. Kloser (University of Notre Dame), Michael Szopiak (University of Notre Dame), Jon Kovach (UCLA)

Abstract

We investigate how collaborative structures within and across grade levels can influence teachers' understanding of student learning trajectories across K-8 science and how teachers align their instruction with the NGSS using vertically aligned and culturally relevant storylines of natural phenomena. The Professional Learning Community (PLC) model also relies on software tools to collect classroom artifacts reflecting instruction and student thinking. We analyze how participation in this type of PLC structure influences teacher planning, practice, and self-efficacy by comparing teacher and student survey responses before and after participation. We find promising evidence of statistically significant increases in teachers' self-efficacy, the relevance of collaboration between teachers across grade levels, and students' exposure to science practices.

Introduction

The Next Generation Science Standards (NGSS) serve as a comprehensive framework for K-12 science education that guides teachers and students through crosscutting concepts, practices, and core ideas in increasing depth and richness across non-consecutive grades. The effective implementation of this modern framework requires teachers to (1) interact and collaborate within and across grades to observe and study how their students' learning trajectories and scientific thinking evolve and (2) align their instruction with the NGSS to shape and guide this evolution. Unfortunately, K-12 science teachers rarely have opportunities, time, and tools to engage in such activities efficiently.

The SPIRAL Project aims to address these critical needs by providing (1) a Professional Learning Community (PLC) structure that involves within (horizontal) and across (vertical) grade-level teacher collaboration, (2) resources in the form of vertically and NGSS-aligned storylines of natural phenomena (e.g., Water, Waves) that incorporate culturally relevant and locally grounded concepts that resonate with students' background knowledge and experiences, and (3) digital portfolio tools specifically designed to allow teachers to capture, organize, and share multimedia classroom artifacts. These artifacts, which showcase evidence of student thinking and engagement in science practices along the natural phenomena storylines, offer a window into each teacher's science classroom and provide an anchor for reflection and collaboration.

During the 2021-22 and 2022-23 academic years, a cohort of 22 science teachers in a California school district participated in this project, which was designed to examine how the collaborative structures, resources, and tools described above enable teachers to better understand students' science learning trajectories and shape their instructional practice to improve student experiences and learning of curricular content. In this study, we investigate at an exploratory level how participation in this professional development experience may have influenced teachers' planning, instruction, and self-efficacy by comparing these teachers' and their students' responses to survey items before and after participation. Specifically, our guiding research questions are to investigate whether there is evidence of pre-post changes in terms of:

1. Teacher self-efficacy around *target* classroom practices (e.g., developing scientific and

engineering practices; supporting students' learning trajectories).

2. Teachers' planning and delivery of science lessons (e.g., alignment to NGSS standards, collaboration between teachers across grade levels, use of artifacts of student thinking).
3. Students' range of exposure to target instructional practices aligned to the NGSS (e.g., asking questions and defining problems, developing and using models, planning and carrying out investigations).

Literature review

This study is rooted in theoretical and empirical evidence highlighting the importance of coherent student and teacher learning across grades. While the NGSS outline a spiraled trajectory of core ideas, concepts, and science practices, they provide little guidance on how classroom instruction can best support these standards and what student thinking looks like across these development bands (Krajcik et al., 2014). The notion that independent decision-making in the classroom can sufficiently support professional reflection around teaching is increasingly in conflict with adult learning theory (Trotter, 2006) and a growing consensus across subjects and disciplines that teaching core ideas and practices requires vertical coherence across grade levels (Pierson et al., 2019).

We additionally draw on evidence of the value of *classroom artifacts* reflecting direct evidence of instruction and student thinking for supporting teacher learning and development (Darling-Hammond et al., 2017). Portfolios that compile these artifacts can support reliable measures of instruction (Martinez et al., 2021) and richly portray teaching and learning. Furthermore, portfolios also represent powerful conduits for professional reflection, providing concrete examples of practice and student thinking to anchor reflection and abstract understanding of teaching and learning (Shulman, 1998; Kloser et al., 2020). Importantly, portfolio collection does not automatically lead to learning unless situated within structures that provide quality reflection and feedback –PLCs offer one such structure with collaborative discussion over multiple encounters focused on "critically interrogating [practice]" (Stoll et al., 2006, p. 229). Electronic portfolios and mobile technology can support these types of interaction among teachers that may otherwise be difficult due to organizational or logistical constraints. However, there are very few examples of active use of e-portfolios to foster reflection and learning in the field.

Furthermore, the PLC model explicitly builds on culturally responsive, community-grounded, and asset-based content and pedagogical approaches, which understand and acknowledge the role of culture in teaching and learning and privilege asset-based perspectives in developing student sense-making around science content (Brown et al., 2018). Under this framework and hypothesizing that high-quality NGSS teaching is not limited to knowledge and practice within one's classroom or grade level, our PLC's horizontal and vertical structures connect classroom communities, leading to more coherent quality science instruction and fostering *vertical* sense-making—revising conceptions of effective instructional practice and student learning trajectories.

Methods and Procedures

The SPIRAL Project

A multidisciplinary team of education researchers, science educators, and software developers collaborated to design, develop, and assemble the SPIRAL Project. The UCLA Science Project at Center X crafted two storylines with science teachers in 2nd, 5th, and 6th grade teaching the Water phenomena storyline and science teachers in 1st, 4th, and 8th grade teaching the Waves phenomena storyline. These storylines represent 12-16 hours of science instruction that participating teachers taught between February and April of each school year (2021-22 and 2022-23).[1]

The UCLA Science Project also established and facilitated two Professional Learning Communities (PLCs), one for each teacher storyline cohort, that met regularly throughout each academic year, amounting to approximately 66 hours of learning and collaboration within this structure. Each school year, in its first five sessions (August-February), the PLC focused on introducing, adapting, and learning each grade-level storyline and familiarizing teachers with the digital tools to capture and upload student artifacts.[2] In these sessions, teachers collaborated horizontally and vertically, mapping the big ideas and potential artifacts to collect, incorporating formative assessments, and discussing how their instruction connects and adds to students' learning trajectories. In the last two sessions of the PLC (April-May), once teachers had already taught the Water and Waves storylines, teachers contributed some of their collected artifacts to a shared digital portfolio. With these artifacts at the center of reflection, teachers discussed (1) how and in what ways their students' knowledge and understanding had changed throughout the storylines, (2) whether their students had taken in the big ideas and what were common misconceptions, and (3) what can students do at the start and end of the storylines across grades.

As mentioned, teachers used digital tools throughout this process to collect, annotate, organize, and share artifacts of student thinking in digital portfolios. These tools include the

[1] Teachers chose how to accommodate and teach the storylines in their classrooms between February and April of each academic year.
[2] In this first phase of the project, teachers met 3 times in person for a full day and twice virtually for two hours. In the 2021-2022 academic year, however, due to COVID-19, teachers only met 2 times in person.

SPIRAL Notebook Project iOS and Android apps and a web portal.[3] This software underwent extensive testing during the first school year (2021-2022), during which we gathered teacher feedback about usability and necessary improvements. At their core, these tools allow educators to upload and annotate artifacts into their portfolios and contribute, react, and comment on shared portfolios with the app, which sends reminders and notifications of new activity.

Participants

Participants comprise a cohort of 22 teachers in a small and ethnically diverse school district in southern California, teaching in grades K-8 during the 2021-22 and 2022-23 school years.[4][5] The district agreed to participate based on its previous collaboration with the research team and its interest in opportunities to support its teachers' science instruction and curriculum. We recruited teachers via pamphlets distributed by the district, describing the project, requirements, and benefits of participation. While participation was voluntary, the district counted it towards contract hours of Professional Development, and teachers received a monetary stipend and an iPad tablet from the project for their work outside of those meetings.

Table 1 presents the distribution of teachers by grade level and storyline, with 13 and 9 teachers initially learning and teaching the Water and Waves storylines, respectively.[6] Participation required teachers to (1) attend the PLC meetings (30-36 hours each school year,

[3] The SPIRAL apps were released to the iOS and Android stores in the fall of 2021.
[4] The district serves a student population that is almost 90% Hispanic or Latino.
[5] The initial 2021-22 cohort comprised 26 teachers, but changes within the district resulted in four teachers not continuing for the second year.
[6] The distribution of teachers by grade and storyline is not a result of pre-defined sample size calculations but a direct result of the district's available number of teachers by grade level and teachers' interest in the project. Furthermore, two 4th-grade teachers moved grade levels at the start of the 2022-2023 school year. One moved to 2nd grade, and the other moved to 5th grade.

including full-day in-person sessions and Zoom meetings), (2) teach the Water or Waves storylines, and (3) use the digital portfolio tools to capture, upload, annotate, share, and discuss artifacts of student thinking. During the project's first year (2021-22), two in-person meetings were postponed due to the rising number of COVID-19 cases in the district, while COVID-19-related absences and student learning gaps presented challenges for some teachers in teaching or finishing the storyline, as other subjects were often prioritized over science. The PLC structure was fully implemented during the project's second year (2022-23). Table 2 describes our teachers' sociodemographic information, including sex, years of teaching science experience, and race and ethnicity.

Data Instruments and relevant variables

We collected several types of data, including (1) pre-post teacher and student surveys, (2) pre-post teacher interviews, (3) video recordings of teachers' collaborations and discussions during the PLC sessions, (4) pre-post grids identifying core ideas and topics across grades, (5) artifacts captured by the teachers in their digital portfolios, and (6) record of feedback and interactions among PLC members through the portfolio platform.

This paper focuses on data from the surveys. All teachers completed the baseline survey at the start of the 2021-2022 academic year (N=26) and the follow-up at the end of the 2022-2023 academic year (N=22). Teachers rated (1) the influence of various elements in planning and delivering science instruction (4-point Likert scale from "Not important" to "Extremely important"); (2) their students' exposure to science teaching practices (5-point Likert scale from "No exposure" to "In-depth exposure"); and (3) their self-efficacy on various teaching practices aligned to the NGSS (4-point Likert scale from "Not prepared" to "Very well prepared").

Student surveys were collected in participating teachers' upper elementary (4th and 5th grade) and middle school (6th and 8th grade) classrooms during the 2022-23 school year.[7] Participating teachers shared a link to the anonymous survey with 324 students whose parents had signed a consent form (25 on average per teacher). Students completed the baseline survey during the Fall 2022 semester (N=251, 77.5% response rate) and the follow-up in the Spring 2023 semester (N=267; 82.4% response rate). Table 3 describes the sociodemographic characteristics of the sample of surveyed students, which reflects the diversity of the district student population. Students rated (1) their exposure to NGSS practices (5-point Likert scale from "I completely disagree, we never did that" to "I completely agree, we did that often"); (2) their exposure to quality assessment practices[8] (4-point Likert scale from "I completely disagree, we never did that" to "I completely agree, we did that often"); and (3) their feelings and perceptions (identity, interest, cost, value) about science (4-point Likert scale from "No, I strongly disagree" to "Yes, I strongly agree").

Analysis

One limitation of our study is that we cannot attribute our results directly to the project, as our research design lacks a counterfactual to isolate the effects of participation. As such, our findings must be understood, at this point, as exploratory, providing relevant signals of what aspects of science instruction the program may help strengthen and improve. Analyzing outcomes such as students' exposure to science practices might be better supported by

---

[7] As the items inquired about complex aspects of classroom instruction, we only surveyed students in upper and middle school grades.
[8] The quality assessment practices are based on Martinez et al. 2012.

examining the student artifacts or conducting classroom observations. As we have a collection of artifacts from each teacher in our cohort, we do not discard the possibility of using these to support our findings in a future study. It is also worth noting that while the grain size of the storylines amounts to 12-16 hours of the science curriculum, the project comprises other resources, such as the PLC's support, that round off our cohort's professional development beyond the storylines and their instruction.

We compare teachers' baseline (Fall 2021) and end-of-year responses (Spring 2023) to explore how participation might have influenced teachers' science instruction (N=22). We also compare students' baseline (Fall 2022) and end-of-year survey responses (Fall 2023) to complement teachers' perceptions about students' exposure to science practices. It is important to note that students' surveys were anonymous to encourage participation and reinforce our message that their responses would not be used to evaluate their teachers or affect their grades. As such, while we can identify each student's school and grade level, we cannot associate their responses with their teacher or match their baseline and end-of-year responses.

For the teacher pre-post comparisons, we use within-subject t-tests, which calculate the difference between each participant's responses, examining whether the average difference in the respective Likert-scale is statistically different from zero at the 5 and 1 percent levels. As students are not identified across pre-post surveys, we conducted between-groups t-tests, calculating and comparing the overall mean group difference across surveys, examining whether this difference is statistically different from zero at the 5 and 1 percent levels. While it is convenient to run a parametric test to analyze mean scores, as the Likert scales are ordinal, the sample size is small, and we examine each item individually, the normality assumption framing this analysis might be violated (Sullivan & Artino, 2013). As such, we also conduct a non-parametric test for paired and ordinal data, the Wilcoxon Sign Test, for teacher pre-post comparisons and a non-parametric test for independent and ordinal data, the Mann-Whitney U test, for student pre-post comparisons. As our teacher sample size is small (N=22), we do not disaggregate these comparisons by phenomena (Water or Waves cohort) or grade level in this paper; we leave these comparisons for future cohorts that include more teachers.

Results

Teachers' self-efficacy

Our first research question aims to answer whether and in what ways their participation may influence teachers' self-efficacy. To answer this question, we compare teachers' pre-post responses to how prepared they feel to mentor or lead conversations with other teachers regarding 15 classroom practices intended to support science learning goals. Teachers rated these items using a four-point Likert scale (from "Not prepared" to "Very well prepared"). In the baseline, on average, the highest-rated item on which teachers felt most prepared was "Eliciting student thinking through multiple forms and media." However, this average score was below the "Somewhat prepared" threshold. Furthermore, none of the other 14 items received an average score between the "Somewhat prepared" and "Very well prepared" thresholds, indicating that, at baseline, teachers felt unprepared regarding all these classroom practices.

Table 4 examines whether the pre-post average differences are statistically different from zero. Some of the items where we especially expect or would like to observe a positive trend given the project's emphasis are (1) supporting students' development of models and explanation of phenomena, (2) supporting students' learning trajectories, and (3) developing students' scientific and engineering practices. We expect positive changes in these items as all teachers collected and reflected on artifacts that showed students' initial and final models of the Water or Waves phenomena; teachers collaborated across grade levels to map the big ideas for each grade; and, besides modeling, the project emphasized other scientific and engineering practices (e.g., asking questions and defining problems, engaging in argument from evidence, constructing explanations and designing solutions).

In contrast to the baseline, in the post-survey, the average scores for all items are between the "Somewhat prepared" and the "Very well prepared" thresholds. Indeed, examining the pre-post average differences, we observe a positive and statistically significant average difference for all 15 items. Furthermore, Cohen's d suggests that all the average differences are large (e.g., above 0.8 standard deviations). Notably, we observe the highest average difference, 1.18, for "Supporting students' development of models and explanations of phenomena;" Cohen's d indicates the pre-post means differ by 1.5 standard deviations. We also observed a 1.18 average difference for developing assessments to measure 3D science learning. While assessment was not at the forefront of the project, the formative assessment process was discussed throughout several PLC sessions to help teachers gather artifacts that showcased their students' learning trajectories. We find the third and fourth-highest average pre-post differences in our other items of interest: "Supporting students' learning trajectories" with 1.05 and "Developing students' scientific and engineering practices" with 1.0. Table 5 shows the results of the nonparametric tests, which find the average pre-post is statistically different from zero for the same items as the parametric tests.

Teachers' planning and delivery of science lessons.

To answer our second research question, we examine changes in the relevance of 10 elements in teachers' science instruction planning and delivery. In the baseline survey, the highest-rated item, on average, was the standards documents, and the lowest-rated item, on average, was classroom texts. Only three items on the baseline survey reached average scores beyond the "Important" threshold (e.g., the standards documents, collaboration with teachers in your same grade level, and artifacts of student work).

As teachers collect student artifacts while teaching the Water/Waves storylines, which were developed to support K-8 students' learning trajectories of these natural phenomena, and consistently work throughout the year with teachers across grade levels to map the big ideas of these storylines, we expect teachers to rate the standards documents, collaboration between teachers across grade levels, and artifacts of student work as significantly more important on the post-survey. Table 6 shows teachers rated collaboration with teachers outside their grade level significantly higher, 2.73, post-survey. Still, their average score in the post-survey remained between the "Somewhat important" and "Important" thresholds as in the baseline survey, and Cohen's d suggests the average difference is moderate in size (0.48 standard deviations). As for the relevance of artifacts of student work in teachers' science planning and delivery, we see a slight and non-significant increase in the post-survey with a score that remains between the "Important" and "Extremely important" thresholds. Regarding the importance of standard documents, we observe a surprisingly negative and significant average difference at the 10 percent level. We hypothesize that the project's emphasis on the storylines might have driven this negative average difference, even though the storylines were designed based on the standard documents. While the average rating remains just above the "Important" threshold, it might be necessary in the future to show and remind teachers of the direct connections between the storylines and the standard documents. Table 7 presents the results of the nonparametric tests, which find the average pre-post is statistically different from zero for the same two items as the parametric tests.

Exposure to science practices

To answer our third research question, whether students' exposure to science practices changed during the year, we consider a battery of items in both the teacher and student surveys. Starting with the teacher survey, teachers rated their students' exposure to science teaching practices (8 items aligned with the NGSS practices, two items associated with traditional instruction, and one item relating science to students' experiences) using a 5-point Likert scale (from "No exposure" to "In-depth exposure"). According to the teachers' average rating in the baseline survey, students do not have "Considerable exposure" or "In-depth exposure" to any of these practices. Regarding the practices associated with the NGSS, the highest rated were, with average scores between "Some exposure" and "Considerable exposure," (1) Asking questions and defining problems, (2) Obtaining, evaluating, and communicating information, and (3) Constructing explanations and designing solutions. As for students, Table 9 describes some summary statistics for 22 science practices in the baseline, which students had to rate based on how often they engaged in class using a 4-point Likert scale (from "I completely disagree; we never did that" to "I complete agree; we did that often"). Only for five items, on average, did students agree they engaged in at least sometimes: (1) Complete investigations that the teacher planned, (2) Discuss ideas about science with others, (3) Make statements or claims and support them using evidence, (4) Find scientific information on the internet, books, or articles, and (5) Use computers or other technology to help me learn science.

We expect a positive and significant difference in students' exposure to developing and using models, as it was the science practice most highlighted by the project, with students developing initial and revised models throughout the storylines, and these models constituting the basis for teacher reflection and discussion in the PLC sessions. Table 8 presents the pre-post comparisons, indicating that students' exposure to five science practices had positively and statistically increased. As expected, modeling had the highest pre-post change, and Cohen's d suggests it is a moderately to large average difference (0.71 standard deviations). The other science practices where students' exposure increased positively and significantly were (1) engaging in argument from evidence, (2) using mathematics and computational thinking, (3) analyzing and interpreting data, and (4) planning and carrying out investigations. Still, even in the post-survey, students' exposure to any of these science practices does not, on average, score beyond the "Some exposure" threshold. It is also worth noting that while the storylines emphasize some of these science practices by motivating student-student interactions, wonderings, and discussions, we do not believe the change in using mathematics and computational thinking is related to the project. Table 9 shows the results of the nonparametric tests, which find that the average pre-post is statistically different from zero for the same six items at the same significance level as the parametric tests.

As for the student surveys, Table 10 shows that 18 of the 22 pre-post differences regarding students' exposure to science practices are statistically significant. Echoing the teacher survey results, the average difference was positive and statistically significant for the three items (i.e., C, D, and E) that refer to developing and using models, with Cohen's d suggesting moderate to large average differences. The items associated with the science practice of engaging in argument from evidence (i.e., M, N, O, P, and Q), analyzing and interpreting data (i.e., I, J, and K), planning and carrying out investigations (i.e., F and G) also present positive and statistically significant average differences further supporting the teacher survey results. Table 11 presents the results of the nonparametric tests (Mann-Whitney for independent samples), which find that the average pre-post is statistically different from zero for the same items and at the same significance level as the parametric tests.

Conclusion

During the 2021-22 and 2022-23 school years, 22 K-8 science teachers in a small California district participated in the SPIRAL project, which provided ample opportunities for teacher collaboration within and across grade levels, with artifacts of student work at the forefront of reflection and discussion. In this study, we examined how participation in this project may influence teachers' science instruction, using survey data to compare pre-post responses. We first analyzed whether there were changes to teachers' self-efficacy over classroom practices intended to support science learning goals, finding positive and significant effects over all the included items, with the largest preparedness increase observed in supporting students' development of models and explanations of phenomena. Then, we analyzed whether the influence of various elements in planning and delivering science instruction had changed over time. While we found a positive and significant difference in the relevance of collaboration with teachers outside their grade level, we did not find evidence of a positive change in the influence of other elements associated with the project, such as the standards documents and artifacts of student work. Finally, we explored if students' exposure to science practices had changed over time, finding evidence in both the teacher and student surveys of increased exposure to modeling, engaging in argument from evidence, analyzing and interpreting data, and planning and carrying out investigations. Given our encouraging exploratory results and the continuous improvement of the storylines and digital tools during the last three years, any escalation of the program, which would allow us to increase the teacher sample size as well, would need to implement a causal research design.

References

Brown, J. C., Ring-Whalen, E. A., Roehrig, G. H., & Ellis, J. A. (2018). Advancing Culturally Responsive Science Education in Secondary Classrooms through an Induction Course. *International Journal of Designs for Learning*, *9*(1), 14–33. https://doi.org/10.14434/ijdl.v9i1.23297

Darling-Hammond, L., Hyler, M. E., & Gardner, M. (2017). *Effective Teacher Professional Development*. Palo Alto, CA: Learning Policy Institute.

Desimone, L. M. (2009). Improving Impact Studies of Teachers' Professional Development: Toward Better Conceptualizations and Measures. *Educational Researcher*, *38*(3), 181–199. https://doi.org/10.3102/0013189X08331140

Kloser, M., Edelman, A., Floyd, C., Martínez, J. F., Stecher, B., Srinivasan, J., & Lavin, E. (2021). Interrogating Practice or Show and Tell?: Using a Digital Portfolio to Anchor a Professional Learning Community of Science Teachers. *Journal of Science Teacher Education*, *32*(2), 210–241. https://doi.org/10.1080/1046560X.2020.1808267

Krajcik, J., Codere, S., Dahsah, C., Bayer, R., & Mun, K. (2014). Planning Instruction to Meet the Intent of the Next Generation Science Standards. *Journal of Science Teacher Education*, *25*(2), 157–175. https://doi.org/10.1007/s10972-014-9383-2

Liddell, T., & Kruschke, J. K. (2017). *Analyzing ordinal data with metric models: What could possibly go wrong?* [Preprint]. Open Science Framework. https://doi.org/10.31219/osf.io/9h3et

Martin, M. S., & Pardo, M. S. (2020). Estimating the Medium-Term Impact of a Scholarship Program Using Standardized Test Data. *Journal of Education and Learning*, *9*(1), 172. https://doi.org/10.5539/jel.v9n1p172

Martínez, J., Borko, H., Stecher, B., Luskin, R., & Kloser, M. (2012). Measuring Classroom Assessment Practice Using Instructional Artifacts: A Validation Study of the QAS Notebook. *Educational Assessment*, *17*(2/3), 107–131. https://doi.org/10.1080/10627197.2012.715513

Pierson, A. E., Clark, D. B., & Kelly, G. J. (2019). Learning Progressions and Science Practices. *Science & Education*, *28*(8), 833–841. https://doi.org/10.1007/s11191-019-00070-0

Shulman, L. (1998). Teacher portfolios: a theoretical activity. In *With portfolio in hand: validating the new teacher professionalism* (pp. 23–37). Teachers College Press.

Stoll, L., Bolam, R., McMahon, A., Wallace, M., & Thomas, S. (2006). Professional Learning Communities: A Review of the Literature. *Journal of Educational Change*, *7*(4), 221–258. https://doi.org/10.1007/s10833-006-0001-8

Sullivan, G. M., & Artino, A. R. (2013). Analyzing and Interpreting Data From Likert-Type Scales. *Journal of Graduate Medical Education*, *5*(4), 541–542. https://doi.org/10.4300/JGME-5-4-18

Trotter, E. (2006). Student perceptions of continuous summative assessment. *Assessment & Evaluation in Higher Education*, *31*(5), 505–521. https://doi.org/10.1080/02602930600679506

Wilsey, M., Kloser, M., Borko, H., & Rafanelli, S. (2020). Middle School Science Teachers' Conceptions of Assessment Practice Throughout a Year-long Professional Development Experience. *Educational Assessment*, *25*(2), 136–158. https://doi.org/10.1080/10627197.2020.1756255

Appendix

Table 1: Teacher distribution by grade level, storyline, and academic year

	2021-2022		2022-2023	
Grade	Number of Teachers	Storyline	Number of Teachers	Storyline
1st	2	Waves	3	Waves
2nd	6	Water	6	Water
4th	5	Waves	3	Waves
5th	4	Water	5	Water
6th	2	Water	2	Water
8th	3	Waves	3	Waves

Note: Two 4th-grade teachers moved grade levels at the start of the 2022-2023 school year. One moved to 2nd grade, and the other to 5th grade.

Table 2: Teacher sociodemographic information

	All	**Water cohort (2022-2023)**	**Waves cohort (2022-2023)**
Number of teachers	22	13	9
Number of schools	9	8	6
Female	14 *63.6%*	9 *69.2%*	5 *55.6%*
Experience Teaching Science			
2-5 years	3 *13.6%*	2 *15.4%*	1 *11.1%*
6-10 years	4 *18.2%*	2 *15.4%*	2 *22.2%*
11-15 years	5 *22.8%*	1 *7.7%*	4 *44.5%*
16-20 years	3 *13.6%*	2 *15.4%*	1 *11.1%*
21+ years	7 *31.8*	6 *46.1%*	1 *11.1%*
Race/Ethnicity			
White or Caucasian	5 *22.7%*	4 *30.8%*	1 *11.1%*
Hispanic or Latino	15 *68.2%*	9 *69.2%*	6 *66.7%*
Mixed heritage	2 *9.1%*	0	2 *22.2%*

Table 3: Students' sociodemographic information

	Fall 2022	Spring 2023
Number of students	251	267
Grade		
4th grade	89 *35.5%*	84 *31.5%*
5th grade	69 *27.5%*	83 *31.1%*
6th grade	25 *10%*	45 *16.8%*
8th grade	38 *15.1%*	32 *12%*
Unspecified	30 *11.9%*	23 *8.6%*
Female	101 *40.2%*	113 42.3%
Race/Ethnicity		
White or Caucasian	10 *4%*	7 *2.6%*
Hispanic or Latino	89 *35.5%*	88 *33%*
Asian	9 *3.6%*	9 *3.4%*
African American	3 *1.2%*	4 *1.5%*
Native American	6 *2.4%*	6 *2.2%*
Other	14 *5.6%*	24 *9%*
Mixed heritage	41 *16.3%*	62 *23.2%*
Prefer not to say	79 *31.4%*	67 *25.1%*

Table 4: Teachers' survey – pre-post comparisons self-efficacy on teaching practices – T-tests (parametric)

	N	Mean Pre	Mean Post	POST-PRE				
				Mean Diff	95% Conf Int.		Cohen's d	
Creating, designing, or articulating learning goals and expectations for students	22	2.45	3.32	0.86	0.58	1.15	1.35	***
Eliciting student thinking through multiple forms and media	22	2.77	3.59	0.82	0.39	1.24	0.85	***
Scaffolding instruction based on students' understanding of prerequisite material	22	2.64	3.23	0.59	0.21	0.97	0.69	***
Differentiating instruction to help students at different levels increase their understanding	22	2.41	3.18	0.77	0.28	1.26	0.70	***
Connecting disciplinary core ideas to phenomena	22	2.41	3.36	0.95	0.58	1.33	1.13	***
Using crosscutting concepts across disciplines	22	2.5	3.45	0.95	0.56	1.36	1.06	***
Developing students' scientific and engineering practices	22	2.23	3.23	1.00	0.69	1.31	1.45	***
Supporting students' development of models and explanations of phenomena	22	2.32	3.5	1.18	0.78	1.58	1.30	***
Developing assessments to measure 3D science learning	22	1.77	2.95	1.18	0.74	1.63	1.17	***
Using student artifacts to adapt and improve instruction	22	2.68	3.5	0.82	0.42	1.22	0.90	***
Providing feedback to help students understand their strengths and needs	22	2.5	3.45	0.95	0.51	1.4	0.96	***
Using students' partially correct ideas as assets to develop complete explanations of core science ideas	22	2.45	3.32	0.86	0.45	1.28	0.92	***
Supporting student learning trajectories of core ideas across grades	22	2.05	3.09	1.05	0.6	1.49	1.05	***
Using apps or other online tools to collect and review classroom artifacts	22	2.45	3.41	0.95	0.51	1.4	0.96	***
Offering structured feedback on student artifacts and classroom practices	22	2.5	3.27	0.77	0.32	1.22	0.76	***

Notes: (1) 4-point Likert scale – 1 "Not prepared," 2 "Somewhat unprepared," 3 "Somewhat prepared," 4 "Very well prepared." (2) *** significant at the 1%, ** significant at the 5%, *significant at the 10

Table 5: Teachers' survey – pre-post comparisons self-efficacy on teaching practices – Wilcoxon Signed Rank Test (non-parametric)

	N	Z-statistic	
Creating, designing, or articulating learning goals and expectations for students	22	-3.91	***
Eliciting student thinking through multiple forms and media	22	-3.41	***
Scaffolding instruction based on students' understanding of prerequisite material	22	-2.73	***
Differentiating instruction to help students at different levels increase their understanding	22	-2.77	***
Connecting disciplinary core ideas to phenomena	22	-3.77	***
Using crosscutting concepts across disciplines	22	-3.78	***
Developing students' scientific and engineering practices	22	-3.97	***
Supporting students' development of models and explanations of phenomena	22	-3.94	***
Developing assessments to measure 3D science learning	22	-3.61	***
Using student artifacts to adapt and improve instruction	22	-3.41	***
Providing feedback to help students understand their strengths and needs	22	-3.53	***
Using students' partially correct ideas as assets to develop complete explanations of core science ideas	22	-3.54	***
Supporting student learning trajectories of core ideas across grades	22	-3.64	***
Using apps or other online tools to collect and review classroom artifacts	22	-3.37	***
Offering structured feedback on student artifacts and classroom practices	22	-3.06	***

Notes: (1) *** significant at the 1%, ** significant at the 5%, *significant at the 10

Table 6: Teachers' survey – pre-post comparisons – influence of various elements in planning and delivering science instruction – T-tests (parametric)

	N	Mean Pre	Mean Post	POST-PRE				
				Mean Diff	95% Conf Int		Cohen's d	
Standards documents	22	3.36	3.05	-0.32	-0.63	0.00	-0.44	*
Classroom texts/textbooks	22	1.77	1.73	-0.05	-0.42	0.33	-0.05	
Artifacts of student work	22	3.14	3.36	0.23	-0.08	0.53	0.33	
My prior content knowledge or experience	22	2.91	2.82	-0.09	-0.50	0.32	-0.10	
Students' prior content knowledge or experience	22	2.27	2.32	0.05	-0.30	0.39	0.06	
Collaboration with teachers in your grade level	22	3.18	3.32	0.14	-0.18	0.45	0.19	
Collaboration with teachers outside your grade level	22	2.32	2.73	0.41	0.03	0.80	0.48	**
Consulting resources outside your grade level	22	2.82	2.59	-0.23	-0.66	0.20	-0.23	
Online resources	22	2.73	2.55	-0.18	-0.67	0.30	-0.17	
Other science resources	22	2.86	2.50	-0.36	-0.81	0.08	-0.36	

Notes: (1) 4-point Likert scale – 1 "Not important or influential," 2 "Somewhat important," 3 "Important," 4 "Extremely important or influential;" (2) *** significant at the 1%, ** significant at the 5%, *significant at the 10 percent level

Table 7: Teachers' survey – pre-post comparisons – influence of various elements in planning and delivering science instruction – Wilcoxon Signed Rank Tests (non-parametric)

	N	Z-statistic	
Standards documents	22	1.93	*
Classroom texts/textbooks	22	0.10	
Artifacts of student work	22	-1.51	
My prior content knowledge or experience	22	0.38	
Students' prior content knowledge or experience	22	-0.09	
Collaboration with teachers in your grade level	22	-0.90	
Collaboration with teachers outside your grade level	22	-2.02	**
Consulting resources outside your grade level	22	1.19	
Online resources	22	0.78	
Other science resources	22	1.72	

Notes: (1) *** significant at the 1%, ** significant at the 5%, *significant at the 10 percent level

Table 8: Teachers' survey – pre-post comparisons - students' exposure to science practices – T-tests (parametric)

	N	Mean Pre	Mean Post	POST-PRE				
				Mean Diff	95% Conf Int		Cohen's d	
Asking questions and defining problems	22	3.55	3.91	0.36	-0.14	0.87	0.32	
Developing and using models	22	2.91	3.91	1.00	0.37	1.63	0.71	***
Planning and carrying out investigations	22	3.05	3.68	0.64	0.15	1.12	0.58	**
Analyzing and interpreting data	22	2.91	3.55	0.64	0.08	1.19	0.51	**
Using mathematics and computational thinking	22	2.50	3.18	0.68	0.15	1.22	0.56	**
Constructing explanations and designing solutions	22	3.32	3.82	0.50	-0.15	1.15	0.34	
Engaging in argument from evidence	22	2.64	3.41	0.77	0.23	1.32	0.63	***
Obtaining, evaluating, and communicating information	22	3.50	3.64	0.14	-0.42	0.69	0.11	
Using informational text as the primary method to learn about key core ideas	22	3.41	3.36	-0.05	-0.60	0.51	-0.04	
Receiving direct instruction or presentation of a disciplinary concept	22	3.18	3.18	0.00	-0.45	0.45	0.00	
Connecting science learning to their own interests and life experiences	22	3.18	3.91	0.73	0.23	1.22	0.65	***

Notes:(1) 5-point Likert scale – 1 "No exposure," 2 "Limited exposure," 3 "Some exposure," 4 "Considerable exposure," 5 "In-depth exposure." (2) *** significant at the 1%, ** significant at the 5%, *significant at the 10 percent level

Table 9: Teachers' survey – pre-post comparisons - students' exposure to science practices - Wilcoxon Signed Rank Tests (non-parametric)

	N	Z-statistic	
Asking questions and defining problems	22	-1.43	
Developing and using models	22	-2.82	***
Planning and carrying out investigations	22	-2.41	**
Analyzing and interpreting data	22	-2.18	**
Using mathematics and computational thinking	22	-2.22	**
Constructing explanations and designing solutions	22	-1.54	
Engaging in argument from evidence	22	-2.49	**
Obtaining, evaluating, and communicating information	22	-0.49	
Using informational text as the primary method to learn about key core ideas	22	0.39	
Receiving direct instruction or presentation of a disciplinary concept	22	-0.14	
Connecting science learning to their own interests and life experiences	22	-2.89	***

Notes:(1) *** significant at the 1%, ** significant at the 5%, *significant at the 10 percent level

Table 10: Students' survey – pre-post comparisons - students' exposure to science practices – T-tests (parametric)

		N Pre	N Post	Mean Pre	Mean Post	Post-Pre				
						Mean Diff	95% Conf Int		Cohen's d	
A	Ask questions from observations of nature	216	241	2.88	3.18	0.30	0.18	0.43	0.44	***
B	Decide if a question is scientific	216	237	2.43	2.80	0.37	0.20	0.54	0.39	***
C	Create drawings or physical models of what happens in nature	216	240	2.94	3.34	0.40	0.24	0.56	0.46	***
D	Revise or update models when I get more evidence or information	215	240	2.73	3.36	0.63	0.46	0.80	0.69	***
E	Convince someone that a model is a good (or bad) explanation of what happened	216	239	2.42	2.88	0.46	0.28	0.63	0.48	***
F	Complete investigations that the teachers planned	214	236	3.16	3.35	0.18	0.01	0.35	0.20	**
G	Plan investigations with other students to answer our own questions	217	237	2.61	2.89	0.28	0.10	0.45	0.29	***
H	Use science ideas to predict what will happen in an investigation	217	236	2.85	3.31	0.47	0.30	0.63	0.53	***
I	Collect and analyze data to answer a scientific question	215	235	2.83	3.32	0.49	0.32	0.66	0.54	***
J	Organize and display information in tables or graphs	218	237	2.71	3.12	0.42	0.25	0.58	0.46	***
K	Use information in tables or graphs to find patterns or relationships	217	238	2.66	2.97	0.31	0.13	0.49	0.31	***
L	Write reports or reflections on investigations I did in class	214	237	2.82	3.09	0.27	0.95	0.45	0.29	***
M	Discuss ideas about science with others	217	239	3.27	3.52	0.25	0.11	0.39	0.32	***
N	Critique or challenge a scientific statement or claim	216	236	2.47	2.97	0.50	0.34	0.66	0.58	***
O	Make statements or claims and support them using evidence	215	236	3.13	3.56	0.42	0.27	0.58	0.52	***
P	Use different types of evidence to explain what happens in nature	217	238	2.85	3.40	0.55	0.40	0.70	0.66	***
Q	Compare different scientific	216	239	2.66	2.93	0.28	0.10	0.45	0.29	***

		N Pre	N Post	Mean Pre	Mean Post	Post-Pre				
						Mean Diff	95% Conf Int		Cohen's d	
	explanations to see which one is better									
R	Find scientific information on the internet, books, or articles	217	235	3.24	3.26	0.03	-0.13	0.19	0.03	
S	Evaluate if I can trust scientific information from the internet or books	217	239	2.48	2.62	0.14	-0.04	0.32	0.14	
T	Memorize science facts and vocabulary	215	240	2.79	2.74	-0.05	-0.24	0.13	-0.05	
U	Apply science ideas to solve real-world problems	218	238	2.65	2.97	0.32	0.15	0.49	0.34	***
V	Use computers or other technology to help me learn science	217	240	3.23	3.33	0.11	-0.06	0.27	0.12	

Notes: (1) 4-point Likert scale – 1 "I completely disagree; we never did that," 2 "I disagree; we rarely did that," 3 "I agree; we sometimes did that," 4 "I completely agree; we did that often." (2) *** significant at the 1%, ** significant at the 5%, *significant at the 10 percent level

Table 11: Students' survey – pre-post comparisons - students' exposure to science practices – Mann-Whitney Test (nonparametric)

		N PRE	N POST	Z-statistic	
A	Ask questions from observations of nature	216	241	4.28	***
B	Decide if a question is scientific	216	237	3.93	***
C	Create drawings or physical models of what happens in nature	216	240	4.60	***
D	Revise or update models when I get more evidence or information	215	240	6.65	***
E	Convince someone that a model is a good (or bad) explanation of what happened	216	239	4.77	***
F	Complete investigations that the teachers planned	214	236	1.81	*
G	Plan investigations with other students to answer our own questions	217	237	2.53	**
H	Use science ideas to predict what will happen in an investigation	217	236	4.91	***
I	Collect and analyze data to answer a scientific question	215	235	5.08	***
J	Organize and display information in tables or graphs	218	237	4.44	***
K	Use information in tables or graphs to find patterns or relationships	217	238	2.91	***
L	Write reports or reflections on investigations I did in class	214	237	2.67	***
M	Discuss ideas about science with others	217	239	3.12	***
N	Critique or challenge a scientific statement or claim	216	236	5.66	***
O	Make statements or claims and support them using evidence	215	236	4.98	***
P	Use different types of evidence to explain what happens in nature	217	238	6.15	***
Q	Compare different scientific explanations to see which one is better	216	239	2.69	***
R	Find scientific information on the internet, books, or articles	217	235	0.73	
S	Evaluate if I can trust scientific information from the internet or books	217	239	1.30	
T	Memorize science facts and vocabulary	215	240	-0.19	
U	Apply science ideas to solve real-world problems	218	238	3.29	***
V	Use computers or other technology to help me learn science	217	240	0.82	

Notes: (1) *** significant at the 1%, ** significant at the 5%, *significant at the 10 percent level