Balancing Standards Alignment with Educator Needs

by Craig Kohn, Abby Helmke, and Joe Hendricks - Waterford Union High School

Subject & Problem

Despite nearly a decade since the release of NGSS, there is little evidence of large-scale changes in US science instructional practices (Banilower *et al.*, 2018; ACT, 2016, 2020). Cohen and Mehta (2017) argue that while standards-based reforms have been effective for changing the rhetoric of educational goals, these efforts have had very limited impact on system-wide student outcomes. Local norms, values, and expectations have significant influence on teacher decisions but are often in tension with three-dimensional approaches that privilege open-ended sensemaking and reasoning in student-driven classroom conditions (Lin *et al.,* 2021).

This paper presents findings from a unique research-practice partnership. The authors of this paper are biology instructors employed in a rural Midwestern high school. We report our experiences and insights gained from developing standards aligned curriculum that is responsive to our local needs. We argue that for NGSS to be successful, implementation must address both the interests of students and the needs of educators. We address two research questions in this paper: 1. What approaches enable NGSS alignment in a manner that is responsive to the needs of local educators? 2. To what extent were these efforts perceived as successful and why?

Theoretical Framework

Our work is informed by design-based research (Cobb *et al.,* 2003; Collins *et al.,* 2004). DBR aims to "blend empirical educational research with theory-driven design of learning environments ... [to understand] ... how, when and why educational innovations work in practice" (Design-Based Research Collective, 2003, p. 5). DBR is valuable for investigating theories in education while generating practical solutions (Abdallah and Wegerif, 2014). Given research and standards are enacted within classrooms, DBR provides a means to incorporate the perspectives and insights of teachers to enable broader adoption of reforms (Farrell *et al.*, 2022).

Design & Procedure

Methods. This bounded case study was conducted at a rural high school enrolling nearly 1000 students in a midwestern state. During the 2021-22 and 2022-23 school years, we worked collaboratively to create our own NGSS-aligned biology curriculum that would be sufficiently responsive to localized constraints. The first author primarily designed the curricular materials through a district position that was partly teaching and partly curriculum development using insights gained from professional experiences in education research. Our goal was to use these materials to prepare 9th grade students for the risk and ambiguity of socioscientific decisions made in the adult world while being responsive to factors such as grading, assessment, engagement, time, and expense. In this time, our efforts have expanded to include curriculum development for all core science classes. In this paper, we focus on the outcomes specific to students in the biology courses.

Curriculum Development: In June of 2021, the first author met with the biology team and district administrators to propose a new curriculum that initially followed the scope and sequence of *Carbon TIME* materials (Covitt & Anderson, 2018). Over the following weeks, we held team meetings to determine our curricular objectives, identify constraints, and develop initial drafts. We continued to iteratively develop the curricular materials over the remainder of the school year and met weekly to discuss changes and determine solutions.

Our resulting curriculum consisted of eight curricular units. Each monthly unit consisted of 3-4 physical packets that guided instruction. Each packet consisted of five parts: 1) an introductory phenomenon-based data dive; 2) core ideas and revision of initial explanations; 3) phenomenon-based investigation; 4) review & assessments; and 5) life connections. Each packet of content required 4-6 class periods (60 minutes) and was guided by scaffolded prompts to support collaborative discourse and sensemaking.

Data Collection: We addressed our research questions primarily using semi-structured interviews with participating staff members as well as an administrator. All interviews occurred after the completion of the fall semester in both school years. The interview questions addressed experiences with the curriculum, its efficacy, modes of implementation, and challenges and affordances. Analysis primarily consisted of identifying common themes across participant responses during the exit interview. These conclusions were confirmed with participants and were triangulated with insights obtained from weekly meeting recordings, field notes from the first author, and curricular artifacts.

In addition, we analyzed student responses to questions in packets and assessments to determine the effectiveness of the curricular materials for achieving three-dimensional science learning outcomes. A total of 63 students were randomly selected from each of our general biology courses during the 2022 fall semester. These responses were coded using learning progression categories adapted from the *Carbon TIME* research project (Lin et al., 2022). Discrepancies in individual coding by the first two authors were resolved through discussion to achieve full agreement.

Findings & Analysis

RQ1: Balancing NGSS alignment & local needs via curricular design. Our overarching goal was to develop and implement NGSS-aligned curricular materials that were responsive to our local needs. We define "NGSS alignment" as implementing curriculum and instruction that enables three-dimensional student outcomes that correspond to NGSS life science performance expectations. We prioritized iterative student discourse that enables increasingly sophisticated sensemaking and reasoning about phenomena. Our primary objective was to improve the science literacy of our students to enable them to act as "competent outsiders" who can understand and respond to established science as well as new evidence as it emerges (Feinstein, 2013).

Numerous research-based curricular options already exist for biology, including *Carbon TIME* (Covitt & Anderson, 2018), *Ambitious Science Teaching* (Stroupe, 2017; Windschitl et al., 2020), *Storylines* (Reiser et al., 2021), and *OpenSciEd* (Edelson et al., 2021). However, we felt these resources would be insufficiently responsive to our local needs, particularly with a wide range of instructor experiences with three-dimensional science learning, the ongoing COVID-19 pandemic, and numerous district initiatives. These circumstances necessitated quick and efficient assessments, remote learning compatibility, low risk and ambiguity, and manageable time commitments. In the following paragraphs, we summarize the design considerations from our materials that were particularly valuable for balancing standards alignment with local needs. These considerations include 1) discourse guides, 2) scaffolding within packets, and 3) grading & assessment.

Discourse Guides: While productive student discourse emerged as a priority, our instructors felt uncomfortable implementing this aspect without explicit guidance. The shift from traditional IRE (initiate-response-evaluate) styles of instruction to academically productive discourse (Michaels & O'Connor, 2012) presented a significant obstacle. To reduce the risk and ambiguity of implementation, we embedded instructions with suggestions for implementation into our materials to guide student discourse. This provided a means to communicate how to successfully implement these materials without compromising teacher autonomy and professionalism. This aspect was enhanced through the use of a 5-part packet structure and flexible but consistent discourse routines within each aspect. We subsequently discussed differences in implementation for each part of the packet during weekly meetings to identify a range of effective options for achieving three-dimensional outcomes.

We had particular success with having students initially work in small groups to critique competing claims about data specific to weekly anchoring phenomena and with prompting students to daily revise their explanations about phenomena. Despite initial concerns that it was "a little monotonous", these aspects enhanced instructor confidence without compromising student engagement. The level of scaffolding in our materials provided an additional benefit of being sufficiently responsive to extensive absences due to factors like COVID-19. Students could use online classroom resources to complete the scaffolded prompts independently during absences, minimizing their impact on student performance in comparison.

Packets & Scaffolded Interactions: Another key challenge was shifting from rote memorization of decontextualized facts to evidence-based reasoning about anchoring phenomena. Our instructors felt uncertain how to effectively choose an anchoring phenomenon to support sensemaking, and how to productively scaffold student interactions to enable three-dimensional outcomes. The 5-part packet template proved to be particularly helpful for this purpose.

To initiate reasoning and sensemaking, students were first prompted to take a stance on three competing claims about a phenomenon at the start of each packet. We observed that students were often surprised at the range of ideas among their peers, and became more engaged as they were prompted to achieve consensus in small groups settings. Students were then provided with data and asked to use this to bolster or reject their claims about the phenomenon (similar to a *claims-evidence-reasoning* approach). We then asked students to develop preliminary explanations and identify similarities and discrepancies across each group's arguments. The inherent structure of this routine reduced the burden on the instructor and provided a roadmap for initiating productive reasoning about the anchoring phenomenon.

In Part 2 of the packet, we provided students with "core ideas" about the anchoring phenomenon. Students then took on a variety of roles within their groups (info seeker, summarizer, writer, and speaker) to develop explanations about phenomena for the rest of the class. Student engagement improved by having students first decide which of the four roles they would assume as this served as a low-risk discussion starter. This scaffolding also prepared students to more effectively revise and improve their explanations from the previous day.

We then had students assess the predictive power of their explanatory models in Part 3 through investigations that prioritized testing hypotheses, data collection, and analysis. The choice of investigation provided a means to connect the anchoring phenomenon to a real-world context. The packet's prompts and scaffolding also implicitly demonstrated the difference between traditional labs (with rote instructions and common universal outcomes) and investigations that foregrounded sensemaking and reasoning.

After revising explanations using data and insights from Part 4, we concluded by connecting the weekly phenomenon to contexts relevant to their local communities in Part 5. This provided guidance for connecting student learning to local contexts and lived experiences. The use of this 5-part structure provided a well-defined but flexible 'roadmap' for instructors to address phenomena in a predictable manner that privileged reasoning and sensemaking in lieu of traditional rote memorization and recitation.

Assessments: Another salient priority were assessments that were fair, consistent, efficient, and that enabled intervention. This is reflective of previous literature that argues that grading plays a predominant role in classroom decision making (Doyle, 1983; Schneider & Hutt, 2014). We initially utilized ungraded formative group assessments as a means to assess weekly progress. However, our students perceived an absence of a grade as signaling limited importance and value, impairing their engagement and effort. We also attempted to use laboratory investigations as means to communicate student progress. However, our emphasis on ongoing revisions for sensemaking made this infeasible - students were reluctant to express their ideas outside of "risk free" conditions.

In response, we added a graded "mastery check" to each weekly packet. This consisted of a mixture of one-dimensional forced choice responses and a three-dimensional diagnostic question. Students who scored below proficient were requested for an intervention session outside of regular class time. This provided a sufficiently quick and effective means for identifying students who needed extra support. The mastery checks acted as 'checkpoints' to assess student readiness for a summative assessment that primarily consisted of three-dimensional performances.

We also added "pre-investigation questions" to the packets which required students to demonstrate proficiency through verbal responses before they could advance to the investigations in Part 3 of each packet. Instructors could provide immediate feedback to identify gaps in comprehension and offer reinforcement where needed. Each of these components provided quick but meaningful options for determining appropriate pacing for implementing the curriculum without onerous grading.

Grading: Our team had early concerns about providing impartial and consistent grading for open-ended responses to three-dimensional prompts. In response, we developed a 3-point grading scale based on completeness, accuracy, and precision of students' explanations and solutions. This provided flexibility to provide partial credit in a manner that was responsive to factors such as effort and progress. We included an evaluation box next to each prompt with checkmarks for these criteria that allowed us to provide quick feedback for each of these three considerations. Space was also provided below each question for more extensive comments if needed. We also independently scored sample responses and then calibrated our scoring through discussion during our weekly meetings. Our end-of-unit packets guided students in using this rubric to critique both sample responses and their previous work. This helped clarify our grading policies and expectations, strengthened buy-in among students and parents, and enhanced their preparation to provide more robust three-dimensional performances.

RQ2: Were these efforts perceived as successful and why? By the end of the first semester of implementation, it was evident in our first exit interview that our materials had become effective for our needs. We also agreed that our materials struck a balance between our goal of NGSS-alignment as well our local needs as instructors. As our third author argued, "I was the most skeptical going in…but it's definitely what I feel we should be doing for NGSS and the kids." Here we will discuss aspects of our work that most contributed to this outcome. These include: 1) ease of implementation, 2) responsiveness to district expectations, 3) student engagement, 4) and student outcomes.

Ease of Implementation: The design of our weekly packets reduced the ambiguity of lesson planning, minimized grading, and provided a flexible roadmap for each instructor with a default option that could be changed. As such, our team perceived these materials as easier to teach than both our prior one-dimensional curriculum as well as other research-based options. As our second author explained, "As a first-year person, I can see how the way it's set up really does allow you to follow that process. It's scaffolded, the steps are sequenced." Our third author agreed, stating that "a book never gives you that," adding, "This is laid out for you. You get to see it ahead of time and what you were getting into." While the extensive scaffolding of our curriculum provided less affordances with regard to student agency, it enhanced instructional feasibility, resulting in what the third author called the "perfect middle point". As the second author argued: "Our discussions in our investigations will drive what we do the next day. It's not like it's predetermined."

While we were "grading way less" compared to previous years, we also felt we were providing more meaningful feedback more regularly to our students. Through ungraded formative assessments and continually revising explanations, we had a "risk-free" means to assess student reasoning and sensemaking. The explicit prompting of regular verbal feedback also provided a means to enhance student outcomes. Our grading scale provided a uniform set of criteria for proficiency in response to open-ended three-dimensional prompts, and our weekly mastery checks enabled us to more easily identify students in need of intervention without negatively affecting their grades. As our second author explained, we now had a feasible means to "call that kid in that did horrible and see that they got better."

Responsiveness to District Expectations: Meeting district expectations emerged as a critical concern for our instructors as this significantly affected administrative perceptions of job performance. Numerous district initiatives necessitated iterative flexibility that generally is not provided by either traditional textbook-based approaches or more progressive research-based options. Furthermore, we used our weekly meetings as an opportunity to collaborate with administrators to adapt the curricular design and reduce the workload of individual instructors. We often referred to this as an "umbrella approach" to curriculum; i.e., our materials ensured our teachers wouldn't also have to worry about individually meeting district initiatives in addition to their regular teaching obligations. This was particularly helpful in our first year due to especially challenging circumstances from COVID-19.

Student Engagement: Compared to our previous materials, we felt that our new materials were far more capable of enabling student engagement in a manner consistent with the expectations of NGSS. As our second author explained, "They hands-down preferred the approach we're doing to [the former] approach." Getting students to engage in discussions was like "pulling teeth" with previous materials. The new curriculum often resulted in productive "non-stop talking" about science phenomena during classroom discourse.

Our conclusions align with Engle and Conant's (2002) criteria for student engagement, such as the capacity to develop deeper questions, identify connections between concepts, and design solutions. For example, the third author observed that "in the past…it was just collect the data and answer the questions." With these materials, he felt that students "actually care what somebody else might say within a group because it's leading them to understand what they want to know."

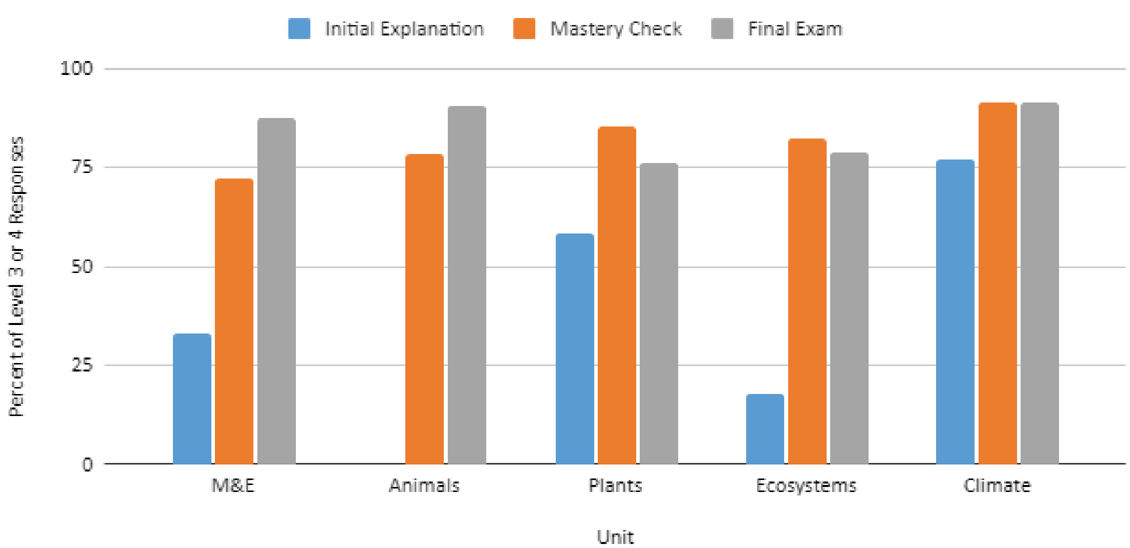
We also observed that the "risk-free" forms of assessment inherent in our instructional model were important for supporting student engagement. This was particularly evident when students could explicitly see how these opportunities supported their preparation for the graded mastery checks near the end of each packet. Design features such ongoing revisions to ideas and the verbal pre-investigation questions increased student engagement while deepening their capacity for reasoning and sensemaking.

We also observed that student perceptions of science learning had become more positive in response to the new curriculum. Our second author argued that in the past, students had regularly questioned the value of their biology lessons. In contrast, she explained that with the new materials, “I didn't get any backlash of what we were learning. Nobody ever was like, ‘This is so stupid. Why would we ever need to know this?’” She partly attributed this outcome to the increased emphasis on student agency: “there's a difference between their peers coming to a decision on how science works and us being adults just telling them.” The third author agreed, adding: “The students are able to have a say in it, not just regurgitate…They're able to talk.”

Perceived Outcomes: Compared to previous years, we perceived our students to be more capable of providing sophisticated explanations and solutions in response to our new curriculum. As our third author explained, “When you look at what we did on the final exam, we could have never asked kids to answer a question like that last year.” He felt that these materials enabled our students to be “better sense makers, link makers, problem solvers, graph readers, etc.” The second author observed a shift in an emphasis on getting a ‘right answer’ to engaging in evidence-based argumentation as they reasoned about phenomena: “Instead of them saying, this is the answer because [the teacher] said so...they're saying how they know, what they know, why they know the answer is correct.”

While it seemed to us that our student outcomes reflected the expectations of NGSS, our approach to grading and assessment were not sufficient to address this question. To assess this consideration more systematically, the first and second authors independently analyzed written performances throughout the fall semester of the second year from 63 randomly selected focus students. We used learning progression categories adapted from the *Carbon TIME* research project (Lin et al., 2022) to code responses as:

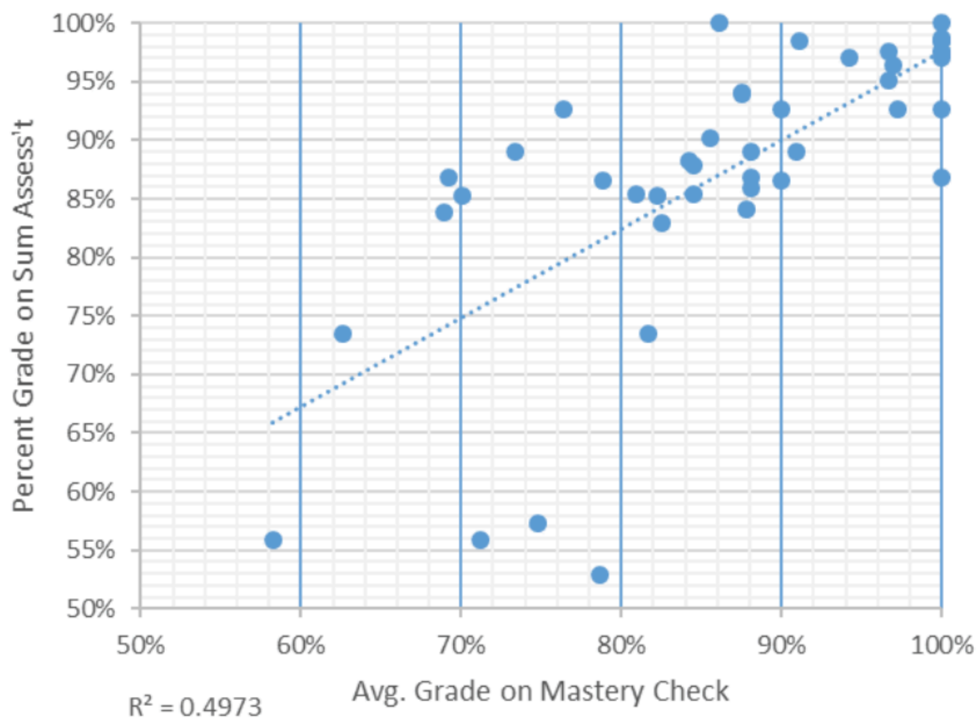
- Level 4: Students fully trace matter and energy through carbon-transforming processes at multiple scales in space and time (equivalent to NGSS performance expectations).
- Level 3: Students show explicit awareness of important scientific principles and of accurate conceptual models but their responses were incomplete or were partly confused.
- Levels 1 and 2: Students' explanations and arguments contain extensive errors and/or consisted of force-dynamic arguments in lieu of scientifically accurate accounts.



We then identified discrepancies in our application of codes and reached full agreement through discussion. Based on this analysis, we found that most students in this class (over 75% for every question analyzed) achieved level 3 or level 4 performances on their final exam; roughly a third (33.2%) of our overall responses were a level 4 performance. Furthermore, we can see ongoing progression from the first response in the first packet of each unit to the first mastery check to the final exam. Student performances in the unit on climate change are particularly noteworthy given the initial level of skepticism we observed among our student population in what is generally a rural and conservative district. While this analysis lacks a control and is not sufficient for making strong claims about the effectiveness of our materials, it does provide preliminary evidence that our materials could be effective for achieving the goals of NGSS.

Additionally, the graded weekly mastery checks provided a means to improve student performances while reducing achievement gaps. By automatically enrolling students in intervention sessions if they were near or below passing, we were able to identify gaps in performance and provide additional support where needed. We were often surprised by the extent to which students expressed gratitude for these interventions (which often occurred in lieu of informal socialization during lunch). Students frequently viewed these interventions as needed support, whereas in the past these were more likely to be perceived as punishment.

Our data also suggest that the use of mastery checks and interventions reduced achievement gaps. With some notable exceptions due to factors such as testing anxiety, students with the lowest initial mastery check scores were more likely to improve on their summative assessments scores compared to students with higher initial mastery check performances.



Administrator Perspectives: In addition to exit interviews with instructors, the first author also conducted a semi-structured interview with a school administrator who oversaw curriculum and instruction among other obligations. He was “really pleased” with the outcomes of the curriculum design and felt that “our students are learning not just at a good level, they're learning at an exceptional level.” He based this conclusion on multiple factors, including:

- Persistence of student outcomes: he observed that students were still capable of effectively addressing questions weeks and months after they initially learned the content (“We have kids that are understanding concepts from quarter one, that in the past they would've forgot by the end of week three.”)
- Staff engagement: he observed that science department staff were “energized” by this work, adding that “the teachers are really excited about it and they're talking about it.”
- Student engagement: he felt that students were more visibly engaged during his classroom observations and that they had more robust interactions with each other through real-world phenomena that enabled improvements to critical thinking and evidence-based argumentation.
- Alignment with district goals: recently the school district shifted to a PLC model that emphasized staff collaboration, student assessment, and intervention to close achievement gaps. This administrator felt that these materials supported collaboration to determine “what we want students to know and how we're gonna assess it” while also providing options for effective intervention and identification of performance gaps.

This administrator echoed sentiments from our second and third author by stating that he observed palpable “apprehension” initially, but firsthand experiences led to eventual acceptance and buy-in among department staff. He felt that the district’s shift towards scheduling extended time for departments to meet and plan was critical for the success of these efforts. He had explicit doubts that these efforts would have been as successful without this dedicated time for collaboration.

Drawbacks & Concerns: Numerous concerns emerged as we developed our materials. First, our students sometimes struggled to grasp that their data dives utilized “real-life data that actually happened.” As our second author explained, “[Some] assume that [this data] is something that we came up with because we want them to think this way…especially when it came to CO_2.” We now better recognize the value of replicating the work of scientists as a means for enabling students to grasp the ‘realness’ of data. Through an additional packet template that emphasized planning and conducting investigations, we have expanded the opportunities for students to engage in more authentic open-ended inquiry. This resulted in more buy-in among students in our second year of implementation without compromising the feasibility of implementation.

Second, we struggled with “quiet kids that don't do very well” in collaborative settings and saw the impacts this had on their classroom performances. We increasingly perceive that there are limits to what an instructor alone can achieve through classroom instruction. We are exploring additional opportunities for more communication and engagement with parents and guardians to address factors outside the classroom that affect student performance.

Finally, we concede that past members of our department might have undermined or even disrupted these efforts. As our second author argued, “if this change would be made in [another] school, administrator support would be needed.” We acknowledge that most districts would struggle to overcome long-standing norms and traditions, particularly among the most veteran members in their science departments. This sentiment was echoed in the administrator interview:

"…you have a new person coming to your team once every three or five years…and that person kind of conforms to what's [already] being done."

The administrator interview also yielded insights into other limitations in this work. In particular, our administrator observed the need for training and planning time to ensure that these materials were properly implemented. He expressed concern that an untrained instructor could assume that "just because I hand this [packet] out now, kids are gonna get it." In particular, he observed a wide range in the extent to which different instructors throughout the department utilized discourse to support reasoning and sensemaking.

He also underscored the need for departments to have access to qualified individuals who could provide ongoing support and training to their respective teams, arguing "they would have to definitely get somebody to coach them and walk through that process to do it well." This suggests that there are limits to what a curriculum alone can achieve regarding three-dimensional learning. Successful implementation likely depends on having qualified individuals to navigate the tension of balancing standards alignment with local needs. This likely exceeds the capacity of most individual schools and would require coordination with post-secondary institutions and government entities to ensure districts have equitable access to training opportunities. This aligns with recent findings (such as Morrison Thomas *et al.*, in press) that argue that teachers depend on ongoing support for enabling three-dimensional student outcomes in addition to their goals of meaningful student engagement through curricular implementation.

Alternative Interpretations, Bias, Reliability, and Validity

It is inevitable that there would be some bias in favor of the effectiveness of our materials and the performances of our own students. We also acknowledge significant limitations in our abilities to make any claims about the effectiveness of our work or its applicability to other classrooms. We are not making any broad assertions about three-dimensional learning, nor are we proposing that each school would develop their own curriculum. Rather, we hope to draw attention to the factors and considerations that supported what we perceived to be successful outcomes in our contexts. We offer this paper as an opportunity to reflect on insights gained from creating a "proof of concept" curriculum that balances NGSS alignment with local considerations. Given the importance of teachers for enabling system-wide reforms in education, we believe that our insights provide an important perspective on the affordances and limitations of NGSS implementation within individual classrooms.

Contributions and General Interest

Achieving the goals of NGSS requires balancing calls for radical reform in science teaching with the ongoing needs of schools and instructors. While the scope of this paper is limited to the experiences and outcomes in one case study, our findings provide insights on the challenge of achieving both standards alignment and addressing local needs. Our work suggests that the goals of NGSS can be achieved in a manner that is responsive to the real-world constraints in the day-to-day realities of teaching. However, for these efforts to be successful, teachers depend on access to unambiguous but flexible curriculum, effective training opportunities, and regularly scheduled collaboration. Support from government and post-secondary institutions are likely a critical component for enabling these outcomes on a larger basis.

References

Abdallah, M. M. S., & Wegerif, R. B. (2014). *Design-Based Research (DBR) in Educational Enquiry and Technological Studies: A Version for PhD Students Targeting the Integration of New Technologies and Literacies into Educational Contexts*. Retrieved from https://eric.ed.gov/?id=ED546471

ACT. (2016). The condition of college and career readiness 2015. https://www.act.org/content/dam/act/unsecured/documents/Condition-of-College-and-CareerReadiness-Report-2015-United-States.pdf

ACT. (2020). The condition of college and career readiness 2019. https://www.act.org/content/dam/act/unsecured/documents/National-CCCR-2019.pdf.

Banilower, E. R., Smith, P. S., Malzahn, K. A., Plumley, C. L., Gordon, E. M. & Hayes, M. L. (2018). Report of the 2018 NSSME+. Horizon Research, Inc. http://horizon-research.com/NSSME/2018-nssme/research-products/reports/technical-report

Cobb, P., Confrey, J., DiSessa, A., Lehrer, R., & Schauble, L. (2003). Design experiments in educational research. *Educational Researcher*, 32(1), 9–13.

Cohen, D. K., & Mehta, J. D. (2017). Why reform sometimes succeeds: Understanding the conditions that pro- duce reforms that last. *American Educational Research Journal, 54*(4), 644–690. https://doi.org/10.3102/ 0002831217700078

Collins, A., Joseph, D., & Bielaczyc, K. (2004). Design research: Theoretical and methodological issues. *Journal of the Learning Sciences, 13*(1), 15–42. https://doi.org/10.1207/s15327809jls1301_2

Covitt, B., & Anderson, C. (2018). Assessing scientific genres of argument, explanation and prediction. In *Language, Literacy And Learning In The STEM Disciplines* (pp. 206–230). Routledge.

Design-Based Research Collective. (2003). Design-based research: An emerging paradigm for educational inquiry. *Educational Researcher, 32*(1), 5–8.

Doyle, W. (1983). Academic work. *Review of Educational Research, 53*(2), 159-199.

Edelson, D. C., Reiser, B. J., McNeill, K. L., Mohan, A., Novak, M., Mohan, L., ... & Suárez, E. (2021). Developing research-based instructional materials to support large-scale transformation of science teaching and learning: The approach of the OpenSciEd middle school program. *Journal of Science Teacher Education, 32*(7), 780-804.

Engle, R. A., & Conant, F. R. (2002). Guiding principles for fostering productive disciplinary engagement: Explaining an emergent argument in a community of learners classroom. *Cognition and Instruction, 20*(4), 399-483.

Farrell, C. C., Penuel, W. R., Allen, A., Anderson, E. R., Bohannon, A. X., Coburn, C. E., & Brown, S. L. (2022). Learning at the Boundaries of Research and Practice: A Framework for Understanding Research–Practice Partnerships. *Educational Researcher*, 0013189X211069073.

Lin, Q., Frank, K. A., Bathia, S., Draney, K., Thomas, J., & Anderson, C. W. (2022). Factors affecting students' learning from a design‐based implementation research project in diverse education systems. *Journal of Research in Science Teaching, 59*(5), 808-840.

Michaels, S., & O'Connor, C. (2012). Talk science primer. *Cambridge, MA: TERC*.

Reiser, B. J., Novak, M., McGill, T. A., & 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.

Schneider, J., & Hutt, E. (2014). Making the grade: A history of the A–F marking scheme. *Journal of Curriculum Studies, 46*(2), 201-224.

Stroupe, D. (2017). Ambitious teachers' design and use of classrooms as a place of science. *Science Education, 101*(3), 458-485.

Windschitl, M., Thompson, J., & Braaten, M. (2020). *Ambitious science teaching*. Harvard Education Press.