Related Paper Set for 2024 NARST International Conference

Building from Strengths and Attending to Context: Supporting Rural Science Teachers' Learning

Contact Information
Abraham S. Lo
E-mail: alo@bscs.org

Related Paper Set Citation

Lo, A. S.*, Gladwell, L., Herrmann-Abell, C., O'Connor, K., Allen, A.-R., Cooper, S. L., Jacobs, J., Cherbow, K., Penuel, W. R., Wingert, K., & Gardner, A. (2024, March 18). Building from Strengths and Attending to Context: Supporting Rural Science Teachers' Learning [Related Paper Set]. NARST 2024 Annual International Conference, Denver, CO.

Individual Paper Citations

Lo, A. S.*, Cooper, S. L., Herrmann-Abell, C. F., Cherbow, K., & Allen, A. (2024, March 18). Lessons Learned from Designing 5D Professional Learning for Rural Science Teachers [Conference Paper]. NARST 2024 Annual International Conference, Denver, CO.

O'Connor, K., Allen, A.-R., & Penuel, W. R.* (2024, March 18). Examining Science Teachers' Conceptions of Student Interest as a Consideration in Designing Assessments [Conference Paper]. NARST 2024 Annual International Conference, Denver, CO.

Herrmann-Abell, C. F.*, Lo, A. S., Cherbow, K., Cooper, S. L., Gardner, A., & O'Connor, K. (2024, March 18). Investigating the Impact of a 5D Professional Learning Course on Rural Teachers' Assessment Practices [Conference Paper]. NARST 2024 Annual International Conference, Denver, CO.

Glidewell, L.*, Jacobs, J., Allen, A.-R., & Wingert, K. (2024, March 18). A Comparative Case Analysis of Rural Teachers' Experience with 5D Professional Learning [Conference Paper]. NARST 2024 Annual International Conference, Denver, CO.

*Corresponding authors

Acknowledgements
We wish to thank the entire team from the 5D Assessment project for their ideas and support in developing this paper: Ari Jamshidi, Jamie Deutch-Noll, Will Lindsay, Lisa Carey, Becca Greer, and Melissa Campanella. This material is based in part upon work supported by the National Science Foundation under Grant Number DRL-2010086. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

Building from Strengths and Attending to Context:
Supporting Rural Science Teachers' Learning

One-third of American public schools are rural and one in five students attends a rural school (National Center for Education Statistics, 2018; Showalter et al., 2019). There is limited support for teachers in rural communities, however, for their own ongoing learning. Studies show that rural educators in STEM fields are less likely to experience rich professional learning experiences than their urban and suburban counterparts (Avery & Kassam, 2011; Banilower et al., 2018).

Characterizing rural educational landscapes, though, requires a greater awareness of the resources within rural communities. "Rural" describes a diverse range of communities (Hartman et al., 2022), but it is too often defined only in terms of population size and in contrast to cities (Azano et al., 2019). In addition, rural areas hold significant "community wealth" that could be leveraged in educational contexts, such as rural communities' resourcefulness, ingenuity, familism, and unity (Crumb et al., 2023).

In particular, the interests, experiences, and identities of rural students are assets for educators to build on and researchers to learn from. Students in rural areas can have a strong attachment to place that can motivate participation in STEM and support involvement in community improvement initiatives (Zimmerman & Weible, 2017). Phenomena of global significance, as well as those from contemporary science, can also be of great interest to rural students (Henson & Penuel, 2023).

Context

This related paper set explores different facets of rural science teacher learning in the context of a professional learning program focused on preparing teachers to design assessment tasks that embody the vision of *A Framework for K-12 Science Education* (National Research Council, 2012). A key goal of the professional learning was to prepare educators to design assessment tasks that are aligned to what we call a "five-dimensional" (5D) vision of science learning and performance, with a focus on the needs and concerns of rural teachers and their students. This vision draws from the *Framework*, which outlined a vision for meaningful "three-dimensional (3D) science learning," whereby students develop and use understandings of disciplinary core ideas (DCIs), science and engineering practices (SEPs), and crosscutting concepts (CCCs) to explain science phenomena or solve engineering problems. Implicit in the *Framework's* multi-dimensional vision are two additional latent dimensions, *interest* and *identity* (Bell et al., 2016). Central to this 5D vision involves students developing meaningful understandings for how to use the three dimensions (DCIs, SEPs, and CCCs) to make sense of phenomena and problems that are not only engaging to them, but also support the development of students' practice-linked identities (Nasir & Cooks, 2009; Nasir & Hand, 2008) as knowers, doers, and users of science.

The study findings presented here build from a rapid ethnography conducted by the project team (Wingert et al., 2022) that identified key strengths of rural teachers and resources in their contexts. That study revealed a strong value among teachers of strategies for anchoring teaching in locally relevant phenomena and project-based learning to bolster student interest in science. At the same time, rural teachers voiced a need for more science-specific professional learning. To address these needs, our project team engaged in design-based research to develop an intervention that promoted shifts in teachers' assessment practices in ways that 1) enhanced teachers' access to professional learning and 2) their ability to collaborate and develop relationships with peers.

Our design work involved six different stages: 1) a rapid ethnography of the professional learning needs of rural teachers (Wingert et al, 2022); 2) an initial adaptation of an 2-day existing in-person workshop designed to support teachers in designing 3D assessments (Penuel, Lo, et al., 2019); 3) soliciting feedback from a panel of rural teacher educators, state organization leaders, and assessment specialists; 4) a pilot study with rural teachers; 5) an experimental study to test the efficacy of the revised course; and 6) a delayed treatment course for control teachers. Figure 1 shows the timeline of this design work. During each stage of the design and implementation work, changes were made considering real-time feedback and the analysis of teacher experience data.

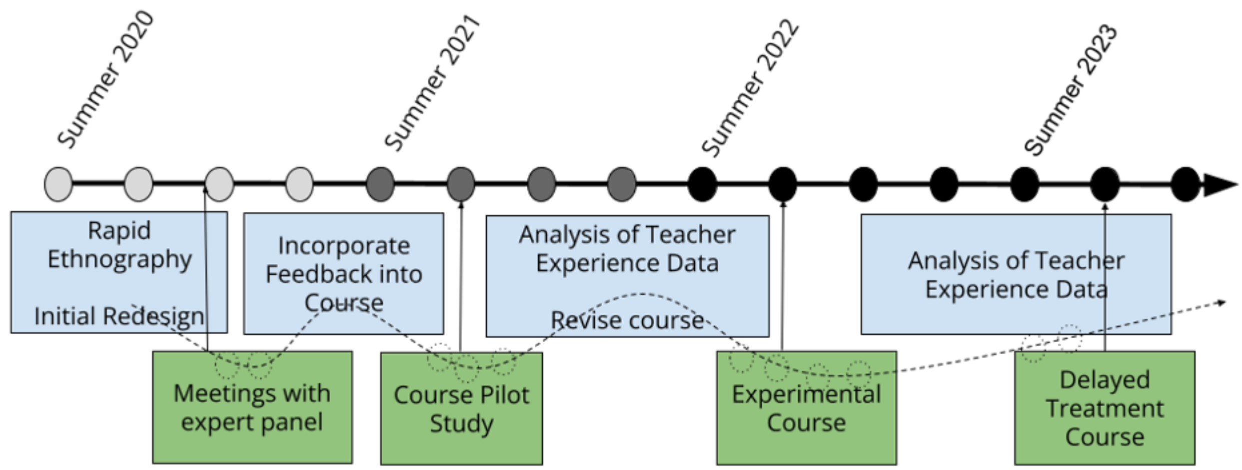


Figure 1. 5D Assessment Course development timeline

Organization and Contributions for Related Paper Set

The papers in this related paper set will show how our team used findings from our design research to adapt a professional learning program to address the needs of rural teachers and documents its impact on teachers' learning and assessment practices.

Papers 1 and 2 describe findings from our pilot study, which involved the initial adaptation and pilot of our 5D assessment course. In Paper 1, Lo et al. (2024) describe the process used to design the course that was the basis of our experimental study. They describe how we used findings from the initial ethnographic study and pilot study data to design supports to meet the needs of rural teachers and promote shifts in their assessment practices. In Paper 2, O'Connor et al. (2024) describe how teachers from the pilot study used student surveys to choose phenomena that reflected their students' interests and helped students recognize the relevance of science to their lives and their communities.

Papers 3 and 4 describe findings from our experimental study examining the effectiveness of the revised course on teachers' assessment practices. In Paper 3, Herrmann Abell, Lo, et al. (2024) present findings from our experimental study that highlights how the revised course led to improvements in teachers' assessments. In Paper 4, Glidewell et al. (2024) present a comparative case study analysis highlighting how variation in teacher backgrounds, rural contexts, and instructional priorities help account for differences in shifts in their vision for science teaching and instructional practice.

This paper set will be of interest to both those seeking to support rural educators' learning and to those interested in how best to support teachers in implementing a "5D" vision for science learning. It includes insights into both what is distinct about rural contexts and what are common needs of teachers seeking to design assessment tasks that reflect the vision of the *Framework.* Some key takeaways for working with rural teachers are the importance of attending to variation in teachers' contexts and experiences and the value of intensive, online professional learning for developing an expanded vision for science learning and repertoire for task design.

Paper 1: Lessons Learned from Designing 5D Professional Learning for Rural Science Teachers

Abraham S. Lo[1]*, Sara L. Cooper[2], Cari F. Herrmann Abell[1], Kevin Cherbow[1], and Annie Allen[2]
[1]BSCS Science Learning and [2]University of Colorado Boulder

Abstract

This research study investigated the extent to which an online course supported rural teachers in developing assessments that provide students with the opportunity to use the 3Ds to explain phenomena or solving engineering problems that engage student interest and support science-linked identity development. The course is an adaptation of a 2-day, in-person workshop and was designed to enhance rural teachers' access to high quality professional learning. Efforts were made to build community among rural teachers who often lack peers with whom to collaborate. This study documents the specific challenges rural teachers face in designing tasks to elicit students' understanding of the 3Ds in ways that connect to their interests and identities and how we used what we learned to iterate upon our design to create a more coherent and effective experience for our teachers.

Grounded in Wingert et al. (2022)'s research about the professional learning needs of rural science teachers in Colorado, we developed a three-month, online professional learning course to enhance rural teachers' access to high-quality professional learning and develop teachers' pedagogical design capacity (PDC; Brown, 2009) to develop and use assessment tasks that are aligned with a five-dimensional (5D) vision for learning and performance (Bell et al., 2016; Lo et al, 2022). 5D vision-aligned assessment tasks provide students with opportunities to use disciplinary core ideas (DCIs, D1), science and engineering practices (SEPs, D2), and crosscutting concepts (CCCs, D3) in integrated ways to explain phenomena or solve problems that are engaging and important to students and their communities (D4). By choosing engaging phenomena or problems, we create authentic reasons for students to use the 3Ds in meaningful ways, therefore contributing to the development of students' practice-linked identities (D5), as they are doing the work of scientists.

Developing teachers' capacity to design 5D vision-aligned assessment tasks is needed because teachers do not often have access to assessments that are aligned with NRC (2012) *Framework*-aligned standards (Banilower et al. 2018). Lo et al. (2022) identified five key areas of PDC, namely teachers' capacity to 1) understand what phenomena are and how they can be used to frame instruction and assessment, 2) "unpack" the targeted 3D understandings found in the standards, 3) learn about their students' interests and prior experiences with science, 4) choose phenomenon contexts that are engaging and productive for students to develop and demonstrate targeted understandings, and 5) develop assessments that prompt students' integrated use of the 3Ds to make sense of phenomena or problems. These areas are interconnected, as teachers' unpacking of standards and understanding of their students' interests influence how they choose phenomena and design prompts that support student sensemaking and science-linked identity development.

We engaged in design-based research (Design-Based Research Collective, 2003) to understand how to design an accessible 5D assessment professional learning program for rural science teachers. Designing assessments that both invite student sensemaking related to key science ideas and practices and that are culturally relevant and personally meaningful to students is not a straightforward process. In addition, teachers can benefit from tools that support the iterative development of tasks and from feedback from more expert task designers (Lo, Penuel, & Wingert, 2022; Penuel, Lo, et al., 2019; Penuel, Turner, et al., 2019).

Our target audience was rural science teachers because they often lack systematic infrastructure to support science-specific education (Wingert et al., 2022; Wingert & Penuel, 2019). Rural teachers experience different models for standards and curriculum adoption, with some teachers supported by larger districts with more funding and other teachers teaching three to four content areas as the only science teacher in their districts. We strived to design a course *for* rural science teachers that did not stereotype or essentialize the interests of rural students and teachers and represented the diversity of student experiences and cultural practices within rural communities. Our 5D approach to supporting teacher learning and shifts in classroom practices involved developing lenses through which teachers actively elicited and used information about their students and their local contexts and communities to make learning more meaningful to students, while addressing the contextual and systemic factors that impact rural science teachers' ability to access and engage in high quality professional learning to support shifts in classroom practices.

In this paper, we describe features of our course design and how we used data and feedback from our expert panel and pilot study to inform revisions to better support rural science teacher professional learning. Our research was guided by two research questions: 1) *In what ways did the 5D course support rural teachers in adopting 5D assessment practices?* and 2) *What were the successes and challenges encountered by teachers?* We used our findings to revise the course, which was evaluated in our experimental study. Herrmann-Abell, Lo, et al. (2024) share the impact of this work in the third paper of this related paper set.

Intervention Development

Our initial design work focused on adapting resources from an existing 2-day, in-person workshop designed to help teachers design 3D assessments. Teachers found developing assessments that involved students using the targeted 3Ds in integrated and grade-band appropriate ways to be challenging and benefited from opportunities to receive and act on feedback (Penuel, Lo, et al., 2019). In light of these findings, our team developed an online professional learning course that would provide teachers with support and opportunities to apply what they had learned over an extended period of time to develop their PDC and sustain shifts in classroom practice outside the context of professional learning.

Developing course storyline

Using a storylined approach (e.g., Edelson et al., 2021; Reiser et al., 2021), we designed sets of course sessions and tools to develop related areas of PDC and leverage what teachers had learned from previous sessions to answer the course-level question, *How do I design my classroom to support the vision for meaningful science learning and performance?*. In what follows, we describe the three session sets of the pilot 5D Assessment Course that was enacted during the fall semester of 2021. Table 1 summarizes the questions and learning goals that guided the design of each session set.

Session Set 1: What does 5D teaching look like? In the first set, teachers participated in a series of activities to motivate the importance of and relationships between key components of the 5D vision for meaningful science teaching and learning. For example, teachers engaged in a learner hat experience and watched videos of actual classrooms to see how phenomenon or problem-driven instruction could support meaningful science learning. Teachers then used language from the *Framework* to describe components of the 5D vision and how they were evident in the learner experience and classroom video. For teachers who were new to *Framework*-aligned instruction and standards, these activities provided the opportunity for teachers to explore the value of phenomenon-based instruction and how it could be used to develop the targeted 3D understandings in ways that were engaging to students and supported students as knowers, doers, and users of science. Taken together, the goal was for teachers to consider engaging student interest (D4) and identity (D5) as co-equal dimensions with the three NGSS dimensions.

We then motivated the need for teachers to explicitly collect information about their students so they could design 5D learning and assessment opportunities. Teachers had the opportunity to modify and administer the *Student and Community Interest and Identity Inventory* to their students, which consisted of 30 questions that elicited information about the questions or issues that interested their students and community, prior experiences using and doing science in science class, prior experiences in which they desired to learn more about a topic in science class, and a description of their ideal science classroom. Through the analysis of this data, teachers could use this information to develop potential entry points for motivating interest to issues that moved beyond individual or local relevance, and to consider the relevance of regional or global issues (c.f., Henson & Penuel, 2023). In addition, teachers could find out more information about students' prior experiences with science to affirm or repair past experiences with science.

Through this student lens, teachers then analyzed the understandings required to demonstrate the targeted 3D standards. The *Essential Unpacking Tool* was designed to support teachers in 1) analyzing the individual elements that comprised a 3D standard to understand the specific components of grade-level appropriate student performance, 2) understanding how those elements work together to provide students with opportunities to demonstrate complex understandings of science, and 3) identifying prior student experiences that teachers could build on to support students' use of the targeted 3Ds. We created opportunities for teachers to use grade band boundaries found in the appendices from the NGSS (NGSS Lead States, 2013) and the *Framework* to help teachers clarify the aspects of DCIs, SEPs, and CCCs that they are responsible for developing and assessing. In addition, teachers considered students' prior experiences, examples of real-world phenomena or problems where these explanatory ideas are relevant, and the questions or issues that students might have when designing 5D instruction and assessment opportunities.

Session Set 2: How can we use phenomena to frame instruction and assessment? Building on teachers' experiences with phenomena, teachers clarified their understanding of what phenomena are (and what they are not) and co-developed explicit criteria for choosing phenomena that will support meaningful, engaging 5D learning and assessment opportunities. Throughout this session set, teachers engaged with a phenomenon from a 5D vision-aligned high school biology unit that allowed us to characterize the ways in which phenomena can anchor a unit of instruction, support students in investigating key ideas using the 3Ds, and provide an opportunity for students to demonstrate what they had learned in the context of an assessment. For example, observations from a phenomenon should be presented in ways that highlight what is puzzling and pose a big question for students to authentically investigate using the targeted 3Ds. However, when designing assessments, teachers should choose sets of observations that can be explained within a shorter period of time.

Through the analysis of example assessments, teachers develop criteria for designing engaging 5D assessment scenarios that frame student sensemaking. These scenarios include information or data that make visible what is puzzling and the question that motivates students' authentic use of the 3Ds to figure out the puzzling phenomena or problem. Follow-up prompts would then scaffold and support sensemaking. Teachers then worked in small groups to consider revisions to these assessment scenarios to better align with the 5D vision.

We then introduced teachers to the *Choosing Candidate Phenomenon Tool*, which outlined four steps that embodied the developed criteria to design their own 5D assessment scenarios. In Step 1, teachers considered the explanatory power of the targeted DCIs and brainstormed phenomena that would engage their students' interest. Teachers were encouraged to use their analysis of their students' interest inventories and additional information about their students to inform their decisions. In Steps 2 and 3, teachers vetted each candidate phenomenon by checking its alignment with the targeted DCIs and whether the explanation of the posed question could be sufficiently answered using the targeted DCIs at grade band. In Step 4, teachers brainstormed data or information that could help students make visible what was puzzling about the phenomenon and use the targeted SEP and CCC elements, together with the targeted DCIs, to make sense of the assessment scenario.

Taken together, this work prepared teachers to develop principles that could be used to adapt existing assessment scenarios to make them more aligned with the 5D vision or develop their own 5D vision-aligned assessment scenarios.

Session Set 3: How can we develop and use tasks to assess student understanding? In the third session set, teachers crafted 3D prompts to scaffold student sensemaking of the assessment scenario. These targeted understandings elicited by these prompts are 3D, while the assessment as a whole is 5D because we integrate student interest and identity in the ways in which we choose and frame authentic student sensemaking of engaging phenomena or problems. Penuel, Lo, et al (2019) found that teachers had challenges developing multidimensional prompts that supported students' integrated use of the 3Ds. Part of the reason involved the use of unidimensional tools that presented teachers with example prompts to support students' use of the SEPs (Van Horne, et al., 2016) and CCCs (Penuel & Van Horne, 2016). Although useful for providing language for how teachers might elicit students' use of SEPs and CCCs, these tools did not provide explicit guidance for how the example prompts could be used to elicit grade band aspects of the dimensions or how to integrate students' use of the other dimensions to explain the scenario. To support an integrated vision for eliciting students' use of the 3Ds, we encouraged teachers to use their unpacking to envision how the targeted 3Ds could scaffold student sensemaking of the assessment scenario. Grounded in that initial vision, we then introduced teachers to the *Prompt Development Guide*, which framed how teachers might use language from the SEP and CCC tools to generate their own integrated prompts. It's not to say that all prompts needed to be 3D; rather, the sequence of prompts should lead students to integrate what they had learned related to the 3Ds to make sense of the phenomenon or problem.

Teachers worked with their coaches to finalize their assessments and administer them to their students during the month of November. Teachers then used student work to develop a 3D analytic rubric to evaluate their designed assessment's ability to elicit the targeted 3D understandings and alignment with the 5D vision. Teachers then considered how they might use this analytic rubric to provide 3D feedback to their students.

Although equity had been an explicit focus throughout the course, we culminated the course by considering ways to support diverse sensemaking opportunities through our assessments. Teachers used the *Supporting Diverse Student Sensemaking Checklist* to evaluate the extent to which they created diverse opportunities for students to represent and express their thinking and consider revisions to make their assessments more equitable. We did this intentionally because our prior research demonstrated the challenge of developing 3D assessments (Penuel, Lo, et al., 2019), and we were wary of adding an additional lens to the assessment development work.

Table 1
Course session questions and learning goals

Questions	Learning Goals
Session Set 1: What does 5D teaching look like?	
● What does meaningful science learning look like in the classroom? ● How do we currently attend to the 5Ds in our instruction? ● What guidance do the standards provide for designing 5D instruction and assessment opportunities?	● Develop an explicit understanding of components of the 5D vision ● Understand how phenomenon-based instruction can support our vision of meaningful, 5D science learning ● Understand how to use the NGSS and state standards documents to identify targeted 3D understandings
Session Set 2: How can we use phenomena to frame instruction and assessment?	
● How can phenomena be contexts for developing and assessing 3D understanding? ● How do we choose phenomena to frame instruction and assessment? ● How can we ensure that phenomena support opportunities for student sensemaking?	● Identify criteria for choosing productive phenomena to develop and assess 3D understanding ● Identify features of 5D assessment scenarios that create opportunities for students to demonstrate 3D understanding
Session Set 3: How can we develop and use tasks to assess student understanding?	
● How do we write prompts that create meaningful opportunities for students to use the 3Ds to make sense of phenomena or problems? ● How can we make sense of what students have learned and provide meaningful feedback? ● How can we support diverse sensemaking using assessments?	● Understand how to use criteria and tools to develop accessible prompts to make sense of assessment scenarios. ● Understand how to use student work to identify the incremental build of students' 3D understanding. ● Identify ways to make assessments more accessible and better support diverse sensemaking

Outcomes for the course. In total, participating teachers engaged in 25 hours of professional learning spread across 3 months. Each session was scheduled to be 2.5 hours, with select sessions lasting 3 hours due to the complexity of the work. Since teachers often have limited time outside of class due to competing demands (Wingert et al., 2022), we provided time during the course for teachers to work and receive feedback from their coaches and peers. The desired teacher course outcomes included shifts in teachers' understanding of the 5D vision and the development of 5D assessment practices, as evidenced by the presence of 5D assessment features in their designed or modified assessments. Acknowledging the challenge of developing phenomenon-driven, 5D vision-aligned assessments, we decided to focus on supporting the design of phenomenon-driven assessments rather than incorporating the additional layer of solving engineering problems. While coherence with classroom instruction was important, we did not explicitly support the adoption of aligned instructional practices due to time constraints. Throughout the course, we intentionally used examples from free open educational resources with educative features that could support teachers in implementing 5D vision-aligned pedagogical shifts in classroom instruction if desired.

Features to support the needs of rural teachers. Grounded in Wingert et al (2022)'s rapid ethnography and feedback from our panel of rural educators and assessment experts, we designed features to support the learning of rural science teachers. We created digital tools and community spaces for teachers to access resources and share their ideas. To foster online collaboration and work, we created a "virtual learning space," which contained hyperlinks to all digital resources and placed to document shared sensemaking and work to be completed between course sessions. In addition, each teacher had a Google Drive folder where they had copies of completed tools and developed their assessments. All course participants and coaches had shared access to these folders to foster collaboration.

Because teachers often have competing demands on their time and sometimes serve as the only science teacher at their school (Wingert et al., 2022), we wanted to ensure that teachers recognized the practical application of what they were learning and had opportunities to collaborate with other rural science teachers during the course who were teaching similar content areas or aged students. Teachers were placed into affinity groups based upon grade level, discipline, and chosen performance assessment targets. The consistency supported more productive small group discussions and facilitated effective task feedback since group members were familiar with each other and the work being presented. Members of the research team and facilitation team served as coaches to cultivate relationships and provide feedback on teachers' work. Coaches were assigned to affinity groups based upon their experience with grade level and discipline content.

Methods

To study how our course supported the development of rural science teachers' PDC and inform future revisions, we conducted a pilot study during the 2021-22 academic year. We recruited eleven teachers from two rural school districts (three middle and eight high school science teachers). Teachers were experienced, having taught an average of 14.8 years, and had access to a moderate amount of professional learning to learn about the 2020 Colorado Academic Standards for Science (11.9 hours) and relevant assessment practices (6.3 hours) prior to their participation in the 5D Assessment Course.

To analyze shifts in teachers' assessment practices, we asked teachers to choose a unit for which they could develop or modify an existing assessment to make it more aligned with the 5D vision. Teachers submitted an assessment to us before and after experiencing the course that they felt aligned with the 5D vision. Teachers shared the rationales for their design decisions, scoring guidance, and examples of student work. We developed and used a rubric to analyze teachers' assessments before and after the course in five areas: 1) teachers' use of phenomena to support 3D sensemaking, 2) opportunities for students to use the targeted 3Ds in integrated and grade band appropriate ways, 3) mechanisms for providing 3D feedback to students, 4) teachers' intent to engage student interest, and 5) opportunities to enhance accessibility of assessment. Students completed a survey to share their experiences taking the assessment.

Members of the research team were also members of the course design team, which ensured that the desired 5D assessment elements were supported by the course. Due to the complexity of the work, we engaged in paired scoring and resolved differences among pairs through consensus discussions. The findings from the assessment analysis then informed revisions to the course.

To understand teachers' experiences with the course and explain observed shifts in teachers' assessments, we collected: 1) field notes of course sessions, 2) course artifacts, 3) course exit tickets 4) semi-structured interviews with teachers, and 5) an end-of-course survey. We analyzed field notes for themes related to the role of phenomena, 3D standards, and student identity and interest. We coded interview transcripts and exit ticket data with the same themes to understand teachers' uptake, engagement, and concerns with the foundational ideas and tools of the course.

Findings

In what follows, we describe our analysis of teachers' designed assessments, how teachers' course experiences could explain these findings, and describe revisions that were made considering teachers' experiences.

Table 2
Pre and post assessment score analysis

Assessment Practices Outcome		*Mean*	*Std. Deviation*
Using Phenomena for Sensemaking (23 pts)	*Post*	*.56*	*.32*
	Pre	*.52*	*.36*
Designing to engage student interest (1 pt)	*Post*	*.64**	*.51*
	Pre	*.27*	*.47*
Targeting integrated & grade-appropriate use of 3Ds (32 pts)	*Post*	*.46*	*.18*
	Pre	*.57*	*.13*
Providing 3D Feedback (3 pts)	*Post*	*.39*	*.33*
	Pre	*.18*	*.35*
Enhancing accessibility (4 pts)	*Post*	*.52*	*.26*
	Pre	*.48*	*.28*

**Statistically significant at 0.05 level $t(10) = 2.19$; Cohen's d effect size = 0.72*

Analysis of Teachers' Assessment Practices

We used our rubric to evaluate the presence and use of key 5D assessment features. For each category of related elements, we calculated the percentage of points earned. Table 2 shows the mean calculated scores for the pre and post assessments for each assessment practice outcome. Paired-samples t-tests were conducted to compare the pre- and post-scores of the design study participants. There were positive shifts in each category except for integrating opportunities to use the 3Ds, which continued to be a challenge for teachers. Of the other outcomes, the intent to engage student interest was statistically significant on the .05 level ($t(10) = 2.19$). The Cohen's d effect size for this difference was 0.72.

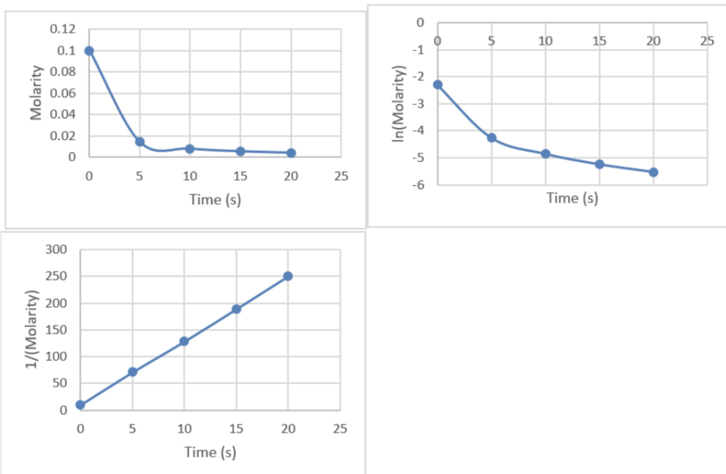
Supporting students' use of phenomena for sensemaking. When comparing teachers' post assessments to their pre-assessments, we saw more instances of teachers requiring students to use what they have learned to explain a phenomena or problem (pre=7, post=8). There was greater use of specific events or problems rather than general class of phenomena. More teachers used real data or information to make visible what was to be explained about the phenomena or problems (pre=4, post=8). However, teachers continued to have challenges with presenting what was puzzling about the phenomenon in ways that would engage student interest and designing multiple prompts to scaffold students' sensemaking of the phenomenon or problem. For example, one teacher developed a post-assessment that required students to read and gather information from an article on coral reef loss since 1950. This article featured various figures illustrating the declines of coral reef cover over time. Students developed a claim supported by evidence and reasoning from the article to explain this coral reef loss. The assessment presented real data and information about coral reef loss to make the phenomenon visible. However, the assessment lacked any framing that explicitly invited students to wonder about coral reef loss. The task began with the article without providing any direction or information motivating them to read it. Furthermore, any puzzling aspects of the coral reef phenomenon were immediately resolved within the context of the article.

Opportunities for students to use the targeted 3Ds in integrated and grade band appropriate ways. Although opportunities for sensemaking often involved the targeted 3Ds, these opportunities were not always integrated or grade band appropriate. Below grade band use was often due to not intending to align with the dimension at the elemental level. When teachers shared with us the 3D assessment targets, teachers were more likely to identify DCI elements, whereas targeted SEPs and CCCs were often at the dimensional level. All the teachers (n=11) specified a DCI element in their post-assessment, while only 6 teachers specified an SEP element and 5 teachers specified a CCC element. These differences in intent to align with dimensional elements was consistent with what was observed in teachers' assessments, in which there were greater opportunities for students to use the DCIs at the elemental level and the SEPs and CCCs at the dimensional level. Thus, greater professional learning support was needed to help teachers to consider not just 3D sensemaking opportunities, but to consider what specific aspects were required at middle or high school to address the performance standard and ensure that teachers were building upon earlier grade bands and supporting increasingly sophisticated 3D understandings.

Because the CCCs were the least familiar to teachers, greater support may be needed to help teachers explicitly use the CCCs in intentional ways to explain phenomena. Only three post-assessments explicitly prompted the use of CCCs in their questions. The other tasks either did not use the CCCs at all or involved implicit use, where the CCCs were not an explicit goal for completing the assessment. For example, one teacher aimed to target the CCC, Stability and Change, with their post-assessment on kinetics and catalysts. Students were tasked with determining the rate of reaction and the mechanism for that reaction using Concentration-Time graphs (see Figure 2). These questions implicitly referenced stability and change as they considered how the chemical reaction changed and then regained stability. However, this CCC was not noted in the questions and was not explicitly assessed in the scoring guide.

Use the following graphs and equation for Questions 4-5:

OVERALL REACTION: $2A + B \rightarrow C + D$



4. Based on the graphs above, what is the overall order of the reaction?
    a. Zero
    b. First
    c. Second
    d. Cannot be determined

5. Which of the following could potentially be mechanisms for the reaction above taking into account the graphs provided and the overall reaction written above?

a. $A + B \rightarrow E$ (slow)
   $A + E \rightarrow C + D$ (fast)

b. $A + B \rightarrow E$ (slow)
   $E \rightarrow C + D$ (fast)

c. $A + A \rightarrow E$ (fast)
   $E + B \rightarrow C + 2D$ (slow)

d. $B + B \rightarrow E$ (slow)
   $E + A \rightarrow C + D$ (fast)

Figure 2. Post-assessment example.

Mechanisms for providing 3D feedback to students. This category examined the extent to which the guidance used to score students' assessments or provide feedback to students was multidimensional. There was an increase in the use of multidimensional rubrics in teachers' post-assessments (n=6) compared to in their pre-assessments (n=3). For example, one teacher did not provide a scoring guide for their pre-assessment, only noting correct responses. In the post-assessment, they provided a scoring guide that explicitly assessed all three targeted dimensions, identifying the dimensions targeted in each question and outlining the specific qualities of thinking and performance related to these dimensions. In addition, there was a shift towards including more specific, rather than generic, guidance in teachers' post-assessment (n=6) scoring guidance compared to the pre-assessment (n=2). However, only two scoring guides from the pre- and post-assessments were three-dimensional. The remaining multidimensional rubric targeted DCI and SEP together in all but one case (SEP and CCC).

Greater specificity in teachers' scoring guidance meant that students were provided with concrete ways in which they could improve their use of the 3Ds in the context of the actual assessment rather than general guidance. For example, one teacher provided a generic Claim-Evidence-Reasoning rubric to accompany their pre-assessment on natural selection. However, this rubric did not specify how students engaged in argumentation in the assessment. In their post-assessment on genetics, the teacher provided a scoring guide that differentiated various levels of performance for specific assessment questions related to development and use of Punnett squares and explanations of genetic inheritance.

Teachers' intent to engage student interest. This category examined the extent to which the teacher identified specific ways in which they attempted to engage their students' interests when designing their assessments. Specificity required teachers to use specific information about their students rather than general guidance about student interest (e.g., all kids like fireworks). Teachers had provided more specific ways and intent to engage student interest when designing their post-assessments (n=7) compared to in their pre-assessments (n=3)

Opportunities to enhance accessibility of assessment. More teachers identified ways in which they sought to enhance the accessibility of their post-assessments (n=8) compared to their pre-assessments (n=5), which included the use of multiple modes for students to express and represent their understanding. However, our analysis of teachers' tasks continued to show evidence of barriers to access.

In summary, our analysis of teachers' post-assessments showed growth in many key areas of teachers' PDC to develop 5D assessments. For example, our pilot teachers grew in their ability to design phenomena-driven assessments that engaged student interest and supported 3D sensemaking and provided students with actionable feedback. Despite this growth, there was still significant room for improvement in each area. Most importantly, there continued to be persistent challenges with presenting the phenomenon in ways that made visible what was puzzling about the phenomenon and providing students with the opportunity to use the 3Ds in grade band appropriate ways.

Using teachers' experiences to inform course revisions

In what follows, we describe how the teachers' course experience could explain observed assessment scores and justify the design of course revisions to further develop teachers' capacity to develop 5D assessments.

Greater support needed to develop teachers' understanding about what the 5Ds are and how they could be used in assessments. To develop the 5D vision, teachers experienced and made sense of the Roadkill phenomenon (NextGen Science Storylines Team, 2019), which involved students wondering what happened to a dead badger that was found at the side of a road. We coupled our learner experience with three videos of classroom instruction to see how students' initial ideas about the phenomenon could drive future learning and identify evidence of student engagement and students acting as knowers, doers, and users of science. We engaged with this phenomenon to help teachers to see 1) how phenomenon-driven instruction could support engaging, 5D learning and 2) support teachers in thinking about how each dimension was distinct, yet powerful when used in integrated ways for sensemaking. Our hope was that by outlining this vision early in the course, teachers could leverage this integrated vision when developing their assessments.

Although pilot study teachers found this work to be engaging, it was not clear to teachers how this visioning work was relevant for helping them design an assessment. Two teachers reflected that they would have liked more time earlier in the course for assessment design, and suggested a shorter focus on developing 5D vision to make room for assessment design sooner. Further, the need to motivate teachers to consider phenomenon-driven instruction was significantly reduced, since the use of phenomena to drive instruction and assessment had gained broader acceptance within the field. In response, we incorporated assessment as a more explicit component of this visioning work – articulating a 5D vision of meaningful science learning *and performance* as evidenced in classroom instruction and in the products that students produce to demonstrate their understanding. Grounded in teachers' initial ideas, we used contrasting examples of assessments from each grade band and content area to clarify our vision for meaningful performance, with a focus on the role that phenomena played in these assessments. Through this work, teachers considered how phenomena could be contexts for students to demonstrate their understanding.

Pilot study teachers were also less familiar with the 3Ds of the *Framework* than had been expected. We assigned readings from the NRC (2012)'s *A Framework for K-12 Science* to familiarize teachers with the standards. We had intended for much of this work to happen outside of class to focus our time on 5D assessment development. However, we found out quickly that this support was not sufficient for many of our pilot study teachers, as some of them were brand-new to the *Framework* and 3D standards and required more in-class support. For example, one teacher said her district had not taken up 3D standards, and she was not sure if her rural school was "behind the time;" explaining, "I don't even know what we're missing as rural teachers…" (Pilot teacher interview, 1/13/22). In response, we took time during the class to process the readings and use examples to help teachers understand concretely what the SEPs and CCCs were and how they were evident in the curriculum examples used in the course. Without this foundation, it would not be possible for teachers to unpack the3D standards and choose aligned phenomena for their assessments. Even for teachers who came into the course with some familiarity with the standards, they said the attention to the 3D standards helped them see "at what rigor and at what degree should we be teaching these new standards. Anybody can look through these new standards…but without the class, I wouldn't dig into the SEPs and CCCs as much" (Pilot teacher interview, 1/18/22).

In our revised course, we designed differentiated, asynchronous pathways for teachers to receive the support they needed to complement the revised 5D vision foundation work and become familiar with their states' 3D standards and its alignment with the NGSS. To help teachers develop a common language and leverage resources developed for the national standards, we created a standards crosswalk document that helped teachers navigate similarities and differences between each state's standards and the NGSS, so that teachers could access valuable NGSS resources to develop their classroom practices. We encouraged teachers from similar states to work together and work with their coaches to provide further support.

Identifying more productive, exemplar phenomenon contexts. Since our course involved secondary teachers, we used a phenomenon-driven assessment grounded in a high school life science standard that involved explaining how the wingspan of swallows changed over time (Northwestern University and Inquiry Hub Partnership team, 2018). To show how students would be prepared to complete such an assessment, teachers engaged in a second anchor phenomenon experience to understand how a population of Juncos changed over time. Although well-intentioned to meet the needs of our target audience, this high school life science example was not accessible for all the teachers, particularly those who were not as familiar with high school life science. Furthermore, the model assessment task did not display all the features of high-quality assessments that were central to the course goals (e.g. grade-appropriate use of data).

We decided to focus our work on the Roadkill anchor phenomenon, because it was more engaging to students and involved a 5th grade standard (5-LS2-1) that was more accessible to our diverse learners. In addition, it allowed teachers to think about how these ideas became more sophisticated in later grade bands. A significant amount of design work was needed to develop a model assessment for the Roadkill unit that would allow us to build upon the learner experiences and demonstrate coherence between instruction and assessment more clearly. The assessment scenario involved trying to figure out why whales are so big and what happens to whales when they die. Because the assessment writer was one of our course facilitators and designers, we were able to show how the course tools could be used to design assessments (e.g. choosing phenomena that are informed by student interest and aligned with 3D standards, finding data to motivate 3D sensemaking, and how to develop a coherent task storyline) and make visible the iterative, and often messy, thinking process that contributed to the final assessment design.

We hoped that revising the course around this exemplar context would enhance teachers' understanding for how each grade level builds upon one another and better support teachers in choosing phenomena that would elicit the required 3D understandings and developing assessment scenarios that would support meaningful student sensemaking.

Dedicated time for collaboration among rural science teachers. The design of the pilot course was intended to provide rural science teachers with high-quality, *Framework*-aligned PL focused on assessment. Teachers in the course reported that they valued this opportunity to work with colleagues teaching in schools and contexts like theirs. One said, "I don't really get to work with other people very often" and the online course provided "the opportunity to take a course that we don't have offered around here" (Pilot Teacher Interview, 1/19/22). Another teacher said, "it is nice to be able to talk shop with people who teach the same subject" (Pilot Teacher Interview, 1/13/22). A third teacher reflected that they wanted more time to collaborate with colleagues in the course, saying "we didn't really have enough time to talk with each other about what worked and what didn't work" and that particularly for rural teachers, the opportunity to "actually talk with other teachers about the content and how it worked and how it rolled out" is a valuable professional experience (Pilot Teacher Interview, 1/18/22).

Enhancing access and use of online resources. During the pilot study, we created a virtual landing page that contained links to various session resources for teachers to modify collectively or make copies of files for their individual sensemaking. Teachers found this landing page to be overwhelming, as teachers needed to have many tabs open each session. With limited screen space, resource management during course sessions was a challenge. Pilot study teachers suggested using Google Classroom to manage the course and teacher-edited resources that could be accessed by coaches and peers for feedback. Because teachers' internet connections were not always consistent, teachers could access Google Classroom and catch up with the rest of the group by utilizing cues in slides to what resources we were using at a given time.

Teachers felt that the course tools were valuable. Since tools were introduced as the course progressed, teachers did not always leverage work from prior tools to inform later work in the assessment development process, which may have hampered the coherence in teachers' assessment development work. In addition, after the course concluded, teachers became overwhelmed in their attempts to recall which tools they were to use when and how. Thus, the lack of usability of the tools without the support of the course may have hindered the development of pilot teachers' post assessments. A tool that teachers found to be both valuable and difficult to use was the initial version of the unpacking tool, which the design team then simplified. One teacher reflected "the first unpacking tool, that one just felt really hard to use, I don't think I would go back to that one. But I would use a paired-down version" (Pilot Teacher Interview, 1/20/22). Another teacher shared that the tool she "got the most out of, even though it took the most time, was the unpacking tool because it just makes it so clear what the instruction should be about" (Pilot Teacher Interview, 1/20/22).

To help teachers use the tools in a more systematic way, we consolidated all the tools into a single electronic document called the *5D Assessment Task Development Tool*. We organized tool use around seven phases of work so that teachers understood the purpose for using each tool and how they could be useful for developing the 5D assessments. The seven steps were:

- *Step 1: Evaluate potential for engaging student interest and identity*
- *Step 2: Evaluate alignment of the phenomenon with target DCIs*
- *Step 3: Identify data or information for sensemaking*
- *Step 4: Draft the scenario*
- *Step 5: Construct assessment storyline*
- *Step 6: Brainstorming prompts to support 3D sensemaking*
- *Step 7: Constructing ideal responses for prompts*

Steps 1-3 evolved from the *Choosing Phenomenon Tool* and were renamed to make explicit teachers' use of the 5Ds to brainstorm and evaluate candidate phenomena. In addition, we included explicit alignment and accessibility checks to ensure that the chosen phenomenon would involve the targeted 3Ds, engage student interest, and be accessible to teachers' students. *Step 4: Draft the scenario* was needed as an explicit step because teachers expressed challenges with presenting the data or information to students so that it was clear 1) what students were to make sense of using the 3Ds and 2) had the big question in mind to frame student sensemaking. Teachers were presented with additional guidance for crafting this assessment scenario and received an *Assessment Scenario Checklist* that summarized key features. *Step 5: Construct assessment storyline* was created to enhance the coherence of teachers' assessments so that the prompts developed in *Step 6: Brainstorming prompts to support 3D sensemaking* would support students' sensemaking and answering the big question posed by the scenario. Teachers used *Step 7: Constructing ideal responses for prompts* to construct ideal student responses to evaluate the coherence of the assessment from the student perspective. Teachers used a *Task Prompt Checklist* to evaluate their prompts.

Supporting opportunities for grade-band appropriate use of the 3Ds. When analyzing teachers' post-assessments, we realized that teachers were still having issues with creating opportunities for students to use the targeted dimensions in grade-appropriate ways. There are several potential reasons for why this could have occurred. First, when doing the unpacking, we asked teachers to choose 3D standards and unpack the foundational elements that informed the design of that standard. However, it may not have been as explicit to teachers that the elemental use of the dimension was critical for ensuring that students' use of the dimension was grade-band appropriate and attending to the specific ways in which each dimension was intended to work together to support the targeted 3D performance. In addition, it may not have been clear to teachers *how* and *why* their choice of phenomena and the design of the assessment scenarios needed to connect to their elemental unpacking. For many teachers, identifying phenomena that aligned with the targeted dimensions, let alone the dimensions at the elemental level, was a challenge.

In addition to reorganizing the tools, we also created an additional course session to complement the unpacking session to help teachers explicitly think about grade-band appropriate use of the targeted dimensions. Through this work, teachers recognized how each dimension varied by grade band (i.e., How are elementary, middle, and high school use of the dimensions different and related to one another?). Thus, there was more attention in the revised course to establishing the grade band context that motivated teachers to closely examine the targeted performances at the elemental level. We hoped that the revised organization of the tools and the explicit connection for how the unpacking should inform the evaluation of choices of phenomena and the choices for data and information would improve students' opportunities to use the targeted dimensions at grade level.

Supporting greater student engagement in designed assessments. A central part of our work involved motivating the need to consider students' interests and identities when designing 5D assessments. Teachers quickly recognized the value of using phenomena to support more "locally relevant" and "authentic" student learning. However, teachers' students did not report statistically significant shifts in the quality of their engagement with assessments. We found that many teachers did not analyze and consistently use specific information about what they learned about their students to inform their design decisions. Part of the issue was having sufficient time to both administer and analyze the data from the *Student and Community Interest and Identity Inventory*. In our revised course, teachers administered the inventory after Session 1, which allowed more time for teachers to analyze and make sense of their students' data. In addition, we revised our tools to include explicit prompts for teachers to use their understanding about their students throughout the design process. For example, when choosing phenomena to frame student sensemaking, teachers were reminded to consider what might be puzzling about the phenomenon from the student's perspective and consider ways to sustain their students' interest throughout the assessment.

Supporting opportunities for 3D sensemaking. Our pilot study teachers made progress with thinking about the usefulness and value of choosing and using phenomena for use in assessments. However, when it came down to designing assessment scenarios that made visible what was puzzling about the phenomenon, issues continued to persist. Because the example assessments we shared during the pilot course included adapted data from a scientific paper, teachers felt the need to find scientific data to use in their assessment scenarios and expressed concern about the amount of time that they had spent "finding data". Although using scientific data is one important way to foster authentic sensemaking, teachers did not always use their unpacking or the big question they wanted kids to figure out to guide their search for data. We also did not want teachers to think that finding data from the experiment was the only way to make visible what was puzzling about the phenomenon.

In the revised *5D Assessment Task Development Tool*, there was stronger coupling between the teachers' unpacking and the evaluation of candidate phenomena to ensure that students needed to use the targeted DCI elements to make sense of it. When developing their assessment scenarios, teachers brainstormed the type of data or information they would need to make visible what was puzzling and ensure that the chosen data and information would require students to use the targeted 3Ds. Further, they would ensure that the chosen data or information was presented in an accessible way to ensure that students could see what was to be explained and connect how they could use what they learned to make sense of it. To facilitate this work, we revised and reorganized the 5D Assessment Development tool to include links to relevant work to ensure that teachers had access to their unpacking and insight about their students to inform the design and the coherence of their assessment work.

Another improvement involved providing resources to support the design of more coherent assessments so that each prompt worked together to support students in figuring out the big question posed by the assessment scenario. To facilitate this work, we asked teachers to brainstorm the types of questions that students would have about the proposed assessment scenarios and begin chunking these questions together in ways that allowed students to engage in sensemaking in service of answering the big question. Teachers were invited to think about what data or information from the scenario would be needed to investigate these questions – or whether additional information would be needed. This ensured that the prompts were connected to figuring out the assessment scenario and prevented the use of phenomena or the presence of data or information from being used as "contexts" only.

Use of examples to illustrate design thinking. The pilot study version of the course included completed examples of each of our tools. For example, in the pilot study, we engaged in a notice and wonder activity to analyze a completed unpacking tool for a MS and HS standard. However, it was not always explicit the decision making that was involved to populate the tools.

In our revised course, we shared contrasting examples of a completed unpacking tool for a single standard, which showed how unpacking could look different and invited teachers to think about key considerations that they may want to include in their unpacking. It also facilitated conversations about grain size and what was needed to ensure that teachers understood what it meant by the different elements and what it would look like for students to demonstrate this understanding. We hope that this revised method of introducing unpacking to teachers may lead teachers to dig into the dimensions at the elemental level and understand what is involved and why it is important.

Similar to the goals of supporting teachers' unpacking of the targeted standards, we hoped that providing examples of how a task designer used our tools could highlight key decision making principles that could support teachers' own design work. We provided a completed 5D Assessment Tool for the Roadkill unit transfer task and provided additional professional learning resources during the course to make visible the decisions that led to the choice of phenomenon and the data and information used to make visible what was puzzling and support student sensemaking.

Focus on providing feedback rather than developing an analytical rubric. During the pilot study, we did not foreground methods for grading or evaluating student work; rather, we engaged teachers in designing a 3D analytical rubric to assess what student understanding was elicited with the goal of improving assessment design. This analysis work occurred *after* teachers administered the assessment. Teachers shared with us that they wished that we had discussed how to analyze student work before administering their assessment. Our intention was to focus on using student work to inform their learning. However, we unintentionally did not meet teachers' need to evaluate student work with a 3D lens.

In the revised course, teachers used their ideal student responses to develop a feedback tool to analyze student understanding and provide students with feedback on how they could improve. We provided teachers with options for providing feedback, which included evaluating the presence of features of the ideal response and developing 3D leveled rubrics for each prompt or the assessment as a whole. We hoped that these revisions would enhance the three-dimensionality and specificity of teachers' scoring guidance.

Supporting more accessible assessments. Due to the challenging task of designing phenomenon-driven assessments, we hesitated with adding the additional layer of incorporating opportunities for diverse sensemaking in teachers' assessments. Although ideas related to accessibility were discussed in earlier sessions, it was not until the last session that teachers used the *Supporting Diverse Sensemaking Checklist* to evaluate the accessibility of their assessments and consider future revisions. Teachers found the checklist helpful and wished they had access to it earlier to inform their design work, as many teachers had students with diverse needs. When revising the course, we incorporated assessment checks related to diverse sensemaking in all our tools so that teachers are considering the accessibility of their assessments throughout the design process.

Conclusion

Our pilot study documented the specific challenges that rural teachers faced when designing tasks that elicited students' understanding of the 3Ds in ways that connected to their interests and identities. Analyses of teachers' engagement with the course led to important revisions to the design of the 5D course. These revisions included 1) grounding our visioning work in what the 5Ds looked like in the context of both instruction and assessment, 2) more extensive focus on teacher learning related to the five dimensions; 3) choosing more productive and accessible exemplars to make visible key features and design thinking; 4) enhancing teachers' access to and usability of online tools and methods for collaboration; 5) integrating student interest and accessibility more explicitly in all aspects of the development process; and 6) more explicit support for designing and assessing 3D sensemaking opportunities. These revisions were made possible through learning from teachers' experiences and feedback about the course. Herrmann Abell, Lo, et al (2024) will discuss the impact of these revisions on teachers' ability to design 5D assessments.

Paper 2: Examining Science Teachers' Conceptions of Student Interest as a Consideration in Designing Assessments

Keelin O'Connor, Anna-Ruth Allen, William R. Penuel*
University of Colorado Boulder

Abstract

A key goal of science education articulated in *A Framework for K-12 Science Education* is to create opportunities for students to answer questions about the world that connect to their interests, experiences, and identities. When developing assessments, teachers have the opportunity to develop scenarios that will connect to their students' interests, yet we do not understand well whether and how teachers consider student interest in assessment task design. In this paper, we analyzed data from a pilot study focused on preparing teachers to design five-dimensional tasks that connect to student interests and identities. Our data sources were teachers' descriptions of their design decisions about the phenomena used to anchor assessment, their designed assessment tasks, and interviews with them about those decisions. We found that interest was accepted as an important consideration for assessment design among teachers in our study, but they considered student interests in different ways that changed over time. Specifically, we found that some teachers shifted their views of what it meant to engage student interest in the context of assessment design over the course of their participation in professional learning, and that most teachers made decisions about what they believed their students were interested in based on their own conception of student interest or based on their knowledge of students or beliefs. Course tools to support elicitation of student interest and community priorities provoked changes to some teachers' thinking, even if they were not widely used as supports for designing assessment tasks.

According to *A Framework for K-12 Science Education* (National Research Council, 2012), a key goal of science education is that teachers regularly provide opportunities for students to systematically investigate issues and questions that relate to their interests. One motivation for connecting science instruction to students' interests in school is that it can help develop a "sustained attraction for science" and help students appreciate how science can be "pertinent to their daily lives" (p. 28). Another reason is that it can be a tool for promoting equity, to the extent that the interests of students from systematically marginalized underrepresented groups are prioritized (p. 28). A third reason is that cultivating interest is a potentially powerful strategy for broadening participation in STEM, since early interest in science is associated with pursuing advanced coursework in STEM in high school and beyond (Maltese & Tai, 2011; Tai et al., 2006).

While research on student interest in science is abundant (see Potvin & Hasni, 2014 for a review), research on how teachers conceptualize and learn about students' interests is not. It is not uncommon for teachers to elicit students' relevant interests and experiences when presenting science phenomena to students at the beginning of a unit or exploration of a topic (e.g., Cowie et al., 2010; Patro, 2008). It is also true that teachers attempt to connect what students are learning with interests that teachers imagine students to hold about a topic as they present it (Hagenah & Thompson, 2021). Some students have reported that their teachers actively shape their instruction around interests they've learned about through their relationships with students (Basu & Calabrese Barton, 2007). What is missing from this literature, though, is research on how teachers consider interest in the design and planning of instructional and assessment tasks.

Assessment design is an important opportunity for studying whether and how teachers consider interest as they plan activities for the classroom. That is so, in part because teachers' classroom assessment practices affect students not only cognitively, but also affectively (Cowie et al., 2010). Further, there is evidence that when students feel a connection to scenarios presented in tasks, their performance is enhanced (Taylor et al., 2016; Walkington, 2013). To be sure, it is also common that there is poor alignment between theories of interest and motivation and the assessments used in classrooms and in the high-stakes tests on which teachers model their assessments (Shepard et al., 2018).

With the intention of supporting teachers engaged in designing assessments to connect to students' interests, we undertook a design-based research study of an online course for rural secondary science teachers. In the course, we introduced tools for eliciting students' interests and using them to inform the design of their assessments. As part of the research on the course, we became aware of and curious about the different ways that teachers were conceptualizing interest and considering its role in the design of assessment tasks. In this study, we consider evidence from the assessments they designed for the course, their rationales for their design, and interviews to answer two questions: *How did teachers conceptualize their students' interests? How did they consider student interest when designing an assessment task?*

Key Characteristics of Next Generation Science Assessments

The consensus volume, *Developing Assessments for the Next Generation Science Standards* (National Research Council, 2014), articulated a number of major changes to assessment from those commonly in use at the time. These include the need for assessment tasks that include multiple components that, taken individually might assess only one or two dimensions but, when considered as a whole need to support inferences about how students are making connected use of science and engineering practices together with disciplinary core ideas and crosscutting concepts. Such assessments would need to reflect the learning progressions articulated in the *Framework* and in the Next Generation Science Standards (NGSS; NGSS Lead States, 2013), which was developed to embody the vision of the *Framework.* They would also, the committee concluded, need to include ways to interpret variable student responses and tools to inform next steps of instruction.

The report also indicated the need for the questions in a task to be "interrelated" (NGSS Lead States, 2013, p. 3), and while phenomena and problems were not named as the basis for constructing tasks, subsequent guidance from the field did articulate the need for tasks to be anchored in phenomena (Achieve, 2018; NextGen Science & EdReports, 2021). That is to say, inferences about students' ability to *use* the three dimensions are expected to be based on their ability to do so while making sense of phenomena or problems they have not seen, but that are related to phenomena and problems they have studied (Penuel, Turner, et al., 2019). Here, phenomena refer to "observable events that occur in the universe and that we can use our science knowledge to explain or predict" (Achieve et al., 2017), while problems refer to concrete design challenges that students use science and engineering knowledge and engineering practices to address. To promote accessibility, phenomena or problems are presented to students using multiple modalities and given sufficient information needed to use what they have learned up front or as part of the task (Achieve, 2018; Furtak et al., 2020).

It is not sufficient, though, for tasks to simply present phenomena or problems to students. The scenario used to describe the phenomena needs to *problematize* the phenomenon or problem, that is, make clear what is puzzling and what is "at stake" in making sense of the phenomenon or problem (Reiser, 2004). This is so for two reasons: it helps to motivate students' own efforts to answer the interrelated questions that follow, and it helps give coherence to the assessment task itself. This same strategy is used to establish coherence in instructional units (Reiser et al., 2021) and to motivate persistence in learning (Blumenfeld et al., 1991).

What is presented in the task—whether it is a qualitative description of problem or phenomenon, or a set of data to analyze—needs to provide students with opportunities to use targeted elements of all three dimensions of science learning present in the standards: disciplinary core ideas, science and engineering practices, and crosscutting concepts (National Research Council, 2014; NextGen Science & EdReports, 2021). There are many phenomena that are potentially interesting and compelling to students and potentially useful for eliciting student understanding of targeted standards; however, not all tasks present students with the chance to use the three dimensions together to make sense of phenomena and problems. In addition, some descriptions of phenomena and problems are inconsequential for being able to answer the questions that follow. Framing phenomena and problems, as well as the questions about them, to elicit students' understanding of the three dimensions, is challenging in designing assessment tasks.

Meeting the demand that assessments consider where students are in a targeted learning progression in the standards demands attention to information provided in the NGSS Appendices (E, F, and G) regarding grade-band expectations. These Appendices specify key "elements" that are important targets of assessment (NextGen Science & EdReports, 2021). That is so, because they are intended to build over time—that is, to mark key steppingstones in the development of more sophisticated understandings of disciplinary core ideas, science and engineering practices, and crosscutting concepts over multiple years. A key assumption of the *Framework* is that students' understanding of the dimensions grows incrementally over the course of many years, and that students will need multiple opportunities to develop three dimensional science proficiency (National Research Council, 2012, p. 26). Assessments that reflect this assumption thus provide information useful to teachers and leaders as to student progress toward those multi-year learning goals, particularly if they also provide opportunities for providing students with feedback on their progress (National Research Council, 2012).

Accessibility and fairness are integral features of assessments of the Next Generation Science Standards as well (National Research Council, 2014). Tasks should be aligned to what students have had the opportunity to learn; put another way, tasks should target performance expectations that students have had opportunity to achieve, to demonstrate their ability to understand and apply the knowledge they developed as part of a unit or sequence of learning activities. Further, tasks need to be accessible to a range of learners, including those who are emergent multilingual learners (Fine & Furtak, 2020).

A novel attribute of some NGSS assessments is that they are intended to be *interesting* to students. This intent derives from the general importance of interest to persistence in learning (National Academies of Sciences Engineering and Medicine, 2018), as well as to the emphasis in the *Framework* on connecting what students study to their own interests, experiences, and identities (National Research Council, 2012, p. 28). What makes an assessment task potentially interesting to a student is, in principle, no different from what might make tasks interesting to students more broadly, features such as novelty, complexity, and incongruity (i.e., problematization) (Berlyne, 1970; Harackiewicz et al., 2016), as well as making clear how tasks might be relevant or useful to students (Hulleman & Harackiewicz, 2009). As we elaborate below, there are several potential approaches to discovering and eliciting interest from students that could inform the design of assessment tasks that focus on phenomena or problems that students find interesting and that are presented as a way so as to elicit or trigger student interest. However, we have little understanding today of how teachers conceive of the goal of making tasks interesting, and how they go about it; hence, the focus of the current study.

Conceptual Framework: Designing Assessments that Connect to Students' Interests

Teachers can potentially increase students' engagement with assessment tasks and performance within them by making them more relatable to students, that is, by connecting them to their interests (Kang & Furtak, 2021). From the perspective of our framework, there is no single way to help make tasks interesting to students; rather, there are multiple ways, each with both affordances or possibilities, and each with risks and possible downsides to be considered. Below, we describe three broad strategies that past scholars have used to design tasks—primarily for instruction—that connect to students' interests.

Eliciting Students' Interests

For several decades, scholars in science education and other disciplines have used techniques drawn from the funds of knowledge approach (Moll et al., 1992) to design learning experiences that build bridges between students' everyday lives in their families and communities with disciplinary learning. Funds of knowledge refer to the bodies of knowledge and skills that have been historically accumulated and culturally developed that are essential for individual and household functioning and well-being (Vélez-lbáñez, 1988; Vélez-lbáñez & Greenberg, 1992). Teachers, in the fullest expression of this model of teaching, are first researchers, who are introduced to qualitative methods of study, including ethnographic observations, developing questionnaires, writing field notes, interviewing, and data management and analysis (González et al., 1993).

One strategy within the funds of knowledge is to use surveys to elicit sources of family and community expertise that could be relevant to teaching. As an example, one group of investigators created a funds of knowledge survey to give out to undergraduate engineering students. The purpose was to elicit their knowledge and experience in three domains relevant to engineering: tinkering, perspective taking, and helping mediate conflict between people with different points of view, with the intent of helping build students' connections to practices of engineers (Verdín et al., 2021). In another design study (Tzou & Bell, 2010), researchers guided teachers to elicit student interest through the technique of photo-elicitation (Clark-Ibañez, 2004). In that study, students formulated questions to pursue in class after taking photos of ways their families stayed healthy. The teacher integrated students' questions into a kit-based science unit that had been chosen after ethnographic research revealed prevalent chronic health problems among students and their families in the school community, and so it was judged to be of great potential relevance. The study found that teachers could skillfully elicit student questions and use them in instruction, but they did not always align with disciplinary learning goals.

This and other studies focused on teachers' elicitation of students' familial and cultural practices have identified challenges with doing so. It is possible, for example, to reinforce essentialized and even deficit notions of particular communities, in terms of their interests (Kirchgasler, 2019; Llopart & Esteban-Guitart, 2017). Further, students' own practices may be conceptualized as nonscientific (Sengupta-Irving et al., 2021) or present tensions for teachers as they attempt to help students navigate between everyday and dominant scientific ways of thinking (Braaten & Sheth, 2017). Therefore, it is critical to attend to variation *within* groups (Gutiérrez & Rogoff, 2003) when making sense of survey data and to build a classroom culture that values heterogeneous ways of knowing and an expansive conception of what counts as science (Warren et al., 2020).

Connecting to Places and to Lands and Waters

Another strategy for promoting interest is to connect learning specifically to a local area that students know well. Place-based education emphasizes the need to ground teaching in local phenomena, where students can easily make personal connections between what they are studying and the people, nonhuman kinds, and natural and built environments they know through their direct experience (Gruenewald, 2003; Smith, 2007). A key assumption within place-based education in science is that grounding teaching in the local activates students' affective relationships to particular places, that is, their sense of place (Haywood, 2014; Semken & Freeman, 2008). Place-based education in science can also invite students into grappling with critical priorities in their communities, such as environmental racism and climate justice (Aikenhead et al., 2006; Eppley, 2017; Morales-Doyle, 2017; Segura et al., 2021; Zimmerman & Weible, 2017).

Indigenous educators have advocated for the use of the terms *lands* and *waters* to refer to places where Indigenous tribes assert sovereignty (e.g., Bang et al., 2012). Land-based pedagogies seek to sustain and repair relations between human communities and lands and waters, as well as with more-than human relatives in such places (Calderon et al., 2021; Corntassel & Hardbarger, 2019; Styles, 2019). Centering learning within lands and waters can help bring to the foreground young people's responsibilities, the impacts of prior decisions on the land, and the ethical need to anticipate consequences of present decisions (Learning in Places Collaborative, 2022). Land-based pedagogies necessarily invite a critical reckoning with colonial legacies of places and how those legacies are alive within current relationships and with views of what counts as knowledge (Bang et al., 2012). As such, they are more than simply about interest in the sense of a personal connection to a topic in science, such pedagogies invite a reimagining of relations between learners and the places and communities where they live, though they share with place-based pedagogies a concern for cultivating affective and ethical relationships to specific places.

While place-based educators have suggested that such approaches can be made congruent with current standards-based approaches to teaching science, others have expressed skepticism about the prospect. Those who do suggest it is possible to do so emphasize the need for flexibility in interpretation and bundling of standards (Semken & García, 2021) and for alignment to standards that reflect Indigenous values (Aikenhead et al., 2006). Those that are skeptical point to the fact that many of the standards are silent on the role of human activity in shaping the very phenomena and problems that might be studied in science (Morales-Doyle et al., 2019). Choosing sociopolitical issues to focus on, as well as considerations about how to integrate different ways of knowing and Indigenous perspectives, require considerable care, and they benefit from close ties between schools and communities. In practice, there are often strong cultural disconnects between schools and communities (Aikenhead et al., 2006).

Developing Deep and Caring Relationships with Students

There is a strong evidence base supporting the idea that deep and caring relationships between teachers and students are valuable for motivation, learning, and civic outcomes (Schindel & Tolbert, 2017; Valenzuela, 1999; Wanders et al., 2020; Wentzel, 1997; Wubbels & Brekelmans, 2005). Caring relationships that couple high expectations and supports to meet them, alongside an appreciation for the specific ways that discrimination and marginalization shape the lives of youth from systemically marginalized groups (e.g., Black students, LGBTQIA+ students), can help establish bonds that build students' commitment to their own learning and a sense of belonging (Antrop-González & De Jesús, 2006; DiNicolo et al., 2017). Enacting critical care in this way can also support teachers gain self-awareness and relate in more skillful ways to their students in ways that support students' identity development (Kumpulainen & Rajala, 2017; Trout, 2018).

Although relationships are widely understood to be at the heart of teaching, it is often assumed that they develop without needing explicit planning, attention, or effort (Kim & Schallert, 2011). However, how teachers respond with care in moments of vulnerability in science classrooms reveals that it requires intentionality and a commitment to building and maintaining a culture in which people show care for one another (Krist, 2020; Krist, in press; Krist & Suárez, 2018). When students perceive there to be a culture of caring in the classroom, students are, moreover, more likely to show interest in science (Singleton et al., accepted, with major revisions). Further, planning for relationship-building can help to improve teachers' knowledge of students, as well as and students' relationships with teachers. A study by Potvin (2021) used a modified form of a "cultural probe" (Gaver et al., 2004), a design technique for eliciting an inspirational, personal response from someone that reveals something about their preferences, with a group of teachers as part of cycle of teacher inquiry. The approach shifted teachers' practices and gave them a better understanding of students they did not know as well in their class.

There are also challenges with approaching intentional relationship-building as a means for connecting instruction to students' interests. A challenge is that purposeful reflection on relationship quality can be emotionally challenging for teachers, if confronted with having to reflect on different interactions or student feedback on the quality of their relationships (Krist, in press; Potvin, 2018) or on how well lessons connect to students' personal interests (Penuel et al., in press). Further, how to integrate knowledge of students into instruction in science presents similar challenges as other strategies for eliciting interests, in terms of identifying opportunities to link such interests to standards.

The Current Study

The current study is a qualitative, descriptive study focused on teachers' ideas about the importance of interest in assessment task design and how they considered interest when developing tasks intended to elicit three-dimensional performance expectations of their own choosing from the NGSS. It is part of a design-based study (Design-Based Research Collective, 2003) of an online course designed to help teachers construct assessment tasks that reflect the vision of the *Framework for K-12 Science Education* (National Research Council, 2012) for teaching, learning, and assessment. The larger project is also undertaking an experimental study of the course to examine whether it can improve the quality of teachers' assessment tasks. This study of teachers' own reflections on their tasks and task design, derived from artifacts they provided in the research and interviews, was intended to inform future iterations of the online course.

Design of the Online Course

The focus of this course was on helping secondary (grades 6-12) science teachers from rural areas to develop what we refer to here and elsewhere (Penuel, Turner, et al., 2019) as *transfer tasks.* Transfer tasks are multicomponent tasks to be administered at the end of a unit and that target performance expectation(s) that were focal in the unit. It is a test of students' ability to transfer knowledge (Bransford & Schwartz, 1999), in that they have to make sense of and answer questions about a new phenomenon or problem they have not yet seen using elements of the three dimensions from the performance expectation being assessed.

The initial version of the online course built from an earlier, briefer, face-to-face professional learning workshop (Penuel, Lo, et al., 2019), and it had five broad learning goals. First was to support teachers in "unpacking" or analyzing three-dimensional understandings of a focal standard or performance expectation (Krajcik et al., 2014). The second was on helping teachers understand what phenomena and problems are and how they can be used to provide coherence to assessment tasks. The third was to have them develop scenarios where a phenomenon or problem is presented in a way that will be interesting to them and productive for eliciting and demonstrated targeted understandings. The fourth was to develop assessments that prompt students' integrated use of the three-dimensions to explain phenomena. The fifth and final goal—which received somewhat less attention—was to support teachers in developing tasks that their students would find interesting.

The course was designed to be online via Zoom to ensure that participating teachers had access to the professional learning without the need to travel. Online tools were created to facilitate the sharing of resources. In total, participating teachers engaged in 25 hours of professional learning spread across 3 months. Each session was scheduled to be 2.5 hours, with select sessions lasting 3 hours due to the complexity of the work. Teachers were grouped in small groups based on grade band and content area to allow teachers the ability to collaborate. Teachers received feedback on their work from their peers and a coach with expertise in the content area. Breakout rooms were used to facilitate small group discussions. To help teachers apply what they had learned to their classrooms, teachers focused their assessment work on a unit that they would be assessing at the beginning of the last month of the course. Sessions were scheduled to allow time for teachers to work and receive feedback on their work from a member of the research team. Teachers had a three-week window to administer their assessments to students and collect student work to reflect on their assessment work.

As part of the course, teachers administered to students a Student and Community Interest Inventory at the beginning, which was intended to support their consideration of interest in the choice of phenomena or problem to anchor assessments. The student interest survey included several questions relating to students' experiences and interests learning science, as well as a question connected to local community issues: "What are two issues that members of your community are concerned about? For instance, students in California might be worried about wildfires." After receiving survey responses from their students, teachers analyzed responses in the 5D class. Teachers were encouraged to draw on results from the survey in their assessment design.

Sample

A key focus of the project was to support rural teachers, whose access to sustained professional learning opportunities in science is often limited (Zinger et al., 2020). A total of 11 middle and high school science teachers from rural districts participated in the study. Of these, three were middle school teachers, and eight were high school teachers. They came from two high schools and two middle schools in mountainous rural communities where students' families work primarily in service occupations serving vacationers. Among the high school teachers, two taught biology, two taught chemistry, one taught physics, one taught Earth science, two taught ecology or environmental science, and one taught botany/zoology and biotechnology. Of the middle school teachers, all taught integrated science.

We received data from a total of 328 students in teachers' classes that we used to describe the backgrounds of students. Of those that provided information about their gender, 141 identified as girls, 163 as boys, and 5 as gender nonbinary or genderfluid. Among those that identified their race or ethnicity, 201 were white, 95 were Latine, seven were Asian or Asian American, five were Native American, and one was African American/Black. Six identified as bi- or multi-racial. A total of 213 students said they spoke primarily English at home, while 18 spoke primarily Spanish, and another 78 said they spoke both English and Spanish. Other languages spoken in the home were French (n = 8), German (n = 2), Vietnamese (n = 2), Japanese (n = 1), Russian (n = 1), Polish (n = 1), Hebrew (n = 1), and Zapotec (n = 1).

Measures

We drew on three different sources of data for the study: teacher-designed assessment tasks, a submission form that explained their design decisions and rationale, and teacher interviews.

Teacher-designed assessments and assessment submission form. After completing the course, Teachers were asked to submit a 5D assessment they gave to their students, that they created or modified outside of the 5D course. Teachers submitted their post-assessments using an assessment submission form, in which they were asked questions about how they developed their assessment, including "When designing/modifying this assessment, how did you use what you know about your students to make it ENGAGING and RELEVANT for your students?"

Teacher interviews. We conducted semi-structured interviews with each individual teacher after the online course was complete. In interviews, we followed a semi-structured protocol, where we asked teachers to expand on topics related to their experiences in the class, take-aways from the course, as well as opportunities and challenges they see in implementing the 5D course vision in practice. We recorded all interviews and created transcripts using Zoom.

Approach to Analysis

We conducted analyses first to characterize teachers' assessments themselves, as to whether the assessment had a phenomenon or not, and if so, if the phenomenon was local or explicitly related to students' responses to the Student and Community Interest Inventory as interpreted by teachers.

We next used a mix of theoretically driven and open coding of teachers' responses to the question about how they used what they knew about students to focus their assessment. Theoretically driven codes pertained to whether teachers referenced data from formal methods for eliciting interest (e.g., the inventory or a survey they gave), referenced a focus on place or land, or referenced their knowledge of students. Open-ended codes developed from the data were three: claims that "everyone loves" a particular topic (or science in general), "implicit" and "don't mention interest." In the "implicit" responses, teachers did not state why a certain phenomenon or topic might be of interest, or they assumed that a certain topic would be of general interest. Slightly different from implicit assumption of student interest in a phenomenon, "everybody loves" responses involved explicitly statements that a certain topic was of universal interest to all people or students.

With respect to interviews, we first identified mentions of interest or of identity in the verbatim transcripts of interviews, and then conducted the same integration of theoretically driven and open coding of the data to identify specific topics related to how teachers considered and learned or knew about student interests. For each topic, we conducted a thematic analysis of responses (Braun & Clarke, 2006). To further illustrate these themes, we developed portraits of three teachers whose conceptions and approaches to using interest varied.

Findings

All the teachers reported some consideration of interest in their design of assessment task, though two of the 11 teachers did not anchor their assessments in a phenomenon at all. Of those who did consider interest, as Table 1 below shows, the most common consideration was teachers' reported knowledge of their students (n = 5). Of the other theoretically derived codes, only one teacher referenced using data from the inventory, and one used the idea that the topic was locally relevant was a consideration. The one teacher who did use data from the inventory noted the phenomenon they chose was related to what students said was interesting on their inventory. Otherwise, the remaining phenomena (n = 8) were unrelated to data on the inventory, and all the phenomena (n = 9) were nonlocal, that is, not specific to the place or lands where students attended school.

Below, we elaborate on the conceptions and considerations related to interest in task design for three different teachers, Katie, Chris, and Evelyn. We chose these teachers because they represent different approaches to considering interest. Katie said that she valued a focus on interest and thought that local phenomena would be most interesting but chose a nonlocal phenomenon for her assessment. Chris said his ability to consider interest was constrained by his curriculum, and—and as interviews revealed—he was resistant to taking up the specific aspects of task design promoted in the course. Evelyn, by contrast was compelled to consider student interest, and she did so from the perspective of her reported knowledge of students and consideration of place in ways that were inspired by the course tools.

Katie

Katie is someone who held that what made something compelling in science was that it relied on authentic, local datasets; nonetheless, she chose a phenomenon for her genetics assessment that is common for this set of big ideas in science, a rare genetic disease. According to Katie, using a real data set "opens the door to research or studies that are being done in your local area." By employing real datasets – something emphasized as a criterion for good assessments in the course, "you've hopefully got the interest piece also because it's real, it's authentic." But the phenomenon she chose for her assessment did not include local datasets at all. In her assessment, she presented students a TED talk given by an adolescent suffering from progeria, a rare genetic condition that causes rapid aging in children and youth. Her rationale focused on the potential for the phenomenon to show the importance of resilience in the face of adversity, that is, how the adolescent "cope with the disease yet still have a happy life." Notably, this ethical lesson is framed in the teacher's terms, rather than in terms of student interests or concerns.

Table 1

Teachers' Phenomena and Ways of Considering Interest in Assessment Task Design

Teacher	Phenomenon	Reasoning Related to Interest	Illustration of Reasoning
Bella	*What happens to matter when you make ice cream* Nonlocal Unrelated to inventory	Everyone loves	All students love food, specifically ice cream.
Maria	*The Blue Fugates* Nonlocal Unrelated to Inventory	Knowledge of Students	My kids like the strange the out of the normal the things that we don't experience up here in [place]. They like playing detective and piecing things together especially when it comes to genetics and the human condition.
Katie	*Rare genetic disorder as experienced by a teen* Nonlocal Unrelated to Inventory	Implicit	They learned about the disease progeria by watching a TED talk by an adolescent The TED Talk focused on his life as a high schooler and how he can cope with the disease yet still have a happy life.
Lonnie	*How fireworks get their colors* Nonlocal Unrelated to Inventory	Everyone Loves	The phenomenon was engaging to students because everyone loves fireworks.
Darla	None	Implicit	The students seemed very interested in trying to understand why a natural resource can have different properties then when combined with other natural resources.
Chris	*Demonstration of a rollback car where two wheels connected by pieces of wood and a rubber band between the wheels* Nonlocal Unrelated to Inventory	Everyone Loves	My students are interested in science, so I tried to use the demo of the rollback car as an interesting phenomenon which can then show their understanding of energy.
Brooke	*PKU results in people not being able to digest protein and developmental differences* Nonlocal	Implicit	interest in genetic diseases and mutations have been shown so I chose this topic for this reason

	Unrelated to Inventory		
Fran	*Explaining why Chile home to the largest earthquake and longest mountain range* Nonlocal Unrelated to Inventory	Knowledge of Students Implicit	I knew they wanted to learn about rocks in different places and since it was a real example of a location that most of my students had heard about, I figured it was engaging and relevant.
Nancy	None	Knowledge of Students	I knew that students were already engaged in the topic about ozone and environmental concerns so this provided an extension of that topic.
Evelyn	*Why Earth is habitable but other planets are not* Nonlocal Unrelated to Inventory	Knowledge of Students Connection to Place/Land	They also had many questions about Venus and Mars while we were studying the atmosphere part of this unit. The modeling fit in perfectly. The knowledge-based questions are based on phenomena they see locally.

One reason why Katie did not turn to tools in the course like the Student and Community Interest Inventory to learn about her students' interest was that she did not think her students were knowledgeable about their interests. She commented that asking students, "What are you interested in learning about?" was not "best way to approach" the problem of connecting teaching and assessment to students' interests. Katie said that her students, when asked what they were interested in, often did not know. Katie trusted more the idea that a local phenomenon would interest her students:

> By making the learning relate to things that are happening currently or locally, I think that you can't help but capture student interest and by creating lessons and assessments where the students are acting and thinking like a scientist, whether they know it or not, they're creating a level of identity as a person who knows how to do science.

Even though she did not use or value the Student and Community Interest Inventory, she said it did provoke her thinking, but did so in a way that pushed her toward thinking that engaging students in using science and engineering practices to engage with local phenomena would be best for "engaging their interest in helping them create a sense of identity."

Chris

Chris illustrates a teacher who sees interest as important and a valuable part of his science teaching. Interest, he says in his interview, is a given in his instruction, "It's got to be there." This view is consistent with the justification for his phenomenon choice of a rollback car: "My students are interested in science, so I tried to use the demo of the rollback car as an interesting phenomenon which can then show their understanding of energy." Here, any phenomenon might have followed the statement, "My students are interested in science," so long as the phenomenon was a scientific one, but for him, interest in science is a given, and the topic doesn't necessarily matter for garnering and sustaining students' interest.

Chris did appreciate two phenomena that were introduced in the course for how interesting they were to him. One was a time-lapse video showing an animal decomposing over several days on the side of a road, and a second featured data about a swallow population's adaptations to a new highway built through their habitat. He appreciated that these phenomena were "more than a cool hook" and could motivate learning through instruction or assessment. At the same time, Chris said that he prioritized creating a caring environment—that is, one that is "caring, loving, and good" in his own classroom over the goal of adjusting his curriculum and teaching to promote interest in science.

In fact, Chris saw his curriculum as a constraint limiting his ability to adjust his teaching to reflect student interests. In fact, he suggested that the materials themselves constrained the topics he could teach but were also inherently interesting to students:

> I love knowing what my students are interested in in the essence of, "It was kind of cool." My students, when they did this [survey, completed as part of the course], they were like, "We kind of like science." That was kind of nice to hear, but I'm still doing PEER physics and I'm not going to build my class around say an interest in space even if ten kids have it. For me, that's just not going to happen.

At the same time, he acknowledges a pull toward adjusting his teaching to focus on topics related to students' interests but says that it's not possible for him to do. He says,

> I would love to be that teacher that could pull that off. I can't do it. It's kind of like we got this, we're going to roll with this, I'll make it as good as possible, but if you're a kid who just love space, sorry, dude, you're here.

Chris does not elaborate on how he would "make it as good as possible," meaning interesting to students, in his response, though it's evident he sees it as his responsibility to do so within the constraints of the curriculum he's implementing. For him, the class did not help deepen his conception of interest: "I don't know if—no offense to the class—I don't think the class really made me think more about interest."

Evelyn

In contrast to Chris, Evelyn is someone whose ideas shifted through the course about interest. Initially, she admitted in her interview, "I didn't care about their [her students'] interests." By the conclusion of the course, however, there was consideration of interest evident her assessment submission form. Further, we characterized her consideration of students as based on her knowledge of students and their questions, and as connected to local phenomena. Her phenomenon focused on the uniqueness of Earth's habitability in comparison to other planets in the solar system. She chose that, because she had observed that students "had many questions about Venus and Mars while we were studying the atmosphere part of this unit." She commented, too, that she incorporated questions to elicit students' understanding of disciplinary ideas "based on phenomena they see locally."

Although she did not use the Student and Community Interest Inventory to select the phenomenon used in her assessment, it did lead to a shift in her thinking. She commented,

> But with that interest survey, it was just super cool…That interest survey was really enlightening to [inaudible] what they thought of science and how they, I guess didn't think they were a scientist at all. I appreciated that.

From this response, we inferred that Evelyn did in fact care that students identified with science but was surprised to find that they did not. Further, Evelyn said she had begun to use interest surveys in her class, in a "pared down fashion," as part of units she is teaching.

The online course also gave her a different way of thinking about science teaching, according to her interview. Her new vision for science teaching was that it should be more "holistic" and connected both to students' own questions and to things students care about:

> To me, it's more holistic way of teaching science, which is how it's supposed to be taught, not just in isolation and this little bit in class. But, how do they see it in their terms and in their lives, and then how do I access that? Not just like, "Here's the content." But, here's why that matters in a bigger scheme and connect that to social studies and, "Hey, remember this from middle school, remember that storm we had a week ago." Or, back to questions that they had. And so, I guess it's not just content, it's not just skills. It's really making it important to them and what they care about.

Here response in this respect contrasts sharply with Chris' response, in that she does not take content as a given, but as something that should be adjusted to accommodate students' interests and to allow for students to connect what they are learning in science to other disciplines, to what they have learned in the past, and to their own experiences.

By the end of the class, she was attending to students' questions more, and beginning to tailor her teaching more to those questions. She commented:

> I really loved hearing things that they were questioning. Because then I could answer that later, I could come back to it and say, "Oh, Mars soil. Let's talk about that for..." And, it allowed me to gear my instruction and even some of those phenomenon towards their wondering statements.

As noted above, it was these questions that inspired her choice of phenomenon for the assessment task she designed for the course. In sum, Evelyn is someone whose consideration of interest shifted significantly in ways she attributed to encounters with a new kind of teaching and with opportunities to try out tools from the course.

Discussion

In this design study focused on how to support teachers in developing assessment tasks that embody the vision of *A Framework for K-12 Science Education* (National Research Council, 2012), we found that most teachers' considerations of interest were based on their own ideas about what all students were interested in or based on the idea that all students liked science. Despite presenting and giving teachers and experience of a vision that emphasized the importance of connecting instruction and assessment to students' experiences and interests and tools for doing so, most teachers did not use any of the three broad approaches identified in our literature review—eliciting students' interests, connecting to place or lands and waters, or developing knowledge of students—to inform the design of their assessment tasks. From our perspective as designers, this result was disappointing, but not entirely surprising. We know that it is difficult for any designer of instructional materials to hold and balance the goals of writing tasks to standards while also considering students' interests and concerns about issues in their communities. Addressing both these goals requires teachers to apply both an understanding of the standards themselves (Krajcik et al., 2014) and their direct knowledge of students' interests and community priorities. The challenges of shifting to engage students meaningfully in science and engineering practices through tasks may eclipse a focus on making connections between tasks and students' funds of knowledge (Carpenter et al., 2023). Teachers are largely left on their own to "connect the dots" among new three-dimensional goals for science learning, students' everyday lives, and teaching (Madkins & McKinney de Royston, 2019). Just as designers do, teachers need well-structured tools and practices to do so (Penuel, Allen, et al., 2022; Penuel, Reiser, et al., 2022), tools and practices that themselves often require iterative re-design before they can prove successful.

Each of the focal case study teachers was influenced by tools and practices introduced in the course, but not necessarily in the ways that we as designers intended. Notably, all three teachers mentioned the Student and Community Interest Inventory, but they oriented to it in different ways. Both Katie and Chris were provoked by the inventory's introduction in the course, but both rejected it as a tool. Katie believed it would be an ineffective means of eliciting students' interests, and while Chris did not challenge this aspect of the tool, he did not believe he could or should use the tool to adjust the content of his teaching. Evelyn used the tool, and even she did not use it to help select a phenomenon for her assessment; in fact, in the sample, only one teacher, Lisa, used the tool for this purpose, but she did not elaborate in her interviews as to why. Other activities did provoke change in teachers' conceptions of student interest. For Katie, it was the idea that tasks should use authentic data so that students got a feeling for what grappling with phenomena with real-world data might mean. And for Evelyn, the presentation of the vision of teaching and learning from the *Framework* gave her a more "holistic" vision of science teaching, as she put it, that gave more room for student questions.

It is tempting to interpret resistance to the tools from our point of view as designers, but an actor-oriented point of view (Lobato, 2012) can be more useful for informing iterative redesign of tools and practices. Although we could see Katie's judgment of her students' inability to articulate their interests as viewing her students from a deficit lens (Valencia, 2010), we choose to interpret her response as a failure of our tools to elicit interests in a way that could be useful for her. An alternate way to elicit interests used in our instructional materials design work (Penuel, Allen, et al., 2022), might be to first frame a set of potential phenomena to use in an assessment and then to elicit students' interest in each. This approach provides the teacher with a "sandbox" of phenomena that are aligned with standards and that students rate in a way that helps the teacher know which one is best. We need a different approach to respond to Chris' resistance here, one that builds upon what he says he takes to be his responsibility—namely to make existing content "as good as possible" for students, that is, to make what is already in the curriculum as interesting as possible to students. While he might not agree to using a phenomenon for an assessment task that was chosen because it was based on evidence of student interest or was local, he might be able to incorporate a brief writing task where students construct their own ideas about why what they study matters to them, a strategy that has been found to enhance secondary students' interest in science (Hulleman & Harackiewicz, 2009). These could be incorporated into an exit ticket type of an assessment that takes students only a few minutes to complete (see, e.g., Penuel et al., in press).

The study did yield promising findings about other course-designed activities for helping to influence teachers' conceptions of interest in the context of learning to design assessment tasks. In particular, for all three teachers, having experiences of learning activities that reflect the vision of the *Framework* provoked teachers to consider some new possibilities for learning, most notably for Evelyn. The design of this activity was based on earlier efforts to introduce teachers to the vision of the *Framework* by putting them in "student hat" while experiencing three-dimensional science learning (Lowell & McNeill, 2020; McNeill et al., 2022). Our qualitative finding that these shifted teachers' beliefs is consistent with other research showing the power of engaging teachers as students in shifting their beliefs (see, e.g., Reiser et al., 2017). It may be that iterations of our course can capitalize on these shifts by encouraging teachers to make connections between their own experiences and that of students, in terms of helping them see the practical possibilities for connecting to students' interests in instruction.

We caution that connecting assessment tasks to student interests and experiences is no straightforward matter, either for teachers or for people with professional expertise in developing assessments. Connecting to students' interests and experiences is not a replacement for grappling explicitly with racism that reproduces exclusion and marginalization from science and engineering (Sheth, 2019). It is critical to ask to what ends we are recruiting student interest, and into what kinds of practices. To that point, we might better invite students to what science and engineering could be, rather than what those fields are now (Penuel, 2020). In addition, it is important to recognize that interests of students change (DiGiacomo et al., 2018), and they are also idiosyncratic (Azevedo, 2018). Identifying good instructional practices and tasks to spark and support interests is hard for teachers, when these interests diverge and change too quickly for the schools' organizational timescales. Even strategies that rely on connections to place and local issues are subject to change in this way and may not connect to all students equally. What may prove most scalable are tools and practices that, in the end, fit within the constraints of teachers' time for planning while also enabling students to have more say in what they study within the constraints of a standards-driven policy environment.

Conclusion

In designing both instruction and assessment, connecting to students' interests and experience is an important consideration, and teachers can call upon a variety of strategies for doing so. Specific tools vary with respect to how teachers perceive their feasibility and utility, however, for eliciting students' interests. Further, experiences of interest-connected instruction can be powerful vehicles for influencing teachers' thinking about how best to connect standards-based instruction and assessment to students' interests and experiences. Caution is needed, though, in considering exactly what we are recruiting students' interests for, and attending to the variety and changing nature of interests presents challenges in its own right. As a field, we need to continue to iterate on tools and practices that can support the vision of the *Framework* that all students should be able to pursue questions in science and engineering related to their own interests, concerns, and communities' priorities.

Paper 3: Investigating the Impact of a 5D Professional Learning Course on Rural Teachers' Assessment Practices

Cari F. Herrmann-Abell[1]*, Abraham S. Lo[1], Kevin Cherbow[1], Sara L. Cooper[2], April Gardner[1], and Keelin O'Connor[2]

[1]BSCS Science Learning and [2]University of Colorado Boulder

Abstract

This paper explores the efficacy of a revised three-month, online professional learning course designed to support rural teachers in designing five-dimensional (5D) assessment tasks that explicitly attend to students' interest and practice-linked identity development. We conducted a randomized control study with 55 rural teachers from 13 states. Before and after experiencing the course, teachers submitted an assessment they developed or modified to make it aligned with the 5D vision. The assessments were scored using a rubric we developed to measure the extent to which the teacher-designed assessments aligned with the 5D vision. Linear regression models of gain scores were used to evaluate the impact the course had on teachers' assessment practices. The treatment teachers had significantly higher gains than the comparison teachers in how they used phenomena for sensemaking, provided opportunities for students to use the three NGSS dimensions in integrated and grade-band appropriate ways, and designed to engage their students' interests. However, we did not see a significant treatment effect on providing 3D feedback to students and enhancing the accessibility of the assessment. These findings show the promise of the online professional learning course in supporting rural teachers in aligning their assessment practices with a 5D vision, but also highlights the persistent challenges that teachers face after experiencing in-depth professional learning.

This research study examined the efficacy of a revised three-month, online professional learning course designed to support rural teachers in designing five-dimensional (5D) assessment tasks that align with a vision for student learning and performance found in *A Framework for K-12 Science Education* (NRC, 2012). This revised course was informed by findings from design-based implementation research as reported in Lo et al. (2024) and O'Connor et al. (2024). Implicit in this multidimensional approach are two additional latent dimensions to the traditional three-dimensions: student interest and identity (Bell et al., 2016). 5D tasks involve opportunities for students to use the science and engineering practices (SEPs), disciplinary core ideas (DCIs), and crosscutting concepts (CCCs) in integrated ways to make sense of phenomena and problems that engage student interest and support students' identity development as knowers, doers, and users of science.

Designing 5D Assessments

In *Developing Assessments for the Next Generation Science Standards* (NRC, 2014), the National Research Council offered a set of recommendations regarding assessments designed for the NGSS. They recommended that assessments should (1) include multiple components to allow students to demonstrate their use of different scientific practices in the context of different disciplinary ideas and crosscutting concepts, (2) provide information that situates students' knowledge and mastery on a learning progression, and (3) include rubrics and other tools to help teachers interpret and use students' responses to adapt instruction and provide feedback to students. Through the Task Annotation Project in Science, Achieve (2018) identified "must-have" features of NGSS-aligned assessments that include: (1) a focus on real-world phenomena, (2) requiring sensemaking using the dimensions, (3) being fair and equitable, and (4) supporting the intended purpose. These resources informed important assessment design principles used in the online course.

Interesting real-world phenomena. 5D assessments tasks should present students with observations or information about a phenomenon or problem for students to explain or solve. This presentation often takes the form of a scenario where students are introduced to the data or information followed by a central question that drives student sensemaking. Scenarios should include an explicit invitation for students to wonder about the central question posed by the scenario that can compel student interest and motivate students to use the 3Ds for sensemaking.

Sensemaking using the dimensions. Prompts in 5D assessments invite students to demonstrate and use disciplinary core ideas (DCIs), science and engineering practices (SEPs), and crosscutting concepts (CCCs) in integrated and grade-band appropriate ways to answer the central question. Supporting integrated use of the 3Ds is important so that students do not view these dimensions as separate content or skills, but part of a broader ensemble of activity that can be used together for sensemaking.

Fair and equitable. 5D assessments should be explicitly developed in ways that enhance students' access to understanding what is to be explained and provide opportunities for students to demonstrate their understanding in ways that make sense to them (CAST, 2018; Fine & Furtak, 2020). Strategies to promote accessibility include presenting the phenomenon using multiple modalities, defining specialized vocabulary, using clear visual representations, and creating opportunities for students to represent their understanding in multiple ways.

Rubrics. For the results of assessments to inform instruction, developing or using 3D scoring guidance and rubrics that support opportunities for actionable feedback are necessary. Rubrics should provide more than just the correct response. They should be clearly connected to the task and describe specific qualities of thinking and performance that can support students in improving their use of the 3Ds.

Designing assessments that fully align to the 5D vision presents several challenges to classroom teachers. For one, they need to create opportunities for students to apply their understanding of DCIs, SEPs, and the CCCs to explain phenomena (for science) or solve problems (for engineering). Additionally, choosing phenomena that invite students to make sense of phenomena using related key science ideas and practices (sensemaking) and doing so in ways that are culturally relevant and personally meaningful to students is not a straightforward process (Lo et al., 2022; Penuel et al., 2022; Penuel, Lo, et al., 2019; Penuel, Turner, et al., 2019). The online professional learning course provided rural teachers with online tools and a learning community designed to help them face these challenges and develop their 5D assessment practices. This study sought to answer the research question, *To what extent did the 5D course impact teachers' assessment practices compared to business-as-usual?*

Study Design

Intervention

Using a design-based research approach, we revised a course designed to develop teachers' pedagogical design capacity to design 5D assessments (see Lo et al., 2024 for description of revisions). The course involved three sets of related sessions that answered the course level question, "*How can I design my classroom to align with this 5D vision for meaningful science learning and performance?"* Sessions 1-3 investigated the question, "What does 5D learning and performance look like?" Sessions 4-6 investigated the question, "How can we use phenomena to frame instruction and assessment?" Sessions 7-10 investigated the question, "How can we develop and use tasks to assess student understanding?" In total, teachers engaged in 25 hours of professional learning over a three-month period. Through these sessions, teachers used information from their students to choose viable and meaningful phenomena and choose data and information to make visible what was puzzling and what needed to be explained using the targeted 3Ds. Teachers then practiced designing an assessment scenario that presented the phenomenon to students and prompts to scaffold student sensemaking using the targeted 3Ds. Teachers also created a 3D feedback tool and used surveys to assess students' experiences using the assessment. After administering the course designed assessment to their students, teachers considered next steps for applying what they learned.

Data Sources

We conducted a randomized control study during the 2022-23 school year. Fifty-five rural teachers from 13 states participated in the study. Table 1 summarizes the demographic information for the teachers by treatment group. Within states, teachers were randomly assigned to the treatment or comparison groups. Treatment teachers (n = 22) participated in the course in Fall 2022 and comparison teachers (n = 33) received a delayed treatment in Fall 2023. To investigate the extent to which the course impacted teachers' assessment practices, we asked teachers to choose a unit for which they could develop or modify an existing assessment to make it more aligned with the 5D vision. Teachers submitted an assessment to us before (Spring 2022) and after (Spring 2023) the treatment course. Additionally, teachers shared the rationales for their design decisions, scoring guidance, and examples of student work. During the treatment course (Fall 2022), comparison group teachers continued with their business-as-usual instructional and assessment practices.

Table 1.
Demographic information for the teachers in the experiment

	Treatment (n = 22)	Comparison (n = 33)
Gender		
Female	82%	91%
Male	18%	6%
Race / Ethnicity		
White or Caucasian	91%	94%
Hispanic or Latino	5%	0%
Asian or Asian American	0%	6%
Grade Band		
Middle School	11	12
High School	15	28
User of storyline materials	45%	15%
Years Teaching	16.9	13.8
Hours of prior NGSS PL	11.9	9.3

Evaluating teachers' assessment practice

We developed a rubric to measure the extent to which the teacher-designed assessments aligned with the 5D vision. The rubric measured five aspects of 5D vision-aligned assessments: using phenomena for sensemaking, targeting integrated and grade-band appropriate use of the three dimensions, providing 3D feedback to students, enhancing the accessibility of the assessment, and designing the assessment to engage student interest. Each aspect was measured by a set of rubric elements. Some elements were scored dichotomously, and others were scored on a three-point scale (0, 1, 2).

Using phenomena for sensemaking. This aspect evaluated teachers' use of phenomena or problems to frame 3D sensemaking opportunities. There were 12 elements that made up this aspect with a maximum number of points possible of 23. Elements focused on how the phenomenon was presented to students including whether the teacher used more than one modality, included real-world accurate data and information, and whether the goal of the assessment was focused on explaining phenomena. There were also elements that evaluated whether the data and information presented in the task provided students with opportunities to use each NGSS dimension individually and together.

Targeting integrated and grade-appropriate use of the 3Ds. This aspect evaluated the alignment of assessed science content with the DCIs, SEPs, and CCCs. There were 5 elements for both the DCIs and SEPs and 6 elements for the CCCs. Across the three dimensions there were 32 possible points. The rubric elements looked at whether the opportunities to use the dimensions were at the element level or if it was more generally aligned to the dimension. Elements also focused on whether students' use of the 3Ds was done in grade band appropriate and integrated ways. There was an additional element that considered how explicit the assessment was in prompting students' use of the CCC(s). Use of each dimension was evaluated both on its own and together with the other dimensions.

Providing 3D feedback. This aspect evaluated the scoring guidance or tools used to provide feedback to students. The elements look at the dimensionality of the scoring guidance and whether the guidance supports opportunities to provide actionable feedback to help students improve their thinking and performance. There are two elements for this aspect with a total of three possible points.

Enhancing accessibility. This aspect evaluated whether the teacher intended to enhance the accessibility of the task and whether there is evidence that the assessment is accessible. Additional elements were included to evaluate whether students had the opportunity to use multiple modes to express their ideas. The four elements that make up this aspect were dichotomously scored for a total of four possible points.

Designing to engage student interest. This aspect included one element that indicated whether the teacher intended to engage their students' interest in selecting a phenomenon for the task. This element was dichotomously scored.

Scoring

The assessments were randomly assigned to one of three pairs of scorers and blinded by condition. Scoring partners independently scored each assessment and resolved disagreements through discussion. Approximately 13% of the assessments were scored by all three groups to evaluate inter-rater reliability and disagreements were resolved through discussion by all scorers. Teams revisited their scoring considering scoring discussions. We calculated scores for each aspect in the rubric by calculating the percentage of points earned out of the total possible points for that aspect.

Approach to analysis of teachers' assessment practice

Because we were interested in comparing treatment and comparison teachers' change over time (Fitzmaurice et al., 2012), we computed gain scores by subtracting the pre-score from the post-score and then compared the gain score of teachers from the two treatment groups. To compare the gain scores, we conducted a two-step hierarchical regression analysis. The first step included only the treatment group as the predictor variable. The second step included the treatment group and other variables we thought might affect the outcomes including the number of years teaching, years since their state adopted 3D standards, hours of prior PL on NGSS, and whether they used storylined curriculum materials. Teachers' use of storylined materials was based on whether the materials' designers claimed that the explanation of phenomena drove student learning. Our team did not analyze alignment of the materials with the 5D vision. Further, identification as a user of storylined materials simply meant that they had experience using these types of materials, but may not use this type exclusively in their instruction.

Analyses on the rubric elements were also conducted to further investigate the impact of the course on teachers' assessment practice. Because the element scoring is categorical, we used chi-square tests to compare the two groups. All the chi-square values presented here are from comparisons of teachers' post-assessment scores.

Baseline equivalence

We established baseline equivalence between the treatment groups by comparing the scores on the assessments the teachers submitted before the online course began. Table 2 shows the average pre-scores for the treatment and comparison groups. Independent samples t-tests indicated that none of the differences in the means were statistically significant ($p > .05$). Additionally, chi-square tests were conducted on the rubric element scores, and no statistically significant differences ($p > .05$) were found. Finally, we compared the means for the number of years teaching, years since their state adopted 3D standards, hours of prior PL on NGSS, and whether they used storylined curriculum. Independent samples t-tests showed no statistically significant differences ($p > .05$) for any of these variables. Based on these analyses, the measures of teacher's assessment practices before participating in the course can be considered equivalent.

Table 2.
Mean pre-assessment scores for the aspects of the 5D vision

Outcome	Group	Mean	Std. Deviation
Using phenomena for sensemaking	Treatment	.42	.35
	Comparison	.41	.30
Targeting integrated & grade-appropriate use of the 3Ds	Treatment	.44	.25
	Comparison	.44	.21
Providing 3D feedback	Treatment	.23	.35
	Comparison	.13	.20
Enhancing accessibility	Treatment	.40	.24
	Comparison	.33	.23
Designing to engage student interest	Treatment	.14	.35
	Comparison	.30	.47

Findings

The average post-assessment scores for each aspect of designing 5D assessments measured by the rubric are presented in Table 3. The treatment teachers scored higher in each category than the comparison group.

Table 3.
Mean post-assessment scores for the aspects of the 5D vision

Outcome	Group	Mean	Std. Deviation
Using phenomena for sensemaking	Treatment	.78	.16
	Comparison	.28	.29
Targeting integrated & grade-appropriate use of the 3Ds	Treatment	.61	.15
	Comparison	.41	.21
Providing 3D feedback	Treatment	.39	.35
	Comparison	.22	.28
Enhancing Accessibility	Treatment	.49	.26
	Comparison	.28	.23
Designing to engage student interest	Treatment	.73	.46
	Comparison	.33	.48

Using phenomena for sensemaking

A primary focus of the online course was supporting teachers in designing assessments that used authentic real-world phenomena to drive sensemaking with the three dimensions. The graph below shows how the treatment and comparison teachers shifted in their use of phenomena in their assessment tasks. The groups had the same average score on their pre-assessments, but the treatment teachers had higher scores on their post-assessments than the comparison teachers.

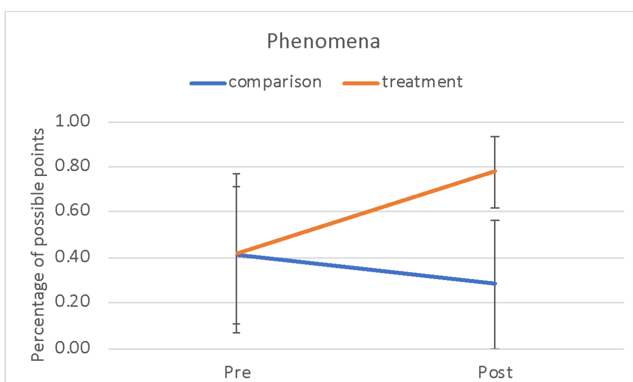


Figure 1. Average scores for using phenomena for sensemaking

The linear regression models of the gain scores showed a statistically significant treatment effect for teachers' use of phenomena (see Table 4). Treatment was the only predictor that was significant. To better understand the practical implications of the significant difference between groups we computed an effect size using Cohen's *d*. Our analysis found a value of 1.16, which is considered large (Cohen, 2013). This indicates that, overall, the treatment teachers shifted in their use of phenomena to drive sensemaking more than the comparison teachers.

Table 4.
Coefficients from the linear regression models for using phenomena for sensemaking

Model		Unstandardized Coefficients		Standardized Coefficients		
		B	Std. Error	Beta	t	Sig.
1	(Constant)	-.132	.071		-1.855	.069
	Treatment	.490	.113	.513	4.350	<.001
2	(Constant)	-.246	.180		-1.366	.178
	Treatment	.473	.120	.496	3.933	<.001
	Storylined curriculum	-.038	.145	-.032	-.262	.795
	Years teaching	.010	.006	.205	1.637	.108
	Years since 3D standards	.001	.027	.006	.050	.961
	Hours prior NGSS PL	-.002	.008	-.036	-.286	.776

Note: $R^2 = .26$ for Model 1, $\Delta R^2 = .04$ for Model 2 ($p > .05$).

Element level analysis. When we look on the element level, we find statistically significant post-assessment differences between the two groups for every element. Every post-assessment from the treatment teachers included a phenomenon compared to only 58% of the comparison teachers' post-assessments ($\chi^2 = 12.52$, $p < .001$). The treatment teachers also more often incorporated real and scientifically accurate data than the comparison teachers (64% and 9%, respectively; $\chi^2 = 20.54$, $p < .001$). The phenomena selected by the treatment teachers were more commonly specific events as opposed to general classes of phenomena ($\chi^2 = 16.59$, $p < .001$) and were described using multiple modalities ($\chi^2 = 17.94$, $p < .001$). Additionally, the phenomena in the treatment teachers' post-assessments were better problematized than the phenomena in the comparison teachers' post-assessments meaning that the treatment teachers better framed the phenomena in a way that invited students to wonder ($\chi^2 = 17.68$, $p < .001$). However, this is still an area of growth because only 23% of the treatment teachers' post-assessments were rated as being problematized.

Regarding the coherence of the assessments, 82% of the treatment teachers' post-assessments focused on sensemaking about phenomena (i.e., more than half of the work done during the assessment involved explaining phenomena), whereas only 9% of the comparison teachers' post-assessments had this focus ($\chi^2 = 30.93$, $p < .001$). While there was a shift towards incorporating multiple prompts about a phenomenon (87% of the treatment teachers' assessments), the prompts tended to not work together in a logical sequence to help students explain the phenomenon. Only 32% of the treatment teachers' assessments were rated as having a coherent series of prompts.

Targeting integrated and grade-appropriate use of the 3Ds

Figure 2 shows the difference between the comparison and treatment groups in how the teachers created opportunities for the students to use the three NGSS dimensions in their assessment tasks. This use could have been in service of explaining the phenomenon or it could have been separate from the phenomenon. We found that the treatment teachers increased from pre to post, while the comparison teachers had little change.

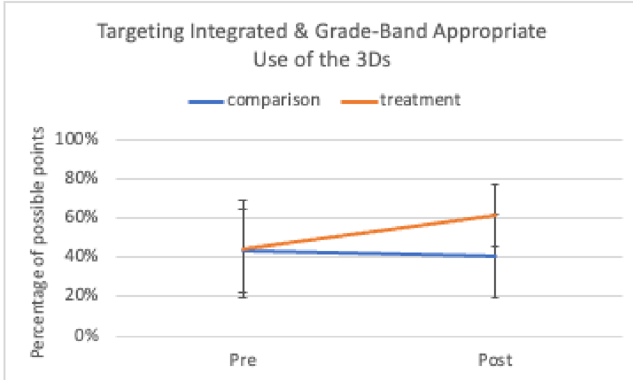


Figure 2. Average scores for targeting integrated and grad-band appropriate use of the 3Ds

Table 5 shows the results of the linear regression of the gain scores for the use of the 3Ds. There is a statistically significant treatment effect with the treatment teachers showing bigger gains than the comparison teachers. Treatment was the only predictor that was significant. The effect size for this difference is 0.87, surpassing the conventionally defined large effect threshold of .80 (Cohen, 2013).

Table 5.
Coefficients from the linear regression models for targeting integrated and grade-band appropriate use of the 3Ds

		Unstandardized Coefficients		Standardized Coefficients		
Model		B	Std. Error	Beta	t	Sig.
1	(Constant)	-.028	.043		-.648	.520
	Treatment	.200	.067	.379	2.978	.004
2	(Constant)	.048	.110		.439	.663
	Treatment	.211	.074	.398	2.867	.006
	Storylined curriculum	-.002	.089	-.004	-.027	.979
	Years teaching	-.002	.004	-.070	-.509	.613
	Years since 3D standards	-.008	.016	-.069	-.514	.609
	Hours prior NGSS PL	-.001	.005	-.018	-.135	.893

Note: R^2 = .14 for Model 1, ΔR^2 = .01 for Model 2 ($p > .05$).

Cross dimensional element analysis. To more deeply understand how the treatment impacted the teachers' use of the three dimensions, the rubric elements for each dimension were combined across dimensions. Gain scores for intent to align, grade band appropriateness use, element level use, coverage, and integrated use were calculated. Independent samples t-tests were conducted to compare the mean gain scores of the comparison and treatment groups. Table 6 shows the means and standard deviations for the gain scores. For all the elements, the mean gain scores for the treatment group are larger than the mean gain scores for the comparison group, which had little to no gains for all the elements.

Table 6.
Mean post-assessment scores for the elements related to targeting integrated and grade-band appropriate use of the 3Ds

Element	Group	Mean	Std. Deviation
Intent to align with NGSS	Treatment	.12	.29
	Comparison	-.06	.36
Grade band appropriateness	Treatment	.03	.37
	Comparison	-.00	.31
Use of targeted elements	Treatment	.18	.38
	Comparison	-.06	.31
Coverage of the targeted elements	Treatment	.11	.36
	Comparison	-.02	.22
Integrated use of the 3Ds	Treatment	.35	.42
	Comparison	-.01	.45

The results of the t-tests and effect sizes are presented in Table 7. The t-tests revealed statistically significant differences for all the rubric elements except grade band appropriateness. The effect sizes for intent to align with NGSS, element level use, and coverage are considered medium ($d \approx 0.5$), and the effect size for the integrated use element is considered large ($d \geq 0.8$) (Cohen, 2013).

Table 7.
Independent Samples t-Tests and Effect Sizes for

Rubric Element	t	df	p	Effect size
Intent to align with NGSS	1.987	53	<.05	.55
Grade band appropriateness	0.342	53	n.s.	.09
Use of targeted elements	2.480	53	<.01	.68
Coverage of targeted elements	1.779	53	<.05	.49
Integrated use of the 3Ds	3.006	53	<.01	.83

The cross dimensional analysis of the rubric elements (see Table 7) showed that the treatment teachers shifted towards aligning to the NGSS dimensions on the element level and towards targeting more of the content contained in those elements. Additionally, treatment teachers' assessments provided more opportunities for students to use the dimensions in integrated ways. The course did not have a significant impact on grade band appropriateness indicating that the use of the targeted dimension did not always match the progression articulated in the NGSS. This could be because many teachers still only indicated alignment to NGSS on the dimension level and were not explicit as to which element within the dimension they were targeting. This was most common for the CCCs where over half of the teachers did not articulate a target CCC element.

There was one additional rubric element for the crosscutting concept dimension focusing on whether the crosscutting concept was explicit in the assessment task or in the scoring guidance. A chi-square test revealed a statistically significant difference between the explicitness of the crosscutting concepts ($\chi^2 = 8.57$, $p < .05$). Forty-five percent of the treatment teachers included explicit reference to the crosscutting concepts in either the prompts or the rubrics, and 18% of the treatment teachers included explicit reference to the crosscutting concepts in both the prompts and the rubrics. On the other hand, the comparison teachers' use of the crosscutting concepts was predominantly implicit (76%).

An analysis of the rubric elements for each dimension showed some differences among the dimensions. There were fewer significant differences for the DCIs than there were for the other dimensions. The post-assessment differences for all elements for the SEPs except the grade-band appropriateness element were statistically significant, and the differences for half of the CCC elements were statistically significant. This suggests that the treatment teachers shifted more in their use of SEPs and CCCs than in their use of the DCIs. We hypothesize that this is due to teachers being more comfortable assessing the DCIs, which were the primary focus of pre-NGSS standards.

Providing 3D feedback

We evaluated whether the scoring guidance the teachers submitted assessed all three dimensions and whether it included specific qualities of thinking and performance for how students can improve related to their use of the 3Ds. Figure 3 illustrates the change in scores for the treatment and comparison groups. Both groups had small increases in scores. The linear regression models of the gain scores showed that there was no statistically significant difference between the gains of the treatment and comparison group for this aspect of 5D assessment design (see Table 8).

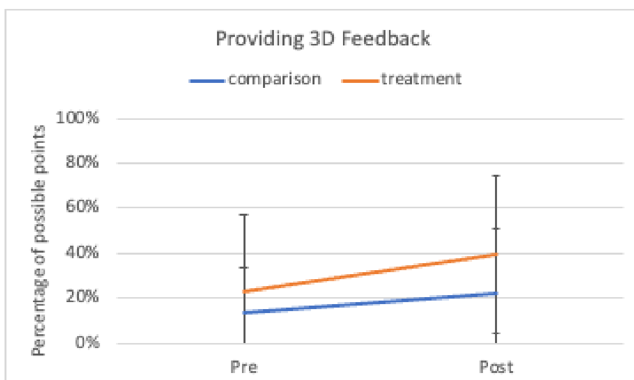


Figure 3. Average scores for providing 3D feedback to students

Table 8.
Coefficients from the linear regression models for providing 3D feedback

Model		Unstandardized Coefficients		Standardized Coefficients		
		B	Std. Error	Beta	t	Sig.
1	(Constant)	.091	.064		1.420	.161
	Treatment	.076	.101	.103	.751	.456
2	(Constant)	.232	.162		1.437	.157
	Treatment	.110	.108	.148	1.014	.315
	Storylined curriculum	-.077	.130	-.085	-.593	.556
	Years teaching	.004	.005	.108	.743	.461
	Years since 3D standards	-.019	.024	-.109	-.774	.442
	Hours prior NGSS PL	-.009	.007	-.195	-1.356	.181

Note: R^2 = .01 for Model 1, ΔR^2 = .05 for Model 2 ($p > .05$).

Element level analysis. There were two elements related to providing 3D feedback. Chi-square tests showed non-significant differences for both elements ($p > .05$). While we saw an increase in the percentage of teachers who designed multidimensional scoring guides (2 or 3D), very few scoring guides were three dimensional (23% treatment, 12% comparison). Additionally, few scoring guides were designed in a way that would allow for specific actionable feedback (27% treatment, 9% comparison).

Enhancing Accessibility

Figure 4 shows how the average scores on enhancing accessibility changed over time for each group. The treatment group showed a small increase, and the comparison group showed a small decrease. Like the opportunities to provide 3D feedback, the linear regression models for enhancing accessibility gain scores showed no statistically significant treatment effect (see Table 9). This indicates that, overall, experiencing the online course did not significantly shift teachers' ability to enhance the accessibility of their tasks.

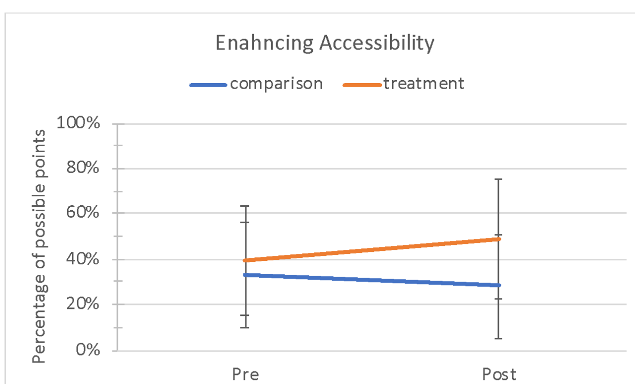


Figure 4. Average scores for enhancing accessibility

Table 9.
Coefficients from the linear regression models for enhancing accessibility

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.
1	(Constant)	-.053	.059		-.893	.376
	Treatment	.144	.094	.206	1.534	.131
2	(Constant)	-.267	.150		-1.780	.081
	Treatment	.102	.100	.146	1.016	.315
	Storylined curriculum	.125	.121	.146	1.038	.304
	Years teaching	.002	.005	.055	.383	.703
	Years since 3D standards	.032	.022	.198	1.426	.160
	Hours prior NGSS PL	.000	.006	.011	.075	.940

Note: $R^2 = .04$ for Model 1, $\Delta R^2 = .06$ for Model 2 ($p > .05$).

Element level analysis. Chi-square tests showed that one of the four rubric elements for enhancing accessibility had a statistically significant difference. That element indicated whether the teacher expressed an intent to enhance the accessibility of their assessment task ($\chi^2 = 7.64$, $p < .01$). Sixty-eight percent of the treatment teachers were intentional in attending to accessibility in their post assessment compared to 30% of the comparison teachers. A common shift in this element was moving from a generic description of ways to improve accessibility to pointing to specific features they used to make the assessment more accessible. One teacher wrote “Based it on discussions from lessons” on the pre-assessment survey and then wrote “I used photos and diagrams that would remind them of the lab they completed in class” on the post assessment survey. Another teacher shifted from “I tried to make the test so that each student had a chance to demonstrate their knowledge” to “I tried to use a variety of questions, so that I could assess all levels of students.” Despite the high percentage of treatment teachers intending to enhance accessibility, we found that half of the treatment teachers’ post-assessment tasks still showed evidence of barriers to accessibility. Although not statistically significant, we did see a small increase in the percentage of treatment teachers whose tasks provided multiple ways for students to express and represent their understanding while responding to the assessment (50% to 59%).

Designing to engage student interest

Teachers were asked to describe how they engaged their students' interest when selecting a phenomenon for their assessment task. Figure 5 shows an increase in the percentage of treatment teachers who attended to their students' interests when designing their assessment task, while the percentage of comparison teachers attending to student interest remained relatively constant.

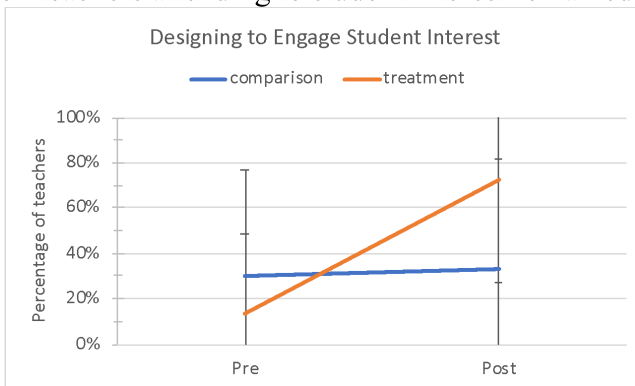


Figure 5. Percentage of teachers expressing an intent to engage student interest

Table 10.

Coefficients from the linear regression model for designing to engage student interest

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.
1	(Constant)	.030	.097		.314	.755
	Treatment	.561	.153	.451	3.674	<.001
2	(Constant)	-.062	.246		-.253	.801
	Treatment	.574	.165	.461	3.483	.001
	Storylined curriculum	-.141	.199	-.093	-.712	.480
	Years teaching	-.006	.008	-.091	-.694	.491
	Years since 3D standards	.018	.037	.063	.490	.626
	Hours prior NGSS PL	.010	.011	.121	.927	.358

Note: $R^2 = .04$ for Model 1, $\Delta R^2 = .06$ for Model 2 ($p > .05$).

The linear regression models of the gain scores confirmed that the difference in gain scores between the groups was statistically significant when controlling for years teaching, years since the state adopted 3D standards, hours of prior professional learning on NGSS, and curriculum used (see Table 10). The effect size was large with a value of 1.03 (Cohen, 2013). A chi-square test also showed a statistically significant difference in the post-assessment scores ($\chi^2 = 8.197$, $p < .01$). On the post-assessment survey, about 73% of the treatment teachers provided specific reasoning for how they tried to engage their specific students' interests in the assessment compared to 27% of the comparison teachers. One treatment teacher shifted from thinking about engaging interest as making the assessment more comprehensible and more closely related to instruction to considering their students' life experiences. For their pre-assessment, they wrote "I got rid of a question that would be confusing to the students. A question was added that covered material we have spent a lot of time on (pedigrees)." On the post-assessment, they wrote

“Hunting is a big part of Northern Wisconsin. Many people see albino deer and so I thought tying in something student's see, would make this relevant.”

Conclusions

Our analysis of the teacher designed assessment tasks showed that the course had a significant impact on teachers’ assessment practices compared to business-as-usual. Overall, treatment teachers had greater gains than the comparison teachers. We found that the course had the biggest impact on teachers’ use of phenomena to drive sensemaking in the assessment task. Treatment teachers shifted to providing students opportunities to use the 3Ds in integrated ways to make sense of scientifically accurate data and information about specific real-world phenomena. Additionally, we found that treatment teachers became more intentional about considering students’ interests when selecting phenomena and enhancing the accessibility of their tasks. However, we did find areas that remain a challenge for teachers including consistently providing opportunities for grade band appropriate use of the 3Ds, developing scoring guides that provide actionable 3D feedback to students, and removing barriers to accessibility in their assessments. This study contributes to the field by demonstrating the promise of the online professional learning course in supporting rural teachers to shift their assessment practices toward a 5D vision. It also highlights the challenges teachers face even after experiencing in-depth professional learning.

Paper 4: A Comparative Case Analysis of Rural Teachers' Experience with 5D Professional Learning

Loraine Glidewell*, Jennifer Jacobs, Annie Allen, Kerri Wingert
University of Colorado Boulder

Abstract

Rural science teachers represent a diverse group with variation in their local contexts, relevant prior experience, and how they engage in and learn from professional development. In this paper, we investigated the extent to which rural science teachers who participated in an online course focused on 5D assessment showed growth in their 5D-aligned vision, assessment and instruction We also examined how differences in the teachers' rural and professional contexts help to explain variation in the ways they showed growth. This mixed-methods paper pulls together data from surveys, interviews, exit tickets, course artifacts and field notes from the full group of teachers who engaged with the course (n=22) and from three selected cases for a comparative case analysis. On the whole, the teachers who took the course exhibited at least some degree of growth in all of the areas under investigation. The three case study teachers entered the course at very different places with respect to their standards-aligned beliefs, skills in creating a 5D assessment, and the extent to which their classroom instruction reflected the 5D vision. An important takeaway from the variation in these teachers' growth trajectories is that the online 5D assessment course was well-suited to support this degree of diversity and move each of them forward in measurable ways. Also evident is the teachers' rural cultural wealth, especially their resourcefulness and community unity, which helped drive their learning from the course.

Rural areas hold significant "community wealth" that could be leveraged in educational contexts, such as rural communities' resourcefulness, ingenuity, familism, and unity (Crumb et al., 2023). Rural science classrooms in particular have been shown to be a rich space for innovation. For example, rural lands and waters can become locations for rural students to engage in deep scientific investigations about sharing a watershed and how to take collective action (Zimmerman & Weible, 2017). Yet rural teachers often do not receive systematic support for this type of instructional innovation.

In the United States, "the unique needs of rural education are often obscured by their urban and suburban counterparts" (Lavalley, 2018, p.1). Supporting rural educational landscapes through avenues such as increased funding and greater accessibility of teacher professional development requires a greater awareness of the existing resources within rural communities. Recent research on critical pedagogies of place help to disrupt the educational inequities experienced in rural places by encouraging the utilization of local places for learning, and by highlighting the importance of attending to students' interests and identities rather than depending on generalized instructional resources that are often urban normative (Gruenewald, 2003; Huffling, Carlone & Benavides, 2017). Standardized tests and curricula are often framed from a metropolitan perspective, excluding the millions of students who live elsewhere. For example, Epply (2015) described a standardized test question that referenced the subway system in which a rural student responded with, "Oh, I know Subway! I eat there with my mom'' (p. 77).

In all locations, teachers are obligated to address common standards; a challenge faced by rural teachers is the limited opportunities to learn about those standards, and also to imagine how to organize teaching and learning opportunities that both address standards and attend to students' interests and identities, particularly those that are tied to place. Within the context of place, science learning research has been organized around a framework of three-dimensional learning, where students learn crosscutting concepts, disciplinary core ideas, and science and engineering practices as they investigate phenomena. Recent work has expanded this 3D model to be "five-dimensional" (5D) and includes as equally important the engagement of students' interests and identities (Bell et al., 2016; Lo et al., 2022). This can be especially valuable for rural students who are rarely acknowledged by urban normative curricula. Avery (2013) argues that "whether playing outdoors or working on the farm, rural children acquire science and engineering skills throughout their daily lives. Although 11.4 million children in the United States grow up in rural areas, compared to 14.6 million in urban areas, relatively little attention is given to rural science education" (p.28). We articulate 5D science learning as compatible with disrupting educational inequities in rural places (Gruenewald, 2003), and positioning the rural context as an asset (Crumb et al., 2023) for science teaching and learning because it locates students' interests and identities in specific places.

We developed an online PL course focused on 5D assessment that was designed to help rural teachers develop phenomenon-based tasks that elicit students' understanding of standards, connect to students' interests, and allow students to feel like scientists or engineers. The course brought together rural teachers from a number of US states and geographic contexts, highlighted the assets of rural students and educators, was conducted synchronously online to ensure access and encourage community building, and was curriculum-agnostic to support inclusivity and meet teachers where they were as professionals using various resources for unique student populations. We used a design-based implementation research approach to enact and study this virtual professional learning course.

In this paper, we investigate the following research questions: 1) In what ways did rural science teachers who participated in the online course show growth in their 5D-aligned vision, assessment and instruction? 2) How do differences in teachers' rural and professional contexts help explain the variation in the ways they showed growth? This mixed-methods paper pulls together data from surveys, interviews, exit tickets, course artifacts and field notes from the treatment group (n=22) and from three selected cases within that group for a comparative case analysis (Yin, 2003). Our hope is to highlight the shared needs and differences that can exist among rural teachers to support 5D vision-aligned science education and its implementation in rural schools (Zinger et al., 2020).

Background

Based on NCES definitions, 42% of the school districts in the U.S. – that serve 15% of students – are designated as rural (Gutierrez & Terrones, 2023). Teachers in these rural schools are less likely to receive professional development (PD) workshops and coaching—particularly in student-centered modes of instruction—than their urban and suburban counterparts (Banilower et al., 2018). However, surveys have found that rural teachers are eager to learn and grow, and are particularly interested in PL involving the science and engineering practices emphasized within *A Framework for K-12 Science Education* (National Research Council, 2012; Wingert & Penuel, 2019). Science-specific, phenomenon-driven, online professional learning holds potential for meeting the expressed needs of rural science teachers, in support of their efforts to implement high-quality 5D instruction and assessment (Wingert et al., 2022).

A Framework for K-12 Science Education makes clear that all students should have access to high-quality, 3D science instruction, yet rural educators receive far less support in making shifts in their teaching. In general, it is a struggle for rural teachers to access professional development and professional learning communities simply because these opportunities are often too far and too expensive (Durr et al., 2020; Lavalley, 2018). However this problem is even more pronounced for rural science teachers as it is increasingly difficult to recruit and retain those who are highly qualified in STEM (Harris & Hodges, 2018; Lavalley, 2018). The rural context often involves challenges for science teachers related to isolation, funding, small student populations and limited STEM course offerings (Sipple & Brent, 2008). For example, "rural teachers with backgrounds in chemistry, physics, or calculus may be unable to teach these courses because the student body is too small to support advanced courses" or because there is not enough resources or support for one teacher to be able to teach multiple STEM courses (Avery, 2013, p.30).

There are also considerably fewer resources allocated to *studying* what kinds of science teaching are occurring in rural schools and the degree to which rural teachers' views are aligned with the NGSS (Long & Avery, 2017; Thier et al, 2021). Due to grant funding preferences that often favor generalizability, replication power, and large impacts, non-rural schools are naturally favored and as a consequence, there are few studies on rural science education (Harris & Hodges, 2018, p. 6). Furthermore, the studies that do include rural education are often either "examined through the perspective and values of metropolitan academics and policymakers" or are using rural education to make a point of comparison to urban education (Johnson & Howley 2015 as cited by Lavalley, 2018, p. 23). As a result, research has mostly generated a shallow understanding of the assets and challenges of science teaching and learning in rural schools.

"Rural" describes a diverse range of communities (Hartman et al., 2022), but it is too often defined only in terms of population size and in contrast to cities (Azano et al., 2020). Studies have shown there are important differences among rural districts in the US in ways that directly impact teachers and students, especially smaller versus larger schools (Drescher et al., 2022). Rural communities are unique and complex, which makes generalizing specific research findings across geographic space "challenging and inappropriate" (Hammack et al., 2022, p.545). Some of the unique challenges associated with rural teaching include limited networking, mentoring, and professional development opportunities, as well as the expectation to teach multiple courses and ability levels at the same time (Lavalley, 2018). More specifically, rural science teachers represent a diverse group with variation in their local contexts and relevant prior experience (Zinger et al., 2020). This is also variation in how they engage in and learn from professional development focused on 5D instruction and assessment practices (Wingert et al., 2022). There is a need to understand how these differences in rural context, professional experience and particularly engagement with 5D practices influence how rural science teachers experience PL.

Conceptual Framework

In the spirit of disrupting place-based inequities (Gruenewald, 2003), the 5D course sought to provide professional learning opportunities to rural teachers who often, as a mere consequence of geography, do not have access to university-based resources. Many teachers in rural places are not only faced with geographic and professional isolation, but they are also faced with stereotypes reinforced by scholarly research that often emphasizes deficits of rural communities

(Hammack et al., 2022). Contrary to deficit-based perspectives, this study applies Crumb at al's (2023) asset-based conceptual framework of rural cultural wealth to both witness and leverage how the rural context influences science teaching and learning. In line with this framework, we view rural places as worthy of their own study, without needing urban comparisons. Rural is more than just "not-urban" (Crumb at al., 2023, p.2). We do not subscribe to the homogenous deficit-based narrative assumed and perpetuated by many non-rural folks, including some researchers and policymakers (Hammack et al., 2022).

The National Center for Education Statistics (NCES) defines rural in terms of distance to an urban center. Urban and suburban categories are dependent upon population size, broken into categories ranging from "city large" with more than 250,000 people to "suburban small" with less than 100,000 (NCES, 2022). Non-urban and non-suburban categories no longer rely on population, and instead categories are dependent upon the distance to an urban center ranging from "town fringe" which is less than or equal to 10 miles from an urban center, to "rural remote" which is defined as a "census-defined rural territory that is more than 25 miles from an urbanized area" (NCES, 2022). While this definition and categorizing structure informed our participant recruitment and relative positioning of participant's contexts, we believe the term rural describes distinct communities across the country that simultaneously may share similarities due to remoteness and yet remain highly variable and unique. Corbett (2016) states "if you have seen one rural community, you have seen...well, one rural community" (p. 278).

We set our gaze on an asset-based understanding of rurality described by Crumb et al. (2023) that explores the strength, resilience and immense variation of rural people and their communities. This is critical to help inform the analysis of our study which aims to investigate the ways in which this unique group of rural teachers had very different experiences of the course. Rural cultural wealth highlights four strengths of rural people that may be shared across communities, rural resourcefulness, rural ingenuity, rural familism and rural community unity (Crumb et al., 2023, p. 1). While aspects of all of these became visible while working with rural teachers, this study hones in on the tenants of rural resourcefulness and community unity. Rural resourcefulness starts with self-determination, and describes a "savvy" nature in which rural people can "overcome socio-contextual adversities" through "taking actions to mitigate limitations" (Crumb et al., 2023, p.5). For example, we can see this tenant come to life from the very beginning of this study when rural science teachers seeking to overcome the limited access to professional development signed up for the course. Community unity describes the ability to help one another, leverage the "composite assets" and "effectively organize and collaborate", especially in times of need (Crumb et al., 2023, p. 6). We can see evidence of this tenant being put into conversation with resourcefulness, in building community among rural teachers in the course we observed a culture of sharing resources and supporting one another when trying to learn the complex task of creating 5D assessments.

Methods

In this section we briefly describe the 5D assessment course developed by members of the research team and provide information about the teachers who participated in the course. Next we outline the measures and data sources that were used to study the impact of the course. Lastly we describe how the case study teachers were selected and how we generated their case profiles and cross-case analyses.

The 5D assessment course

Our research team developed an online course that engaged teachers in 25 hours of professional learning spread across 3 months. This duration allowed time for teachers to apply what they have learned to design a 5D assessment task that they administered to their students. Teachers worked in small collaborative groups based on grade band and content area. Teachers received feedback on their work from their peers as well as a "coach," a member of the research team with expertise in the content area and the 5D tools and processes. Using student work, teachers reflected on their assessment tasks' design and alignment with the goals of the 5D vision.

A central principle underlying our approach is the idea that engaging teachers in principled assessment design activities can be a powerful context for promoting teacher learning (NRC, 2001). Throughout the course, teachers developed an understanding of the components of the 5D vision, analyze the standards, and obtain information about their students' interests and experiences with science. Once teachers have analyzed the standards, they (1) identify candidate phenomena and design challenges that can address the targeted 3D standards in ways that connect to students' interests and science identity development and (2) develop prompts to scaffold students' sensemaking. Teachers used their analysis of the standards to develop scoring guides to draw inferences about students' mastery of the targeted standards and provide feedback to students.

Table 1

Context Information for the Treatment Group (n=22)

US Region (freq.)	*Remoteness (freq.)*	*Population (mean, range)*	*School Building (freq.)*	*State NGSS Adoption (mode, range)*	*NGSS PL Hours (freq.)*	*Storyline Curriculum (freq.)*	*Grade Levels (freq.)*	*Years Teaching (mean, range)*
Mid-west=18 North-west=1 South-west=3	*Town Distant=2 Town Remote=5 Rural Distant=8 Rural Remote=7*	*5,222 (260-25,679)*	*Multiple MS/HS=1 Separate MS/HS=8 Comb. MS/HS=7 Single K-12=6*	*2015 (2013 - 2020)*	*1-5=5 6-10=5 10-15=4 More than 15=8*	*No=12 Yes=10*	*MS=7 HS=11 MS/HS=4*	*17 (1-39)*

Participants

For the larger study, rural teachers were recruited and randomly assigned to the treatment or comparison groups. Each teacher taught science to U.S. middle and/or high school students in rural or "town" designated areas far from urban centers (i.e., US Census categories 33-43). Treatment teachers (n=22) participated in the 5D assessment course in Fall 2022 and control teachers (n=33) were offered the opportunity to take the course in Fall 2023. The three case study teachers that are the focus of this paper were selected from the 22 treatment teachers. Contextual information about the full treatment group is provided in Table 1. As a group, the teachers were predominantly from the midwestern U.S., working in remote areas with small populations, with almost half using storyline curriculum, and a relatively high level of prior professional learning and teaching experience.

Measures and Data Sources

Vision Survey. Teachers completed a vision survey before and after the course in order to examine differences in their baseline perspectives related to science teaching and learning along with any post-course changes. The survey focused on two aspects of teachers' professional vision, their vision for science assessment and their vision for how to support students' agency, interest, and identity in science. Here, professional vision refers to teachers' ideas about what should be happening in their ideal classroom (Hammerness, 2006). The source of constructs for these two aspects of teacher vision are a pair of National Academies' reports: *Developing Assessments for the Next Generation Science Standards* (National Research Council, 2014) and *A Framework for K-12 Science Education* (National Research Council, 2012). Both have served as policy guidance used by state leaders in designing professional learning for teachers in their states (Hopkins & Gates, 2019), and our own course emphasized these elements of professional vision.

Instructional Practices Survey. The survey of instructional practices focused on two aspects of teaching that were focal in the course: practices for connecting to students' interests and identities and engagement in the science practice of engaging in argument for evidence. We focused on practices that we conjectured would be relatively novel for teachers and that would represent significant shifts away from traditional practice if teachers' practice were to change. For example, it is not typical for teachers to gather information from students to choose phenomena, or to have students produce and critique knowledge through developing arguments based on evidence. The scales' constructs are based on ideas from *A Framework for K-12 Science Education* (National Research Council, 2012) and from Bang et al. (2017). The latter identifies several strategies for promoting equity in science instruction that were relevant to the scale related to supporting students' interests and identities, and the *Framework*'s descriptions of practices provided the basis for items related to argumentation.

Designed Assessments. Treatment teachers submitted three assessments as part of their participation in the study. They submitted both a pre-course and post-course assessment in order to investigate the extent to which the course impacted their assessment practices. For these assessments, teachers selected a curricular unit for which they could develop or modify an existing assessment to make it more aligned with the 5D vision, and that they planned to use with their students. Teachers submitted the pre-course assessment in Spring 2022 and the post-course assessment in Spring 2023. For both of these assessments, teachers shared the rationales for their decisions, scoring guidance, and examples of student work. The assessments were rated using a rubric that measured five aspects of the 5D vision: using phenomena for sensemaking, targeting integrated and grade-band appropriate use of the three dimensions, providing 3D feedback to students, enhancing the accessibility of the assessment, and designing the assessment to engage student interest. In addition, teachers developed an assessment as part of the course, during which they received ongoing feedback and guidance from the course instructors and their coach. These course designed assessments were not rated.

Recorded Classroom Observations. To document how teachers' observed classroom instruction was impacted by the course and to provide context for the assessments they developed, teachers self-recorded at least one science lesson before and after they took part in the 5D assessment course. Teachers were instructed to film lessons related to each of the two assessments that they submitted as part of the study (the pre-course and post-course assessments). The research team constructed a classroom observation protocol that aims to determine the degree to which teaching is aligned with 5D instructional practices. In particular, the protocol addressed whether specific aspects of 5D instruction were present in recorded science lessons, regardless of grade levels and scientific domains and topics. The protocol includes a set of four categories that generate information about the nature of students' classroom experiences: (1) phenomenon-driven instruction, (2) the NGSS dimensions (DCIs, SEPs & CCCs), (3) classroom culture, and (4) student interest, experience and identity. All of the codes were applied to 10-minute lesson segments, with the exception of the NGSS dimensions which are whole-lesson codes. For analyses, all coding was aggregated to the lesson level using a binary classification denoting whether a code was present or absent.

Course Field Notes. Detailed field notes were written by members of the research team during each course session. Notes were taken about how teachers participated in the whole group verbally and via the chat, questions teachers asked, teachers working in small groups with or without a coach, noteworthy moments of learning or sharing about practice, and challenges teachers had with course material or logistics. These notes were organized by session and linked to course slides and other documents for later analysis.

Semi-Structured Interviews. The teachers who participated in the course were asked to take part in semi-structured interviews over zoom after the completion of the course. The interview protocol included questions about what motivated them to sign up for the course, their 5D understandings, the challenges and benefits of engaging with course tools, the course organization and how it supported them, the coach interactions that help them, how they planned to use the 5D tool in the future, and any big aha moments they had throughout the course. The protocol also asked them to look at their pre-course "aligned" assessments that they submitted before the course, and describe what was different about that assessment and the assessment they created during the course with support. Twenty one teachers completed the interview, and their responses were recorded and transcribed. Three of the authors reviewed all of the interview transcripts, coded responses based on topic and theme, and wrote summaries of each teacher's reported experience of the course and their learning. Coding of interviews focused on what aspects of the course teachers said they found valuable (e.g., the focus on phenomenon, the tools used to unpack 3D standards), and what forms of support they received from coaches and instructors, and whether and how they might use the assessment design processes and tools in their classrooms and schools. Interview analysis also provided evidence about the teachers' rural teaching context, including whether they have science teacher colleagues they interacted with or not, and how their curriculum materials fit with or did not fit with 3D standards and the vision and practice of the 5D course.

All four coaches from the course were interviewed over zoom after the completion of the course. Questions asked coaches to reflect on each teacher they had worked with regarding their understandings of 3D and 5D aligned-instruction and assessment, the curriculum materials they used, and their trajectory of participation and learning over the course. These interviews were transcribed and an analytic memo was composed.

Case Selection and Analysis

Case selection. The selection of case study teachers for this paper was a multi-step process. Because our focus is on examining the variation in teachers' course experiences and outcomes, we started with a consideration of the full group of treatment teachers on the key course outcomes: self-reported 5D aligned beliefs and practices, 5D assessment design, and 5D classroom instruction. For each measure (teachers' survey responses and ratings of their assessments and instructional practice), we examined changes in the pre and post-course scores for the full group.

Next we looked closely at the variation in growth scores for each measure. One way we attended to variation was to rank order teachers by their amount of growth on each outcome of interest. We then organized the teachers into quartiles (e.g., the first quartile included the teachers with the most growth) to examine patterns across measures and teachers. Teachers in the first and second quartiles, across multiple outcomes measures, were considered as case study candidates. However, upon closer inspection it was apparent that some teachers exhibited higher or lower growth due to scoring very low or very high at baseline on some measures, which could skew an interpretation of their growth. Therefore we considered their placement in the growth quartiles as well as their pre and post outcomes, with improvements in overall assessment quality being a particularly important factor.

As a final step in case selection, we considered teachers' rural context and professional background. Our intention was to complement an exploration of teachers' growth with more nuanced attention to their local environment and educational background, particularly with respect to their local environment, prior knowledge of the NGSS and their use of standards-aligned instructional materials. We ultimately selected teachers from three different regions of the US, with two working in areas designated as "rural remote" and one in a "distant town." The three teachers had very different levels of familiarity with the NGSS and only one of the teachers used storyline curriculum.

Table 2
Case Study Teachers Regional Teaching Context

Case	*US Region*	*Remote-ness*	*Pop.*	*School Building*	*State NGSS Adoption*	*NGSS PL Hours*	*Storyline Curric.*	*Grade Levels*	*Years Teaching*
Dot	*Midwest*	*Rural Remote*	*1290*	*Single K-12*	*2013*	*1-5*	*No*	*HS*	*38*
Mary	*North-west*	*Rural Remote*	*1802*	*Combined MS/HS*	*2016*	*15+*	*No*	*HS*	*15*
Nora	*South-west*	*Town Distant*	*19,419*	*Separate MS/HS*	*2018*	*10-15*	*Yes*	*MS*	*8*

Developing case profiles. For each of the three identified teachers, we developed a detailed case profile by compiling information on their place context, professional background, and course experience. Information about their course experience was based on pre-course data (e.g., pre-course survey, assessment, and instruction ratings), data on their participation in the course (e.g. course field notes, coach interviews), and post-course data (e.g. post-course rating, teacher interviews). We created extensive analytic memos for each teacher that included all of the relevant data, selected quotations, and comparisons to the full group of treatment teachers and the other case study teachers. After several reviews of the analytic memos, we generated case study profiles with the goal of maintaining essential information to illustrate the local context, professional background and course experience of each teacher and how they compared to the others.

Cross case analysis. Once the case profiles were generated, we took both an inductive and deductive approach to looking across the cases. We looked for emergent themes by considering similarities and differences in the teachers' backgrounds, professional contexts, and growth data. In particular, we sought to uncover and explain the variation in their course experiences and learning outcomes. In particular we attempted to delineate each teacher's learning trajectory from pre to post-course, including where they started and ended relative to one another on multiple dimensions. Given that the teachers were selected, in part, due to the fact that they each made gains, we also sought to investigate why those gains might have occurred and the extent to which the likely explanations were similar or different for each teacher.

We used a deductive approach to the cross cases analysis by considering how the teachers' experiences mapped onto the literature on rural cultural wealth, both individually as a group. Working largely from Crumb and colleagues' (2023) asset-based framework and attending closely to the teachers' resourcefulness and community-based supports, we considered how each teacher's growth could be understood and explained within these lenses. Additionally, we looked for similarities and differences among the teachers in terms of their rural cultural wealth, in an effort to understand how variation in this domain pertains to teachers' progression through and growth outcomes from the course.

Findings

Changes in Teachers' Vision, Assessment, and Instruction

To address our first research question about how participation in the online course impacted teachers' 5D-aligned vision, assessment and instruction, we considered the full group's pre and post-course scores in each of these areas. As shown in Tables 3 and 4, on average the treatment teachers exhibited at least some degree of growth in each of these areas. The tables also show how the three case study teachers compared to the full group and to each other.

Table 3

Teachers' Survey and Assessment Scores from Pre- to Post-Course, Treatment Group and Case Study Teachers

		Survey Scores				Assessment Scores			
		Assess. Vision	***Interest & Identity Vision***	***Interest & Identity Practice***	***Argumentation Practice***	***Overall Assess. Quality***	***Using Phenom for Sense-making***	***Grade appropriate use of the 3Ds***	***Engage student interest***
Treat-ment Group	*Pre Mean (range)*	*44.9 (27-56)*	*12.25 (9-16)*	*100.2 (16-560)*	*140.8 (48-420)*	*0.44 (0.13-0.81)*	*.42,*	*.44,*	*.14*
	Post Mean (range)	*47.2 (31-56)*	*13.95 (10-16)*	*270.7 (52-640)*	*230 (60-540)*	*0.69 (0.46-0.89)*	*.78*	*.61*	*.73*
Dot	*Pre*	*35*	*9*	*264*	*60*	*0.35*	*.35*	*.34*	*0*
	Post	*31*	*10*	*228*	*100*	*0.78*	*.86*	*.75*	*0*
Mary	*Pre*	*46*	*11*	*108*	*180*	*0.17*	*0*	*.19*	*0*
	Post	*48*	*13*	*188*	*540*	*0.51*	*.74*	*.44*	*0*
Nora	*Pre*	*41*	*12*	*176*	*140*	*0.65*	*.61*	*.63*	*0*
	Post	*44*	*12*	*148*	*180*	*0.87*	*.96*	*.81*	*1*

Table 4
Teachers' Videorecorded Classroom Instruction Scores from Pre- to Post-Course, Treatment Group and Case Study Teachers

	Phenom *(At least 1 phenomenon)*		***Work*** *(At least some sensemaking)*		***Collab*** *(At least some collaboration)*		***Related Examples*** *(At least some examples)*		***Affirm Ideas*** *(At least some student ideas affirmed)*		***SEPs Rich*** *(At least 1 rich SEP)*		***CCCs Rich*** *(At least 1 rich CCC))*	
	Pre	*Post*	*Pre*	*Post*	*Pre*	*Post*	*Pre*	*Post*	*Pre*	*Post*	*Pre*	*Post*	*Pre*	*Post*
Treat % lessons	*24%*	*33%*	*48%*	*71%*	*48%*	*76%*	*52%*	*57%*	*62%*	*76%*	*33%*	*33%*	*14%*	*43%*
Dot	*no*	*no*	*yes*	*yes*	*no*	*yes*	*yes*	*yes*	*yes*	*yes*	*no*	*yes*	*no*	*yes*
Mary	*no*	*yes*	*no*	*yes*	*yes*	*no*	*yes*	*yes*	*no*	*yes*	*yes*	*no*	*no*	*no*
Nora	*yes*	*yes*	*no*	*yes*	*no*	*yes*	*no*	*yes*	*no*	*yes*	*no*	*no*	*no*	*yes*

Case Study Teacher Profiles

Dot: Learning "The What" of 5D Assessment

Place and Professional Context. Dot teaches in a remote, rural community in the midwestern US, with a population of less than 1,300 residents. This ranching and farming community is over 300 miles from the state capital and about 35 minutes from the nearest Starbucks. Located in the Great Plains, the land is predominantly flat with gentle rolling hills and a stream that runs through the center of town. There is a lake that residents use for kayaking, along with numerous trails for outdoor recreation and birdwatching. The residents of the town are predominantly White (89%) and Latino (8%), with a median household income ($62,000) well below the national median, and a poverty rate of 11%.

Dot works in a school that includes middle and high schoolers from her community, with almost half (46%) qualifying for free or reduced lunch. She teaches science to the high schoolers, spanning grades 10-12. Her course responsibilities include Biology, Chemistry and Physics, both regular and advanced (honors) classes. She uses a curriculum that she categorizes as primarily "self-designed." Although her state was one of the early adopters of the NGSS, Dot entered the course with very limited prior knowledge of the standards and minimal professional development opportunities focused on them. Dot says she still loves teaching even after close to 40 years, particularly "when students make that 'oohhhh' sound because they finally understand something."

Pre-Course Data. Dot's limited familiarity of the NGSS is reflected in her responses to the pre-course survey, suggesting that her beliefs and approach to science instruction in general are not closely aligned with the standards. She scored well below the group average on both of the vision subscales (assessment, interest and identity vision) and on the argumentation practice subscale, indicating that her beliefs on these topics, along with how often she engages students in argumentation, are all misaligned with the NGSS. However, it is notable that her initial scores on the interest and identity practice subscale are actually considerably above the average, and much higher than the other two case study teachers. There appears to be somewhat of a disconnect between how Dot responded to the beliefs and instructional practices questions that were about connecting to students' interests and identities. Although Dot's responses to the four items in beliefs subscale suggest that she placed a lower value on connecting teaching to students' interests and identities, on the six items in the practices subscale she reported frequently engaging in classroom practices that center students' interests and identities.

Dot's observed instructional practice helps to shed some light on the opposing views found in her survey responses. Her pre-course classroom recording showed clear attention to both a positive, student-centered classroom culture and eliciting students' interests and identities. The lesson was focused on light waves, including reflection, refraction, prisms, rainbows, plane mirrors and the like. The lesson moved back and forth between periods of teacher demonstrations (showing different kinds of light rays), student-led activities (using mirrors to reflect light), and teacher lectures (about the index of refraction). Dot did most of the talking and only occasionally engaged students in sense-making, however she brought up real world examples multiple times throughout the lesson and endeavored to relate the content to students' own experiences, such as by showing an x-ray of a student's hand. This lesson recording, combined with Dot's self-reported attention to interest and identity, suggest that she was, in fact, attuned to promoting student engagement and encouraging their genuine interest in complex scientific topics. At the same time, Dot did not connect to students' interest and identity in important ways that are emphasized in the NGSS, such as incorporating scientific phenomena or using science to promote social justice, perhaps leading to the disconnect in the way she answered the vision and practice questions.

Dot's pre-course science assessment earned a score below the group average, but was in-between the scores of the other two case study teachers. The assessment covered different kinds of waves (e.g., light, microwave) and consisted primarily of multiple choice items with a few short response questions. Dot developed both the curriculum and assessment materials on this topic herself. A few of the test questions involved phenomena, some had real world applications (e.g. describe a medical application of ultraviolet radiation, explain how a microwave uses microwaves to cook food), and some asked students to engage in reasoning (what happens when white light shines on green paper). The questions were primarily focused on DCIs, although there were a few opportunities for students to engage with CCCs. Although she was not certain that her students would find the assessment engaging or relevant, Dot did feel certain that they would do well on it because they had mastered the material. She explained, "I taught this unit to senior physics students and then they taught this unit to freshmen physical science students. The seniors knew the material well because they had taught it."

Course Experience. Because of her limited background with the NGSS, Dot entered the course feeling behind many of the other participants and expressed concerns about keeping up with the conversations. She recalled during an interview after the course ended, "At the beginning I was really overwhelmed, because everyone seemed to know a lot more about it than I did. And they were throwing around acronyms that… I didn't even know what they stood for. I had to request that, at least the first time you use one, that you explain what it is, or at least use the full name of it." Dot understood that the standards set forth certain context expectations, but she was not familiar with the three dimensions or how they worked in concert. She explained, "I started out only looking at the disciplinary core ideas and trying to figure out, okay, what am I expected to teach my students? And so… I pretty much have ignored the science and engineering practices and the crosscutting concepts. Haven't really paid any attention to them at all before now." However as the course progressed, Dot felt that she learned a great deal about the 3Ds and reflected, "I got a lot more familiar with how the standards are set up and how to use them. And I think that was the main goal. You know, that was the reason I took it took the class."

Similarly, Dot struggled to follow the course's heavy emphasis on scientific phenomenon and developing assessment tasks anchored in phenomenon. At the same time, Dot was keen to better understand the expectations around incorporating phenomena into her assessments, and to internalize the distinction between engaging hands-on activities, topics that might capture students' attention or interest, and scientific phenomena that students could be asked to explain. Fortunately she was comfortable voicing her questions and reaching out for guidance, which often led to extended conversations. For example during one course session she asked in chat: "What do you mean by phenomenon? I'm having trouble making sense of that word." The course instructors, together with Dot's peers and assigned coach, were all eager to engage with her questions and offer support. Looking back after the course she recalled, "I didn't even know what a phenomenon was….I just flat out had to ask. Because everybody else seemed to know what they were talking about and I had no idea… But I have a feeling that in the future there's probably going to be more and more materials out there that do this. Because I think it's something [that's] really been emphasized in the…new science standards and it's just a different way of looking at things. And, you know, I just had to kind of get used to it."

Dot's persistence throughout the course, in combination with the ongoing encouragement and assistance that she received, culminated in a designed assessment centered on skin cancer, including how cancerous cells form and reproduce. In her planning drafts leading up to the final version of the assessment, Dot worked hard to ensure that the tasks moved beyond the topic (cancer) to specific instances of cancer (phenomenon), were organized by a guiding question with scaffolds, stayed within the grade band assessment boundary for her target DCI, used pictures and representations that seeded information and encouraged student noticings, and was organized in a coherent fashion to help students work through their explanations. Dot was upfront with how challenging she found the process of designing 5D assessment tasks, sharing within her small group: "I feel like I'm further back than everyone else is. I've used multiple choice tests for so many years and am just starting to use phenomena with scenarios. I'm a whole lot better at writing multiple choice questions." Dot's colleagues and her coach provided frequent encouragement, shared relevant resources, and brainstormed ways to focus and improve the assessment tasks.

Despite all of these efforts, once Dot administered the assessment to her students she identified numerous problems with it. One problem she noticed was that it contained too much text and caused her students to become overly frustrated, particularly those with language barriers or who had difficulty reading. The assessment also required the students to engage in more writing than they were used to doing. When she observed students providing very poor written responses as they were taking the test on their computers, Dot remembered what one of the course instructors told her about encouraging students to draw what they know. Dot decided to hand out blank pieces of paper and invited students to explain their ideas using drawings, which resulted in much better responses. Ultimately Dot decided that she would not give this particular assessment again, but rather that she would start over using what she learned from the course. Towards the end of the course Dot told her colleagues, "I don't like the assessment. I'm going to start fresh with a new assessment for this spring and try to make it better."

Post Course Data. The assessment Dot developed after she completed the course required her students to watch a video showing a set of colliding metal spheres in perpetual motion, also called Newton's Cradle. The video is intended to be paused at designated times, so that students can be prompted to answer a set of open-ended questions about energy and energy transfer. Dot found the assessment from a website for teachers, and adapted it for her students. All of the questions require students to apply scientific reasoning to the phenomenon, and there are opportunities for students to engage with DCIs, SEPs and CCCs in integrated and grade-level appropriate ways. Dot. was hopeful that her students would find the topic interesting, especially because it incorporated "a hands-on activity" and the opportunity to "think about energy transformations in their everyday life." Although Dot expressed numerous times that she felt more comfortable using traditional multiple choice assessments, as this post assessment makes clear she was willing to move outside her comfort zone and devote the time to generating new types of activities that support her students to showcase their knowledge. This assessment was scored well above the group average and represents a dramatic improvement from Dot's pre assessment, although it fell short in the areas of accessibility and engaging students' interests. Although all three case study teachers made notable gains across their pre-post assessments, Dot's gain was the largest.

Dot did not demonstrate particularly notable changes on the post-course vision and practices survey, with the exception of the argumentation practices subscale where she increased in her alignment to the NGSS. Dot's post-scores on both the assessment vision and interest and identity practices subscales actually decreased somewhat. Overall the course appears to have had only a minimal impact on her self-reported beliefs and instructional practices, as captured by this survey. In addition, Dot's post-scores were below the group average on all four subdimensions and were the lowest of the case study teachers on three subdimensions.

On the other hand, Dot's post-course classroom recording suggests the class did have some degree of impact on her everyday science teaching. In this lesson, Dot used a variety of wind-up toys to help her students learn about energy. The students were asked to draw a series of energy pie charts to show how energy is stored and then exerted before, during, and after the toys are wound. They also considered energy transformations through various activities - some of which they undertook in small groups - such as rubbing their hands together, observing a bowling ball pendulum, and dropping bouncy balls. Although the lesson did not incorporate scientific phenomena, it did engage students in the work of sensemaking throughout and Dot provided a plethora of real world, related examples. Furthermore, students engaged with both SEPs and CCCs in rich ways – for example as they developed models, posed questions based on their models, and used their models to make sense of how energy flows within a system.

Summary. Overall, Dot took a good deal away from the online 5D assessment course, as evidenced by the large improvements in her post-course assessment and classroom observation. Despite beginning the course with relatively little knowledge about the NGSS and finding it challenging to keep up, due to her own determination and the ongoing support that the course offered, Dot felt like she made substantial progress and increased her understanding of what standards-aligned assessment and instruction looks like. Although she reported that traditional, multiple choice assessments are still likely to play some role in her science classrooms, she is particularly motivated to seek out more phenomenon-based resources. She shared, "I'm watching for them and I'm thinking about them. And before, like I said, I didn't even know what [a phenomenon] was."

Mary: Learning "The How" of 5D Assessment

Place and Professional Context. Mary teaches in a remote rural town in the northwestern US, that is quite similar to Dot's community in many respects. Both have similarly small populations, and are located far from their state's urban areas and the typical conveniences of larger communities. Mary's town has approximately 1800 residents, is over 450 miles from the state capital, and about 80 miles from the nearest McDonald's. The town is just one square mile and is surrounded by vast open space that contains numerous oil fields and natural gas deposits. A nearby state park contains sandstone formations along with dinosaur fossils. Many of the residents are ranchers, with mining, oil and natural gas drilling also contributing to the local economy. The average household income ($76,000) roughly matches that of the US as a whole, but the poverty rate (19%) is relatively high. The vast majority of the residents are White (92%), along with a sizable number of Native Americans (5%).

Like Dot, Mary works in a combined middle school and high school building, and she teaches high school science to students in grades 10-12. Her teaching responsibilities also span across subjects and include both regular and advanced (AP) courses for which she uses "self-designed" curricular resources. Mary has been teaching for 15 years and recalls, "Two weeks into teaching I was immediately smitten and have never looked back!" Her state adopted NGSS a few years after it was released, and Mary quickly became a leader in supporting their implementation across her state. She took part in professional learning opportunities related to the standards even before they were adopted, was on her state's standards writing team, and then led professional development workshops based on the standards. She received the teacher of the year award in her state along with multiple other awards for excellence in teaching.

Pre-Course Data. Given her extensive experience with the NGSS, it is not surprising that Mary's pre-course survey scores indicate strong alignment with the standards, particularly with respect to her beliefs about how best to assess students and the frequency with which she engages students in argumentation. Mary's scores in both of these subscales (assessment, argumentation) were above the group mean and were the highest of the three case study teachers. In terms of connecting to students' interests and identities, Mary's responses roughly matched the course participants' average, suggesting that she both values these connections and strives to make them explicit in her science lessons at least to some degree.

Mary's pre-course recorded science lesson can best be characterized as a lecture on the structure of various organic carbon compounds, with the teacher frequently posing short answer questions to her students to ensure they are engaged and following along. At one point the students were provided with materials to collaboratively build a physical model of butane, again with Mary asking questions, showing the students her own model, and explaining how models show the chemical bonds between the molecules. Although there was no phenomenon in this lesson and Mary did the sense-making work, there were several brief occasions in which she provided real-world examples intended to capture student interest. For example she explained that some compounds, such as optical isomers, are so similar to one another they can be easily mistaken, which can be deadly if they are accidentally used in food or medicines.

Interestingly, Mary's initial assessment scored well below the group average and received the lowest score of all three case study teachers. The assessment was a traditional test of organic compounds and molecular structures that mapped closely onto Mary's recorded lesson. It was not grounded in any phenomenon and consisted almost entirely of true/false, multiple choice, and short answer questions. The assessment primarily solicited rote knowledge, was not aligned to the conceptual nature of the target DCI, and did not provide students opportunities to demonstrate their understanding of the targeted SEPs or CCCs. Mary described the assessment as one she created herself by adapting materials from her textbook and other resources, and that her goal was for students to recall the information they went over in class.

Course Experience. Right from the start, Mary deeply engaged with and thoroughly enjoyed the online 5D assessment course. She especially liked unpacking the standards and looking at the progressions across grade bands, which she had not done before. She shared that these activities helped her to see the NGSS differently despite her many years of experience working with them. She quickly became a leader in her small group sessions, sharing her knowledge with and inspiring teachers such as Dot when they raised questions and expressed concerns. Mary frequently offered her peers tips on where to find specific resources, provided materials that she had on hand, and generally interacted in a way that her coach appreciatively labeled "as a co-facilitator."

Mary eagerly embraced the challenge of designing assessments around phenomena, brainstorming candidate phenomena for her own classes and helping others in the group to generate ideas. After just a few sessions Mary exclaimed, "I'm going to make a goal to pick an anchoring phenomenon for each unit. I'm kind of obsessed with it." She very intentionally considered her students' interests when deciding what phenomenon to use for her course designed survey, ultimately deciding on an ultramarathon runner whose muscles don't get sore regardless of how much they are exerted. Mary shared that many of her students, especially the boys, are really into fitness and working out. She heard them talk about a "muscle man" who quickly metabolizes lactic acid and has been the subject of research studies. Mary told the group that she plans to show a video about this muscle man runner in the hopes that it will encourage her students to notice and wonder about cellular respiration. She predicted that her students will notice that he can run farther than most humans and they will wonder if his muscles ever get sore or cramped. Mary also shared that she would like to use new forms of assessments, especially to support her students who do not score well on traditional tests or writing assignments, ensuring that they have an "on ramp." She explained, "I'm rethinking what an assessment is… I think assessments could be in pairs or groups. I'm trying to be open minded because not all kids are great at taking a test and it doesn't mean they don't understand."

Mary developed her course-designed assessment for her AP Biology students and was careful to ensure that they would have to "figure things out." She thought about what information to include and exclude from the video of the muscle man, how to bring in data and graphing, and how to make sure that "everything points back to answering the big question." All of these considerations were well-aligned with the goals of the online course, including the notion that 5D assessments can be appropriate for students at all achievement levels. Mary was receptive to feedback from her peers as well as her coach, but she sometimes preferred working on her own and the development of her assessment did not garner as much discussion as Dot's. Her coach expressed that he only needed to provide "small nudges" to Mary because she was already well positioned to use multiple dimensions within an assessment storyline that involved an interesting phenomenon.

Mary appreciated that her coach was able to take the students' perspective when providing feedback on her assessment, which she felt helped to ensure its success. In fact, Mary was quite pleased with how the administration of her muscle man assessment went. Her students liked the video and connected to the topic, which solidified for Mary that centering phenomenon is an key component of effective science assessment for all students. When Mary asked her students for explicit feedback about the assessment, she learned that they wanted more data and needed more support related to muscle fermentation – which Mary noted is an important topic in AP biology and might be something that she spends more class time on in the future.

Post Course Data. Similar to Dot's course-designed assessment, Mary's post-course assessment also dealt with cancer, but in Mary's case the topic was breast cancer rather than skin cancer. Following her success with the muscle man video, Mary's post-assessment also included a video about a woman with a non-detectable form of breast cancer. The assessment began with a question asking students to write down their noticings and wonderings. After several in-depth questions about cancer cells and breast tissue, the assessment shifted to a lengthy reading about several different types of breast cancers. The ensuing questions required students to apply their knowledge from the reading to the breast cancer video they had watched. Lastly there were more questions about cancer cells and the assessment culminated with an essay. Mary explained that students worked on portions of the assessment as a whole class, in small groups, and individually, and she assessed both their oral and written contributions.

This post-assessment was scored considerably higher than Mary's pre-course assessment. Interestingly, not only did Mary write the post-assessment herself, she used a range of course provided resources to support her writing process. The assessment included a specific, real world phenomenon that students were asked to puzzle about and explain. Although there were numerous pictures and diagrams that supported students to piece together what might be happening, the assessment was determined to be not entirely coherent, with a DCI that was not always connected with the targeted SEPs, and without explicit CCCs. Furthermore, almost no outside knowledge was required to complete assessment; all the information needed was included in the assessment itself. These issues led Mary's post assessment to be rated below the group average and as the lowest among the case study teachers.

Mary's responses to the post-course survey indicated that she continued to have beliefs that were highly aligned with the NGSS. On the argumentation subscale, she scored the highest of all the course participants, suggesting that the course strengthened her commitment to engaging students in this important practice. Her score on the assessment vision subscale was similar to her pre-course score, and remained the highest of the case study teachers. In terms of her beliefs and practices supporting students' interest and identity, Mary made gains similar to the full group. In particular her self-reported instructional practices suggest increased attention in this area of her teaching.

In her post-course recorded lesson, Mary had her students to consider the process of tasting at a cellular level. She first had them taste a number of different foods, one of which was tea and others that contained different kinds of sugar. Students made observations, noting for example that the tea tasted like dirt. The students then discussed what they thought was happening with their taste receptors, such as how different chemicals might be sending signals to the cells on their tongue. During this portion of the lesson, the students worked to explain and make sense of how tasting occurs, with students doing a good deal of the talking and sharing their ideas. The second half of the lesson was primarily a teacher-directed lecture on mitosis with the students taking notes. Throughout the lesson, both the teacher and students made connections to related, real world topics such as the impact of certain pills on diabetes and other health concerns.

Summary. Mary made identifiable changes to both her assessment and instructional practices after taking part in the course. Although her NGSS background and experience level was one of the most extensive of any course participant, she did not always use this knowledge to inform her approach to assessment and teaching, particularly for her advanced classes. At the same time, likely because Mary had strongly aligned beliefs and prior knowledge, she was very open to innovating and trying out the new approaches as she learned about them from the course. Of note is the fact that she intentionally selected her AP Biology course as the focal period to create assessments for and to experiment with pedagogically. Mary expressed that her biggest take away from the course was a new found passion for using scientific phenomenon to drive her teaching. As she explained, "I didn't really understand the importance of the phenomenon. I'd heard of other teachers doing it, and I was like that just seems, I don't know, shallow.… I teach the advanced classes, and so I thought that a lot of [my students] would be like, can we just get to the get to the meat of it and skip all the song and dance." However after taking the course Mary became fully bought into the idea of using phenomenon to "braid together a lot of content." In fact, she reports being constantly on the lookout for potential phenomena to use with her students, including in her "daily life on the ranch," and actively attending to how to incorporate them in her assessments and instructional practice.

Nora: Learning "The Who" of 5D Assessment

Place and Professional Context. Nora teaches in a rural community in the southwestern US that is different from the other two case study teachers in several important respects. Nora's community is less remote, with a census designation of "town-distant," meaning that it is between 10-35 miles away from an urbanized area. In fact, this community is just under 35 miles from the state's capital, and boasts a relatively wide assortment of coffee shops, restaurants, and local hangouts. The community is surrounded by mountains, mesas and canyons, with a wealth of trails and other outdoor recreation opportunities. The town is near a government-run facility that primarily employees scientists and engineers. These relatively high paying jobs help to explain why the community has an average household income ($118,000) well above the national average and a low poverty rate (5%). Nora's community is more diverse compared to the other two case study teachers, with residents who are 71% White, 18% Latino, 5% Asian, 4% two or more races, and 1% each Black and Native American.

Nora teaches science to 7th and 8th graders in a standalone middle school building, predominantly courses on life science and astronomy. Nora started her career working as an environmental scientist for several different companies, but discovered that she much preferred working with children and has been a science teacher for eight years. Her enthusiasm for teaching shines though in comments like the following: "I feel so humbled and inspired to shape and inspire young minds every day…. We are surrounded by the natural beauty of mountains, volcanic craters, ponderosa pine forests, canyons and mesas. It's a science teacher's dream to bring place-based learning to life for my students."

Although her state initially faced strong opposition to the NGSS before ultimately adopting them, Nora's school district fully embraced the standards early on and her school highly encouraged the use of NGSS-aligned resources. The district has provided numerous opportunities for their teachers to become versed in the NGSS and Nora took advantage of those opportunities for about 10-15 PL hours prior to joining the study. Nora reported used the OpenSciEd curriculum, which is closely aligned with the vision of 5D assessment and instruction. Nora was the only case study teacher to use storylined curriculum, which she adapted to meet the needs of her students.

Pre-Course Data. Based on the fact that Nora had prior PL experience related to the NGSS and used NGSS-aligned curriculum materials, it is not surprising that her scores on the pre-course survey were, for the most part, very close to the group average. On every subscale her scores fell in between the other two case study teachers. On her self-reported frequency of instructional practices that connect to students' interests and identities, Nora scores were slightly higher than the group average.

Nora's pre-study lesson was an online lecture that she recorded for her students, presumably for them to watch asynchronously. In the lesson, she used slides and walked the students through a written investigation that they were expected to complete, explaining what the various questions on their worksheet were asking and providing scaffolds and examples as to how they might respond. The lesson centered on how antibiotics can be used to eliminate bacterial infections, and why it often takes multiple doses of the antibiotics to do so. Nora was one of only five teachers to incorporate a scientific phenomenon in her classroom observation prior to the study, and was the only case study teacher to do so. Because the lesson was simply a recording of Nora speaking, there was no class discussion and no indication of how her students might have responded or what sense making they might have engaged in when they completed the upcoming investigation.

Nora's pre-course assessment was rated quite a bit higher than the group average, and was given the highest rating of all three case study teachers. Similar to her classroom lesson, the assessment was based on the phenomenon of using antibiotics for a bacterial infection and the competitive advantage of antibiotic resistant bacteria. Nora's assessment was an adapted version of the assessment provided by her curricular materials, which as previously noted were closely aligned to the NGSS. Nora reported modifying the assessment by "embedding sentence frames to support struggling writers" and by adding a section for her students to draw models. The assessment required students to explain the phenomenon and integrated all three dimensions in a relatively coherent fashion, but a number of the items were determined to be either above or below grade level.

Nora's Course Experience. Nora decided to participate in the course because she wanted to push herself as an educator, particularly in the area of assessment which she regarded as a key weakness. She reflected, "I think assessment is absolutely critical and it really should drive everything else…. I feel like I'm starting to reach a plateau a little bit, so I needed something to kind of push me and stretch me more." However, despite her enthusiasm and strong desire for self-growth, Nora became overwhelmed by the course and other events in her life, and barely made it to the end. She explained that she felt "completely burnt out" by teaching and was actually struggling with whether to stay in the profession. She ended up missing four of the twelve course sessions, but received a good deal of individual support from the instructors as well as her coach. Although she worked towards designing an assessment during the course, she did not actually end up generating a complete assessment that she could give to her students.

A primary motivation for Nora was to develop assessments that support equity and ensure that all of her students view science as interesting and relevant. For Nora, this meant using authentic phenomena in her teaching and assessments that students can connect to, ask questions about, "offer pathways to understanding." Nora shared during the class that she strives to give students "data for sense-making, like puzzle pieces that can be arranged in various ways for kids to have different interpretations." She talked frequently about attending to equity, embracing each student where they are, supporting English language learners and students with dyslexia, and making sure that no students feel excluded. She commented, "I think it's so important for us as teachers to think about it from our students perspective and like if our students are really going to buy into these experiences we're creating for them. They have to feel like they're worthy of that." Nora's passion for attending to her students' interests and identities, and the fact that the course provided resources to support this endeavor, helped keep her engaged in the course even though she had multiple "meltdowns" throughout. Although she did not attend the last two sessions, she did take part in all of the post-course data collection. Further, her reflections on the course were primarily positive.

Nora noted that she appreciated "just having structured time to sit and really reflect on, you know, how the standards work. How to think about them from an assessment perspective and also a student perspective. I think that was probably the most enlightening component for me, is how do you approach it from a student angle?" She also explained that she had been "desperately trying to find resources like this on my own the past few years." Perhaps most importantly, Nora felt that the course offered her a supportive community at a time when she was "in a dark place." She shared, "It's just so nice to feel connected. It's so isolating in this job, especially in rural communities."

On a more optimistic note, Nora expressed that the course provided her with concrete resources that she took back to her school's science department, which has helped them to get on the same page. She explained, "I've been able to give them those checklists to help them go through and just use them to make their assessments as five dimensional as they can be. And we've been starting to talk about five dimensions as a department, too, and trying to build it into our own language as professionals." In addition, Nora started working closely with a colleague to write 5D assessments to use in their eighth-grade classrooms, something she felt was urgently needed and helped to ensure that, as teachers, they were thinking about assessment in the same way.

Post Course Data. Nora developed her post-course assessment from scratch, rather than modifying assessment resources from her curriculum like she did in her pre-course assessment. The assessment asks students to figure out how two dogs with the same biological parents can have different muscular phenotypes. The students draw models and then provide written explanations to explain how and why the dogs have different traits. Nora created a modified version of the assessment for her students who were English learners, had IEPs, or were dygraphic that included "simple modifications that would not deter from the students still having to show their 3D ability." She also felt that an assessment focused on dogs would encourage students' engagement because "students LOVE their pets and so dogs were an easy way to connect to that." This assessment was among the highest scoring of all the course participants' submitted assessments, including their pre- and post-assessments, and was rated well above the other two case study teachers' assessments. The phenomenon was presented with a specific set of observations and required students to use multiple modalities to make sense of and explain. In addition there was strong coherence throughout and included well-integrated DCIs, SEPs and CCCs.

Nora's survey scores did not show a great deal of movement from pre to post. Her post-scores fell in-between the other two case study teachers on all of the subscales, and were generally near the average for the full group of participants. She had a slight gain in her scores on the argument practices and assessment vision subscales, stayed the same on the interest and identity vision subscale, and decreased slightly on the interest and identity practice subscale. Overall, this self-report survey may not have surfaced the substantial gains that Nora made as well as those evidenced by her submitted assessments and recorded classroom observations.

Similar to her pre-course lesson recording, Nora's post-course lesson was grounded in phenomenon. However this lesson was conducted in-person which showcased her students' active participation and also included opportunities for sense-making. The lesson focused on heredity and genetics, with the driving question being: How do chromosomes cause cattle to have different phenotypes? Nora shows the class a picture of a muscular cow and asks them to consider how a particular set of genes may have led to the cow's phenotype. The teacher distributes four different "family cases" and has students work in groups to identify patterns in the inheritance of traits. Next the students take part in a gallery walk and look at each others' work to confirm whether the patterns they found hold true for all of the cases. Lastly the teacher leads the class in a consensus building discussion about the relationships between alleles, genes, and phenotypes.

Summary. Nora came in with a good deal of prior knowledge and professional learning experiences related to the NGSS, used an NGSS-aligned storylined curriculum in her classroom, and had a strong desire to improve her assessment writing skills. She was especially committed to supporting all students, ensuring that they would find her lessons interesting and relevant to students, and providing equitable opportunities for them to learn and showcase their knowledge. Nora used phenomenon-based instructional materials and assessments prior to the course, but showed clear improvements in both areas from pre- to post-course. Unfortunately she found it very difficult to keep up with the demands of the course, and although she received individual support from the course instructors and her coach, she missed the final two sessions and did not complete the course designed assessment.

After the course ended, Nora's actions demonstrated that she was committed to taking up what she had learned from the course, despite not completing it. She designed a post-course assessment from scratch that was one of the highest rated assessments in the entire group. She also shared what she had learned with the other teachers in her department, including the language around 5D assessment and several of the course resources. In addition, she engaged in collaborative assessment writing with other teachers in her building to generate 5D assessment materials they could immediately use with their students.

Cross Case Comparison: Differences in the Teachers' Learning Trajectories

The three case study teachers entered the course at very different places with respect to their standards-aligned beliefs, skills in creating a 5D assessment, and the extent to which their classroom instruction reflected the 5D vision. Although Dot was one of the most experienced teachers in the entire group, she had some of the lowest scores on all of these measures and by all accounts had far less experience than her peers did in understanding the NGSS. She worked in an isolated rural community, with few science education colleagues, and primarily relied on her own "tried and true" assessment and instruction practices. Her professional background standards in sharp contrast to Mary, who also worked in an isolated rural community but was heavily engaged in state-level PL around the NGSS and held some of the highest pre-course standards-aligned beliefs. Yet Mary exhibited substantial weaknesses in creating a 5D assessment and her instructional practice was not particularly reflective of the 5D vision. Nora, on the other hand, worked in a considerably less remote but still rural community, in a district that provided strong support for the implementation of the NGSS including a storylined curriculum. Nora entered the course with some of the strongest 5D assessment writing skills in the group. The baseline abilities of these three teachers highlights the large variation that exists among rural US science teachers, although they all shared the desire to take part in a 5D assessment course with the goal of becoming more skilled in this area.

All three of these teachers reported experiencing extensive, personalized support during the course that they attributed to their growth. Dot frequently raised questions, voiced concerns about not understanding central concepts, and requested resources from her peers and coach. In response, she received a great deal of individual attention for which she was extremely grateful and also felt a bit guilty about. Mary was one of the teachers who provided the most support to Dot, playing a mentorship role that she was very comfortable with and that helped her stay engaged and motivated despite her extremely busy schedule. Nora struggled the most to keep up with the course, largely due to personal and health-related challenges. However she met off-line with the course instructors and her coach several times, who encouraged her to persevere and meet her learning goals. The three teachers all had unique experiences as they worked their way through the semester-long course, with each facing substantial obstacles that might have otherwise prevented them from experiencing success. Yet they found a way to overcome these obstacles, largely due to the strong community and ongoing support they received and/or provided to others.

The three teachers all left the course with considerably improved skills in developing 5D assessments, especially by incorporating phenomena that prompted students to engage in sensemaking. Increased attention to the use of phenomenon and encouraging sensemaking was also evident in their post-course recorded lessons. At the same time, Dot continued to have doubts about her ability to develop standards-aligned assessments for all of her units and classes, noting that she still saw a place for more traditional multi-choice tests. However, given her limited prior experience and professional resources, this growth in skill is rather remarkable. Mary gained not only stronger 5D assessment writing skills but also became more convinced of their appropriateness for all students, including those in advanced classes. Nora took from the course both an expanded skill set and a collection of resources that she encouraged her middle school science colleagues to engage with, helping her to become a more knowledgeable and confident leader in school-based assessment writing efforts. An important takeaway from the variation in these teachers' growth trajectories is that the online 5D assessment course was well-suited to support this degree of diversity and move each of them forward in measurable ways.

Discussion and Implications

Rural teachers are a varied group in many ways and as our case analyses of three rural teachers exemplify, even the designation of "rural" itself is highly variable. The teachers who were selected as cases in this paper differed with respect to the state they worked in (which has implications for their exposure to NGSS and 3D professional learning opportunities), the remoteness of their schools, the nature of their classroom teaching load, their histories as science teachers, and their personal and professional capacities at the time they took the 5D course. As we have argued, these characteristics influenced how teachers entered into the course as well as how they experienced the tools, training, and support it provided. The unique professional backgrounds and course experiences of the case study teachers challenge the treatment of rural teachers as a homogeneous group because they are "not-urban" (Biddle & Azano, 2016).

Crumb et al. (2023)'s framework of rural cultural wealth lays out four areas of strength in rural communities – rural resourcefulness, rural ingenuity, rural familism and rural community unity. This paper foregrounded two of these four strengths in our case analysis: rural resourcefulness and community unity. We can see their resourcefulness from the very beginning of the study when the case study teachers (along with their colleagues in the full sample) sought to overcome the limited in-person access to professional development for rural teachers by signing up for the online 5D course. Each of the case study teachers was resourceful in their own ways as well. Dot frequently requested additional information and resources from her classmates to support her development of 5D aligned assessments. She commented numerous times that she hoped to “steal” assessment questions that others were developing, which was always met with enthusiastic agreement. Mary, in particular, would often scour the internet for resources that might help Dot in addition to providing assessment and curriculum documents that she had developed herself. Nora’s commitment to self-improvement, particularly in the area of 5D assessment, was palpable. Despite the fact that she faced hurdles all along the way, from serious health challenges to questioning whether she wanted to continue in the teaching profession, Nora moved past her almost insurmountable feeling of “overwhelm” to gain “great clarity on what my intentions are as a teacher.” She not only generated a self-made, high quality 5D aligned assessment after the course was over, but she began working in close collaboration with the other science teachers at her school to support their learning.

Elements of community unity were also evident in our analysis, as the teachers came together to leverage their “composite assets.” Attention to community was designed into the 5D course, for example through its online design, recruiting of rural teachers across multiple states, frequent small group work, and pairing teachers with coaches. The participating teachers quickly developed a culture of offering ideas, resources, and support to one another as they embarked on the complex task of creating 5D assessments. Each of the three case study teachers leaned into the community as an important resource for their learning. All of them described feeling empowered and energized by the support they received from their colleagues, contrasting the course experience with the frequent isolation they feel as rural science teachers. For example, Mary recalled telling the other teachers at her school how excited she was to meet science teachers who also worked in very remote communities and had teaching caseloads that were similar to her own.

Findings from this case study paper offer a number of implications for professional learning opportunities designed to meet the needs of rural educators. First, it is critical to recognize the variety within rural designations, which primarily attend to how far away a community is from an urban center. Although we found that all of the case study teachers felt a degree of isolation due to being in a rural community, some had established much more extensive science education networks than others and had different experiences accessing NGSS-related professional learning and resources. Second, but very much related, is the importance of recognizing that rural teachers have unique professional contexts. The rural teacher in our study varied greatly with respect to their experience levels, course loads, curriculum, and knowledge of current ideas and practices. All of this variation means that rural teachers will not likely enter into any given PL experience at the same level and must be provided with entry points that map closely to their specific needs. Third, professional learning can and should be intentionally designed to leverage rural teachers’ cultural wealth. For example, programs can support rural resourcefulness by anticipating challenges and providing opportunities, tools, and encouragement for rural teachers to work through them productively. Programs can also promote community unity by offering structures that support relationship building and ongoing collaboration (Inouye et al., 2023). We anticipate programs that explicitly recognize and build on rural teachers’ cultural wealth will experience broader and more sustainable success.

Related Paper Set Discussion

As access to 5D vision-aligned instructional materials and assessments continues to grow, there continues to be a need to enhance teachers' pedagogical design capacity. As designers of their own curriculum and often teaching multiple grade levels or content areas (Wingert et al., 2022), this course met a critical need that allowed them to use what they learned to both design new assessments and adapt existing assessments. In addition, teachers recognized characteristics of vision-aligned instruction that could enhance the coherence between instruction and assessment.

This paper set documented the ways in which our course supported the professional learning needs of rural teachers, with a particular focus on enhancing teachers' access to high quality professional learning and fostering collaboration among adult learners. In addition, we designed professional learning materials and tools to support the development of teachers' pedagogical design capacity. These papers documented the ways in which we sought to enhance the accessibility of our tools through revisions and support teachers as they shifted their assessment practices.

This paper set also revealed the complexity of the integrating student interest in teachers' design work. The *Student and Community Interest (and Identity) Inventory* served as a tool that teachers could use to elicit information about their students. For some teachers, the general nature of the tool was not specific enough to help inform their phenomena choices. Although teachers were given the opportunity to adapt the inventory for use in their classroom, few did. For some teachers, the task of needing to administer the inventory was enough to prompt teachers to use what they know about their students to inform their design decisions. We hope that we can provide future support for how teachers might adapt the tools and encourage more frequent surveys of their students so that they can help inform the design of more meaningful phenomenon-driven assessments and instruction. At the same time, we might need to broaden teachers' understanding of the types of phenomena that teachers might consider for use in a phenomenon-driven assessment to better engage student interest (Cooper & Lo, 2024).

Future work includes adapting our course to focus on the adaptation of assessments found in 5D vision-aligned curriculum materials. The teachers from our study were not required to use a particular set of instructional materials. Considering the shifts that we observed in teachers' assessment practices, we posit that situating our work in the context of 5D vision-aligned instructional materials might help us to focus on the areas of 5D assessment design that required further support and focus teachers' efforts on adapting the provided materials to better engage students' interests and identities.

References

Achieve, (2018) *Task annotation project in science (TAPs)*. NextGen Science, WestEd. Retrieved March 2024, from https://www.achieve.org/our-initiatives/equip/tools-subject/science/task-annotation-project-science

Achieve, NextGen Storylines, & STEM Teaching Tools. (2017). *Using phenomena in NGSS-designed lessons and units*.

Aikenhead, G. S., Calabrese Barton, A., & Chinn, P. W. U. (2006). Toward a politics of place-based science education. *Cultural Studies of Science Education*, *1*(2), 403-416. https://doi.org/10.1007/s11422-006-9015-z

Antrop-González, R., & De Jesús, A. (2006). Toward a theory of *critical care* in urban small school reform: Examining structures and pedagogies of caring in two Latino community-based schools. *International Journal of Qualitative Studies in Education*, *19*(4), 409-433. https://doi.org/10.1080/09518390600773148

Avery, L. M. (2013). Rural science education: Valuing local knowledge. *Theory Into Practice*, *52*(1), 28-35. https://doi.org/10.1080/07351690.2013.743769

Avery, L. M., & Kassam, K.-A. (2011). *Phronesis:* Children's local rural knowledge of science and engineering. *Journal of Research in Rural Education*, *26*(2), 1-18.

Azano, A. P., Brenner, D., Downey, J., Eppley, K., & Schulte, A. K. (2020). *Teaching in rural places: Thriving in classrooms, schools, and communities*. New York, NY: Routledge.

Azano, A. P., Biddle, C., "Disrupting dichotomous traps and rethinking problem formation for, & rural educaton", T. R. E., Vol. 40 No. 2, pp. 4-11, doi: 10.35608/ruraled.v40i2.845. (2019). Disrupting dichotomous traps and rethinking problem formation for rural education. *The Rural Educator*, *40*, 4-11. https://doi.org/10.35608/ruraled.v40i2.845

Azevedo, F. S. (2018). An inquiry into the structure of situational interests. *Science Education*, *102*(1), 108-127. https://doi.org/10.1002/sce.21319

Bang, M., Brown, B., Calabrese Barton, A., Rosebery, A. S., & Warren, B. (2017). Toward more equitable learning in science. In C.V. Schwartz, C. Passmore & B.J. Reiser (Eds.), *Helping students make sense of the world using Next Generation Science and Engineering Practices* (pp. 33-58), Arlington, VA: NSTA Press.

Bang, M., Curley, L., Kessel, A., Marin, A., & Suzokovich, E. (2012, April). Muskrat theories, tobacco in the streets, and living Chicago as indigenous land. Annual Meeting of the American Educational Research Association, Vancouver, British Columbia.

Banilower, E. R., Smith, P. S., Malzahn, K. A., Plumley, C. L., Gordon, E. M., & Hayes, M., L. (2018). *Report of the National Survey of Science and Mathematics Education*. Horizon Research, Inc.

Basu, S. J., & Calabrese Barton, A. (2007). Developing a sustained interest in science among urban minority youth. *Journal of Research in Science Teaching*, *44*(3), 466-489. https://doi.org/10.1002/tea.20143

Bell, P., Stromholt, S., Penuel, W. R., & Van Horne, K. (2016, March). *How to make science instruction compelling to students: Designing formative assessments to build on learners' interests and knowledge* National Science Teachers Association (NSTA) Annual Meeting, Nashville, TN.

Berlyne, D. E. (1970). Novelty, complexity, and hedonic value. *Perception & Psychophysics*, *8*, 270-286. https://doi.org/10.3758/BF03212593

Biddle, C., & Azano, A. P. (2016). Constructing and reconstructing the "rural school problem" a century of rural education research. *Review of Research in Education*, *40*(1), 298-325.

Blumenfeld, P., Soloway, E., Marx, R. W., Krajcik, J. S., Guzdial, M., & Palincsar, A. S. (1991). Motivating project-based learning: Sustaining the doing, supporting the learning. *Educational Psychologist*, *26*(3&4), 369-398. https://doi.org/10.1080/00461520.1991.9653139

Braaten, M., & Sheth, M. (2017). Tensions teaching science for equity: Lessons learned from the case of Ms. Dawson. *Science Education*, *101*(1), 134-164. https://doi.org/10.1002/sce.21254

Bransford, J. D., & Schwartz, D. L. (1999). Rethinking transfer: A simple proposal with interesting implications. *Review of Research in Education*, *24*, 61-101. https://doi.org/10.3102/0091732X024001061

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77-101.

Brown, M. W. (2009). Toward a theory of curriculum design and use: Understanding the teacher-tool relationship. In J. T. Remillard, B. A. Herbel-Eisenman, & G. M. Lloyd (Eds.), *Mathematics teachers at work: Connecting curriculum materials and classroom instruction* (pp. 17-36). Routledge.

Calderon, D., Lees, A., Swan Waite, R., & Wilson, C. (2021). 'Crossing the bridge': land education teacher professional development. *Professional Development in Education*, *47*(2-3), 348-362. https://doi.org/10.1080/19415257.2021.1891957

CAST (2018). *Universal design for learning guidelines version 2.2 [graphic organizer].* Wakefield, MA: Author.

Carpenter, S. L., Macias, M., Arevalo, E., Hansen, A. K., Stone, E. M., & Bianchini, J. A. (2023). Coordinated and intersecting: How preservice secondary science teachers understand science and engineering practices and instructional principles for diverse students. *Electronic Journal for Research in Science and Mathematics Education*, *27*(2), 25-56.

Clark-Ibañez, M. (2004). Framing the social world with photo-elicitation interviews. *The Behavioral Scientist*, *47*(12), 1507-1527. https://doi.org/10.1177/0002764204266236

Cohen, J. (2013). *Statistical power analysis for the behavioral sciences*. Academic press.

Cooper, S. L., & Lo, A. S. (2024, March 17). Supporting Teachers in the Selection of Meaningful Phenomena for Assessment Design [Conference Paper]. NARST 2024 Annual International Conference, Denver, CO.

Corbett, M. (2016). Rural futures: Development, aspirations, mobilities, place, and education. *Peabody Journal of Education*, *91*(2), 270-282.

Corntassel, J., & Hardbarger, T. (2019). Educate to perpetuate: Land-based pedagogies and community resurgence. *International Review of Education*, *65*, 87-116. https://doi.org/10.1007/s11159-018-9759-1

Cowie, B., Jones, A., & Otrel-Cass, K. (2010). Re-engaging students in science: Issues of assessment, funds of knowledge and sites for learning. *International Journal of Science and Mathematics Education*, *9*, 347-366. https://doi.org/10.1007/s10763-010-9229-0

Crumb, L., Chambers, C., Azano, A., Hands, A., Cuthrell, K., & Avent, M. (2023). Rural cultural wealth: Dismantling deficit ideologies of rurality. *Journal for Multicultural Education*, *17*(2), 125-138. https://doi.org/10.1108/jme-06-2022-0076

Design-Based Research Collective. (2003). Design-based research: An emerging paradigm for educational inquiry. *Educational Researcher*, *32*(1), 5-8. https://doi.org/10.3102/0013189X032001005

DiGiacomo, D., Van Steenis, E., Van Horne, K., & Penuel, W. R. (2018). The material and social constitution of interest. *Language, Culture, and Social Interaction*, *19*, 51-60. https://doi.org/10.1016/j.lcsi.2018.04.010

DiNicolo, C. P., Yu, M., Crowley, C. B., & Gabel, S. L. (2017). Reimagining critical care and problematizing sense of school belonging as a response to inequality for immigrants and children of immigrants. *Review of Research in Education*, *41*(1), 500-530. https://doi.org/10.3102/0091732X17690498

Drescher, J., Podolsky, A., Reardon, S. F., & Torrance, G. (2022). The geography of rural educational opportunity. *The Russell Sage Foundation Journal of the Social Sciences*, *8*(3), 123-149. DOI: 10.7758/RSF.2022.8.3.05

Durr, T., Kampmann, J., Hales, P., & Browning, L. (2020). Lessons learned from online PLCs of rural STEM teachers. *The Rural Educator*, *41*(1), 20-26. https://doi.org/10.35608/ruraled.v41i1.555

Eppley, K. (2015). "Hey, I saw your grandparents at Walmart": Teacher education for rural schools and communities. *The Teacher Educator*, *50*(1), 67-86.

Eppley, K. (2017). Rural science education as social justice. *Cultural Studies of Science Education*, *12*(1), 45-52. https://doi.org/10.1007/s11422-016-9751-7

Fine, C. G. M., & Furtak, E. M. (2020). A framework for science classroom assessment task design for emergent bilingual learners. *Science Education, 104*(3), 393-420. https://doi.org/10.1002/sce.21565

Fitzmaurice, G. M., Laird, N. M., & Ware, J. H. (2012). *Applied longitudinal analysis*. John Wiley & Sons.

Furtak, E. M., Kang, H., Pellegrino, J. W., Harris, C. J., Krajcik, J. S., Morrison, D., Bell, P., Lakhani, H., Suárez, H., Buell, J. Y., Nation, J., Henson, K., Fine, C., Tschida, P., Fay, L., Biddy, Q., Penuel, W. R., & Wingert, K. (2020). Emergent design heuristics for three-dimensional classroom assessments that promote equity In M. Gresalfi & I. S. Horn (Eds.), *The interdisciplinarity of the learning sciences* (Vol. 3). International Society of the Learning Science.

Gaver, W. W., Boucher, A., Pennington, S., & Walker, B. (2004). Cultural probes and the value of uncertainty. *Interactions*, *11*(5), 53-56.

González, N., Moll, L. C., Floyd-Tenery, M., Rivera, A., Rendon, P., Gonzales, R., & Amanti, C. (1993). *Teacher research on funds of knowledge: Learning from households*.

Gruenewald, D. A. (2003). Foundations of place: A multidisciplinary framework for place-conscious education. *American Educational Research Journal*, *40*(3), 619-654. https://doi.org/10.3102/00028312040003619

Gruenewald, D. A. (2003). The best of both worlds: A critical pedagogy of place. *Educational Researcher*, *32*(4), 3-12.

Gutiérrez, K. D., & Rogoff, B. (2003). Cultural ways of knowing: Individual traits or repertoires of practice. *Educational Researcher*, *32*(5), 19-25. https://doi.org/10.3102/0013189X032005019

Gutierrez, E., & Terrones, F. (2023). *Small and sparse: Defining rural school districts for K-12 funding*. Urban Institute, Center on Education Data and Policy.

Hagenah, S., & Thompson, J. (2021). Teachers' attempts to respond to students' lived experiences. *Journal of Science Teacher Education*. https://doi.org/10.1080/1046560X.2020.1869887

Hammack, R., Stanton, C. R., & Boyle, J. (2022). "Step Outside": A portrait of an exemplary rural K-8 science educator. *Journal of Research in Science Teaching, 60*(3), 544-567.

Hammerness, K. (2001). Teacher's visions: The role of personal ideals in school reform. *Journal of Educational Change, 2*(2), 143-163. https://doi.org/10.1023/A:1017961615264

Hammerness, K. (2006). *Seeing through teachers' eyes: Professional ideals and classroom practices* (Vol. 46). Teachers College Press.

Harackiewicz, J. M., Smith, J. L., & Priniski, S. J. (2016). Interest matters: The importance of promoting interest in education. *Policy Insights from the Behavioral and Brain Sciences*, *3*(2), 220-227. https://doi.org/10.1177/2372732216655542

Harris, R. S., & Hodges, C. B. (2018). STEM education in rural schools: Implications of untapped potential. *National Youth-At-Risk Journal*, *3*(1), 3-12. https://doi.org/10.20429/nyarj.2018.030102

Hartman, S. L., Roberts, J. K., Schmitt-Wilson, S., McHenry-Sorber, E., & Buffington, P. J. (2022). National Rural Education Association Research Agenda–2022-2027: A closer look at the research priorities. *The Rural Educator*, *43*(3), 59-66. https://doi.org/10.55533/2643-9662.1349

Haywood, B. K. (2014). A "sense of place" in public participation in scientific research. *Science Education*, *98*(1), 64-83. https://doi.org/10.1002/sce.21087

Henson, K., & Penuel, W. R. (2023, April). *Attending to student interest and identity in instructional phenomena.* 2023 NARST Annual International Conference, Chicago, IL.

Holton, J. A. (2007). The coding process and its challenges. In A. Bryant & K. Charmaz, *The SAGE Handbook of Grounded Theory* (pp. 265–289). SAGE Publications Ltd.

Hopkins, M., & Gates, Z. (2019). *Findings from a survey of state science leaders: Year 2.* Boulder, CO: National Center for Research in Policy and Practice.

Huffling, L. D., Carlone, H. B., & Benavides, A. (2017). Re-inhabiting place in contemporary rural communities: Moving toward a critical pedagogy of place. *Cultural Studies of Science Education*, *12*, 33-43.

Hulleman, C., & Harackiewicz, J. M. (2009). Promoting interest and performance in high school science classes. *Science*, *326*(5958), 1410-1412. https://doi.org/10.1126/science.1177067

Hunt-Barron, S., Tracy, K. N., Howell, E., & Kaminski, R. (2015). Obstacles to enhancing professional development with digital tools in rural landscapes. *Journal of Research in Rural Education*, *30*(2), 1-14.

Inouye, M., Houseal, A. K., Gunshenan, C., McReynolds, A., & Perkins, M. (2023). Exploring collaborative professionalism as a means of virtually supporting rural teachers. *The Rural Educator*, *44*(1), 14-27.

Kang, H., & Furtak, E. M. (2021). Learning theory, classroom assessment, and equity. *Educational Measurement: Issues and Practice*, *40*(3), 73-82. https://doi.org/10.1111/emip.12423

Kim, M., & Schallert, D. L. (2011). Building caring relationships between a teacher and students in a teacher preparation program. *Teaching and Teacher Education*, *27*(7), 1059-1067. https://doi.org/10.1016/j.tate.2011.05.002

Kirchgasler, K. L. (2019). Strange precipitate: How interest in science produces different kinds of students. In W. Letts & S. Fifield (Eds.), *STEM of desire: Queer theories and science education* (pp. 191–208). Brill. https://doi.org/10.1163/9789004331068_009

Krajcik, J. S., 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

Krist, C. (2020). Building trust: Supporting vulnerability for doing science in school. In M. Gresalfi & I. S. Horn (Eds.), *The interdisciplinarity of the learning sciences* (pp. 270-277). International Society of the Learning Sciences.

Krist, C. (in press). Striving for relationality: Teacher responsiveness to relational cues when eliciting students' science ideas. *Cognition and Instruction*.

Krist, C., & Suárez, E. (2018). Doing science with fidelity to persons: Instantiations of caring participation in science practices. In J. Kay & R. Luckin (Eds.), *Proceedings of the 13th*

International Conference of the Learning Sciences (Vol. 1, pp. 424-431). International Society of the Learning Sciences.

Kumpulainen, K., & Rajala, A. (2017). Dialogic teaching and students' discursive identity negotiation in the learning of science. *Learning and Instruction*, *48*, 23-31. https://doi.org/10.1016/j.learninstruc.2016.05.002

Lavalley, M. (2018). O*ut of the loop: Rural schools are largely left out of research and policy Ddscussions, exacerbating poverty, inequity, and isolation*. Center for Public Education: National School Boards Association.

Learning in Places Collaborative. (2022). *Framework: Complex socio-ecological systems*. Learning in Places.

Lee, O. (2020). Making everyday phenomena phenomenal. *Science and Children, 58*(1), 56-61.

Lee, O., & Grapin, S. E. (2022). The role of phenomena and problems in science and STEM education: Traditional, contemporary, and future approaches. *Journal of Research in Science Teaching, 59*(7), 1301-1309. https://doi.org/10.1002/tea.21776

Llopart, M., & Esteban-Guitart, M. (2017). Strategies and resources for contextualising the curriculum based on the funds of knowledge approach: A literature review. *Australian Education Research Journal*, *44*, 255-274.

Luehmann, A. L. (2009). Accessing resources for identity development by urban students and teachers: Foregrounding context. *Cultural Studies of Science Education, 4*(1), 51-66. https://doi.org/10.1007/s11422-008-9139-4

Lo, A. S., Cooper, S. L., Hermann Abell, C. F., Cherbow, K., Allen, A., (2024, March 17-20). *Lessons Learned from Designing 5D Professional Learning for Rural Science Teachers* [Related paper set presentation]. 97th NARST International Conference, Denver, CO.

Lo, A. S., Glidewell, L., O'Connor, K., Allen, A., Herrmann-Abell, C. F., Penuel, W. R., Wingert, K., & Lindsay, W. (2022). Promoting shifts in teachers' understanding and use of phenomena in instruction and assessment. In C. Chinn, E. Tan, & Y. Kali (Eds.), Proceedings of the 16th International Conference of the Learning Sciences - ICLS 2022 (pp. 1145-1148). International Society of the Learning Sciences. https://doi.org/10.22318/icls2022.1145

Lo, A. S., Penuel, W. R., & Wingert, K. (2022, April). *Supporting teachers in designing assessments aligned to the vision of the Framework: Findings from two design studies* Annual Meeting of the American Educational Research Association, San Diego, CA.

Lobato, J. (2012). The actor-oriented transfer perspective and its contribution to educational research and practice. *Educational Psychologist*, *47*(3), 232-247.

Long, D. E., & Avery, L. M. (2017). Cultural studies of rural science education. *Cultural Studies of Science Education*, *12*(1), 1–5. DOI:10.1007/s11422-016-9744-6

Lowell, B. R., & McNeill, K. L. (2020). Using the student hat to push on multiple goals in teacher professional learning. In M. Gresalfi, I. S. Horn, N. Enyedy, H.-J. So, V. Hand, K. Kackson, S. E. McKenney, A. Leftstein, & T. M. Philip (Eds.), *Proceedings of the International Conference of the Learning Sciences* (Vol. 4, pp. 2231-2244). International Society of the Learning Sciences.

Madkins, T. C., & McKinney de Royston, M. (2019). Illuminating political clarity in culturally relevant science instruction. *Science Education*, *103*, 1013-1346. https://doi.org/10.1002/sce.21542

Maltese, A. V., & Tai, R. H. (2011). Pipeline persistence: Examining the association of educational experiences with earned degrees in STEM among U.S. students. *Science Education*, *95*, 877-807.

McNeill, K. L., Affolter, R., & Reiser, B. J. (2022). Anchoring science professional learning in curriculum materials enactment: Illustrating theories in practice to support teachers' learning. In A. C. Superfine, S. Goldman, & M.-L. M. Ko (Eds.), *Teacher learning in changing contexts: Perspectives from the learning sciences* (pp. 47-68). Routledge. https://doi.org/10.4324/9781003097112-5
Moll, L. C., Amanti, C., Neff, D., & González, N. (1992). Funds of knowledge for teaching: Using a qualitative approach to connect homes and classrooms. *Theory into Practice*, *31*, 132-141. https://doi.org/10.1080/00405849209543534
Morales-Doyle, D. (2017). Students as curriculum critics: Standpoints with respect to relevance, goals, and science. *Journal of Research in Science Teaching*, *55*(5), 749-773. https://doi.org/10.1002/tea.21438
Morales-Doyle, D., Price, T. C., & Chappell, M. J. (2019). Chemicals are contaminants, too: Teaching appreciation and critique of science in the era of Next Generation Science Standards. *Science Education*, *103*(6), 1347-1388. https://doi.org/10.1002/sce.21546
Nasir, N. S., & Cooks, J. (2009). Becoming a hurdler: How learning settings afford identities. *Anthropology and Education Quarterly*, *40*(1), 41-61. https://doi.org/10.1111/j.1548-1492.2009.01027.x
Nasir, N. S., & Hand, V. (2008). From the court to the classroom: Opportunities for engagement, learning, and identity in basketball and classroom mathematics. *Journal of the Learning Sciences*, *17*(2), 143-179. https://doi.org/10.1080/10508400801986108
National Academies of Sciences Engineering and Medicine. (2018). *How people learn II: Learners, cultures, and contexts*. National Academies Press. https://doi.org/10.17226/24783
National Center for Education Statistics. (2018). *Selected statistics from the public elementary and secondary education universe: School year 2015-16*. U.S. Department of Education. Retrieved January 22 from https://nces.ed.gov/pubs2018/2018052/tables/table_04.asp
National Center for Education Statistics (2022). Locale classifications and criteria. Retrieved March 16, 2024 from https://nces.ed.gov/programs/edge/docs/LOCALE_CLASSIFICATIONS.pdf
National Research Council. (2001). *Knowing what students know*. Washington, DC: National Academies Press.
National Research Council. (2012). *A framework for K-12 science education: Practices, crosscutting concepts, and core ideas*. National Research Council. https://doi.org/10.17226/13165
National Research Council. (2014). *Developing assessments for the Next Generation Science Standards.* National Academies Press. https://doi.org/10.17226/1840
NextGen Science Storylines Team. (2019). Why Do Dead Things Disappear Over Time? [Curriculum Materials] https://www.nextgenstorylines.org/why-do-dead-things-disappear-over-time
NextGen Science, & EdReports. (2021). *Critical features of Instructional materials design for today's science standards*. WestEd and EdReports.
NGSS Lead States. (2013). *Next Generation Science Standards: For states, by states*. National Academies Press. https://doi.org/10.17226/18290
Northwestern University and Inquiry Hub Partnership team. (2018). Why don't antibiotics work the way they used to: A next generation science storyline. [Curriculum Materials] http://www.nextgenstorylines.org/why-dont-antibiotics-work-like-they-used-to
O'Connor, K., Allen, A.-R., & Penuel, W. R. (2024, March 18). Examining Science Teachers' Conceptions of Student Interest as a Consideration in Designing Assessments [Conference Paper]. NARST 2024 Annual International Conference, Denver, CO.

Patro, E. T. (2008). Teaching aerobic cell respiration using the 5Es. *The American Biology Teacher*, *70*(2), 85-87. https://doi.org/10.1662/0002-7685(2008)70[85:TACRUT]2.0.CO;2
Penuel, W. R. (2019). Infrastructuring as a Practice of Design-Based Research for Supporting and Studying Equitable Implementation and Sustainability of Innovations. Journal of the Learning Sciences, 1-19. https://doi.org/10.1080/10508406.2018.1552151
Penuel, W. R. (2020). Commentary: To promote equity, prepare students for what science could be. *NSTA Reports*. https://doi.org/https://www.nsta.org/promote-equity-prepare-students-what-science-could-be
Penuel, W. R., Allen, A.-R., Henson, K., Campanella, M., Patton, R., Rademaker, K., Reed, W., Watkins, D. A., Wingert, K., Reiser, B. J., & Zivic, A. (2022). Learning practical design knowledge through co-designing storyline science curriculum units. *Cognition and Instruction, 40*(1), 148-170. https://doi.org/10.1080/07370008.2021.2010207
Penuel, W. R., Lo, A. S., Jacobs, J. K., Gardner, A., Stuhlsatz, M. A. M., & Wilson, C. (2019, April). *Tools for supporting teachers to build quality 3D assessment tasks.* NARST Annual Meeting, Baltimore, MD.
Penuel, W. R., Raza, A., Salinas del Val, Y., Salinas-Estevez, R., Williamson, E., Smith, J., & Gill, Q. (in press). Making formative use of student experience data to promote equity. *Science Scope*. https://doi.org/10.1080/08872376.2024.2314675
Penuel, W. R., Reiser, B. J., McGill, T. A. W., Novak, M., Van Horne, K., & Orwig, A. (2022). Connecting student interests and questions with science learning goals through project-based storylines. *Disciplinary and Interdisciplinary Science Education Research*, *4*(1), 1-27. https://doi.org/10.1186/s43031-021-00040-z
Penuel, W. R., Turner, M. L., Jacobs, J. K., Van Horne, K., & Sumner, T. (2019). Developing tasks to assess phenomenon-based science learning: Challenges and lessons learned from building proximal transfer tasks. *Science Education, 103*(6), 1367-1395. https://doi.org/10.1002/sce.21544
Penuel, W. R., & Van Horne, K. (2016). Prompts for Integrating Crosscutting Concepts Into Assessment and Instruction. STEM Teaching Tools: Teaching Tools for Science, Technology, Engineering and Math (STEM) Education. http://stemteachingtools.org/brief/41
Potvin, A. S. (2018). *Designing for teacher-student relationships: An investigation into the emotional and relational dimensions of co-design* University of Colorado Boulder].
Potvin, A. S. (2021). "Students speaking to you": Teachers listen to student surveys to improve classroom environment. *Learning Environments Research*, *24*, 239-252. https://doi.org/10.1007/s10984-020-09330-1
Potvin, P., & Hasni, A. (2014). Interest, motivation, and attitude towards science and technology at K-12 levels: A systematic review of 12 years of educational research. *Studies in Science Education*, *50*(1), 85-129. https://doi.org/10.1080/03057267.2014.881626
Reiser, B. J. (2004). Scaffolding complex learning: The mechanisms of structuring and problematizing student work. *The Journal of the Learning Sciences*, *13*(3), 273-304. https://doi.org/10.1207/s15327809jls1303_2
Reiser, B. J., Michaels, S., Moon, J., Bell, T., Dyer, E., Edwards, K. D., McGill, T. A. W., Novak, M., & Park, A. (2017). Scaling up three-dimensional science learning through teacher-led study groups across a state. *Journal of Teacher Education*, *68*(3), 280-298. https://doi.org/10.1177/0022487117699598

Reiser, B. J., Novak, M., McGill, T. A. W., & Penuel, W. R. (2021). Storyline units: An instructional model to support coherence from the students' perspective. *Journal of Science Teacher Education*, *32*(7), 805-829. https://doi.org/10.1080/1046560X.2021.1884784

Quinn, H. (2021). Commentary: The role of curriculum resources in promoting effective and equitable science learning. *Journal of Science Teacher Education, 32*(7), 847-851. https://doi.org/10.1080/1046560X.2021.1897293

Schindel, A., & Tolbert, S. (2017). Critical caring for people and place. *The Journal of Environmental Education*, *48*(1), 26-34.

Segura, D., Morales-Doyle, D., Nelson, S., Levingston, A., & Canales, K. (2021). Sustaining community-school relationships around shared visions of climate justice and science teaching. *Connected Science Learning*, *3*(5). https://www.nsta.org/connected-science-learning/connected-science-learning-september-october-2021/sustaining-community

Semken, S., & Freeman, C. B. (2008). Sense of place in the practice and assessment of place-based science teaching. *Science Education*, *92*(6), 1042-1057. https://doi.org/10.1002/sce.20279

Semken, S., & García, Á. (2021). Synergizing standards-based and place-based science education. *Cultural Studies of Science Education*, *16*, 447-460. https://doi.org/10.1007/s11422-021-10020-4

Sengupta-Irving, T., Tunney, J., & Macias, M. (2021). Stories of garlic, butter, and ceviche: Racial-ideological micro-contestation and microaggressions in secondary STEM professional development. *Cognition and Instruction*, *39*(1), 65-84. https://doi.org/10.1080/07370008.2020.1812612

Shepard, L. A., Penuel, W. R., & Pellegrino, J. W. (2018). Using learning and motivation theories to coherently link formative assessment, grading practices, and large-scale assessment. *Educational Measurement: Issues and Practice*, *37*(1), 21-34. https://doi.org/10.1111/emip.12189

Sheth, M. (2019). Grappling with racism as foundational practice of science teaching. *Science Education*, *103*(1), 37-60. https://doi.org/10.1002/sce.21450

Singleton, C., Deverel-Rico, C., Penuel, W. R., Krumm, A. E., Allen, A.-R., & Pazera, C. (accepted, with major revisions). The role of equitable classroom cultures for supporting interest development in science. *Journal of Research in Science Teaching*.

Showalter, D., Hartman, S. L., Johnson, J., & Klein, B. (2019). *Why rural matters 2018-2019: The time is now*. Rural School and Community Trust. http://www.ruraledu.org/WhyRuralMatters.pdf

Sipple, J. W., & Brent, B. O. (2008). Challenges and opportunities associated with rural school settings. In H. F. Ladd & E. B. Fiske (Eds.), *Handbook of research in education finance and policy* (pp. 612-629). New York: Routledge.

Smith, G. A. (2007). Place-based education: Breaking through the constraining regularities of public school. *Environmental Education Research*, *13*(2), 189-207.

Styles, S. (2019). Literacies of land: Decolonizing narratives, storying, and literature. In L. T. Smith, E. Tuck, & K. W. Yang (Eds.), *Indigenous and decolonizing studies in education* (pp. 24-37). Routledge.

Suárez, E., & Bell, P. (2019). *Supporting expansive science learning through different classes of phenomena.* NARST Annual International Conference, Baltimore, MD.

Tai, R. H., Liu, C. Q., Maltese, A. V., & Fan, X. (2006). Planning early for careers in science. *Science*, *312*, 1143-1144. https://doi.org/10.1126/science.1128690

Taylor, C. S., Kokka, K., Darling-Hammond, L., Dieckmann, J., Pacheco, V. S., Sandler, S., & Bae, S. (2016). *Student engagement: A framework for on-demand assessment tasks*. Stanford Center for Opportunity Policy in Education.

Thier, M., Longhurst, J. M., Grant, P. D., & Hocking, J. E. (2021). Research deserts: A systematic mapping review of US rural education definitions and geographies. *Journal of Research in Rural Education*, *37*(2).

Trout, M. (2018). Embodying care: Igniting a critical turn in a teacher educator's relational practice. *Studying Teacher Education*, *14*(1), 39-55. https://doi.org/10.1080/17425964.2017.1404976

Tzou, C. T., & Bell, P. (2010). Micros and Me: Leveraging home and community practices in formal science instruction. In K. Gomez, L. Lyons, & J. Radinsky (Eds.), *Proceedings of the 9th International Conference of the Learning Sciences* (pp. 1135-1143). International Society of the Learning Sciences.

Valencia, R. R. (2010). *Dismantling contemporary deficit thinking: Educational thought and practice*. Routledge.

Valenzuela, A. (1999). *Subtractive schooling: U.S.-Mexican youth and the politics of caring*. SUNY Press.

Van Horne, K., Penuel, W. R., & Bell, P. (2016). Integrating Science Practices Into Assessment Tasks. STEM Teaching Tools: Teaching Tools for Science, Technology, Engineering and Math (STEM) Education. http://stemteachingtools.org/brief/30

Vélez-lbáñez, C. G. (1988). Networks of exchange among Mexicans in the U.S. and Mexico: Local level mediating responses to national and international transformations. *Urban Anthropology*, *17*(1), 27-51. https://www.jstor.org/stable/40553124

Vélez-lbáñez, C. G., & Greenberg, J. (1992). Formation and transformation of funds of knowledge among U.S. Mexican households. *Anthropology & Education Quarterly*, *23*(4), 313-335. https://doi.org/10.1525/aeq.1992.23.4.05x1582v

Verdín, D., Smith, J. M., & Lucena, J. (2021). Funds of knowledge as pre-college experiences that promote minoritized students' interest, self-efficacy beliefs, and choice of majoring in engineering. *Journal of Pre-College Engineering Education Research*, *11*(1), 192-213.

Walkington, C. (2013). Using adaptive learning technologies to personalize instruction to student interests: The impact of relevant contexts on performance and learning outcomes. *Journal of Educational Psychology*, *105*(4), 932-945. https://doi.org/10.1037/a0031882

Wanders, F. H. K., Van Der Veen, I., Dijkstra, A. B., & Maslowski, R. (2020). The influence of teacher-student and student-student relationships on societal involvement in Dutch primary and secondary schools. *Theory & Research in Social Education*, *48*(1), 101-119. https://doi.org/10.1080/00933104.2019.1651682

Warren, B., Vossoughi, S., Rosebery, A., Bang, M., & Taylor, E. (2020). Multiple ways of knowing: Re-imagining disciplinary learning. In N. i. S. Nasir, C. D. Lee, R. Pea, & M. McKinney de Royston (Eds.), *Handbook of the Cultural Foundations of Learning* (pp. 277-293). Routledge.

Wentzel, K. R. (1997). Student motivation in middle school: The role of perceived pedagogical caring. *Journal of Educational Psychology*, *89*(3), 411-419. https://doi.org/10.1037/0022-0663.89.3.411

Wingert, K., Jacobs, J. K., Lindsay, W., Lo, A. S., Herrman-Abell, C. F., & Penuel, W. R. (2022). Understanding the priorities and practices of rural science teachers: Implications for designing professional learning. *The Rural Educator*, *43*(3), 26-40. https://doi.org/10.55533/2643-9662.1338

Wingert, K. & Penuel, W. R. (2019). Summary of Colorado science teachers' vision for science education. Unpublished manuscript. Boulder, CO.

Wubbels, T., & Brekelmans, M. (2005). Two decades of research on teacher–student relationships in class. *International Journal of Educational Research*, *43*, 6-24. https://doi.org/10.1016/j.ijer.2006.03.003

Yin, R. K. (2009). *Case study research: Design and methods* (Vol. 5). Sage.
Zimmerman, H. T., & Weible, J. L. (2017). Learning in and about rural places: Connections and tensions between students' everyday experiences and environmental quality issues in their community. *Cultural Studies of Science Education*, *12*(1), 7-31. https://doi.org/10.1007/s11422-016-9757-1
Zinger, D., Sandholtz, J. H., & Ringstaff, C. (2020). Teaching science in rural elementary schools: Affordances and constraints in the age of NGSS. *The Rural Educator, 41*(2), 14-30. https://doi.org/10.35608/ruraled.v41i2.55