# Balancing Standards Alignment with Educator Needs

## 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 school year, we worked collaboratively to create our own NGSS-aligned biology curriculum that would be sufficiently responsive to our needs. The first author primarily designed the curricular materials through a district position that was 50% teaching and 50% curriculum development using insights gained from professional experiences in education research. Our goal was to use these materials to prepare 9$^{th}$ 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.

*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 'weeks', each with 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 'week' of content required 4-6 class periods (60 minutes) and was guided by a physical packet with scaffolded prompts to support collaborative discourse and sensemaking.

*Data Collection*: We addressed our research questions using weekly meeting recordings, field notes, curricular artifacts, and responses to semi-structured exit interview questions. The interview questions addressed our 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. Conclusions drawn from these responses were triangulated with the remaining teacher data and were confirmed with all participants.

## 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.

*Scaffolded Discourse & Iterative Revisions:* 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. We had particular success with having students work in small groups to critique competing claims about data specific to weekly anchoring phenomena and prompting students to daily revise their explanations about phenomena. Additionally, we used a repeating weekly structure to minimize ambiguity. Despite initial concerns that it was "a little monotonous", it enhanced instructor confidence without compromising student engagement. While these packets provided more scaffolding than most other NGSS-aligned options, it centered student ideas as a driving factor. This approach provided an additional benefit of being sufficiently responsive to extensive absences due to COVID-19. Students could use online classroom recordings to complete the packets independently, reducing the impact of absences.

*Data-based Phenomena:* Another key challenge was shifting from an emphasis on memorization of decontextualized facts to evidence-based reasoning about anchoring phenomena. Each of our weekly packets used visual data and competing claims to initiate student discourse and sensemaking about a weekly phenomenon. We then asked students to develop preliminary explanations and identify discrepancies across each group. On the following day, we provided students with "core ideas" and asked them to revise and improve their explanations. We then had students assess the predictive power of their explanatory models through investigations that prioritized testing hypotheses, data collection, and analysis. This was followed by another round of revisions to explanatory models. After review and assessments, we concluded by connecting the weekly phenomenon to contexts relevant to their local communities. This approach provided a well-defined but flexible 'roadmap' for instructors to address phenomena in a predictable manner that privileged reasoning and sensemaking.

*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). In response to district initiatives, we initially utilized ungraded formative group assessments, but our students perceived an absence of a grade as signaling limited importance and value. We also intended to use laboratory investigations to communicate student progress. However, an emphasis on ongoing revisions for sensemaking made this infeasible, resulting in initial inaccurate signals about student performances. 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. This provided a sufficiently quick and effective means for identifying students who needed extra support. In contrast, our summative assessments mostly entailed three-dimensional written prompts. These were largely based on the content of *Carbon TIME* materials and modified to be more responsive to our content. Students also completed ungraded three-dimensional "check in" questions prior to the exam and were provided with additional reinforcement opportunities if they scored below proficient.

*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 that allowed us to provide quick feedback for each of these three considerations, and independently scored sample responses to calibrate our scoring. Our end-of-unit packets also 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 our exit interview at the end of the first semester, we felt our approach was successful for our needs. 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.

*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 limited 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. 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. This was particularly helpful in an especially challenging year of teaching due to 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." 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 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 student performances from the fall final exam from one class (*n = 26*). We used learning progression categories 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 awareness of important scientific principles and of models but have difficulty connecting across scales and applying principles consistently.
- Levels 1 and 2: Students' explanations and arguments focus on actors (e.g., animals, plants, people) and enablers (e.g., food, water, sunlight).

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 80% for every question analyzed) achieved level 3 or level 4 performances on their final exam; over half (55%) of our overall responses were a level 4 performance. 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.

*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 appreciate the value of replicating the work of scientists to enable students to grasp the 'realness' of data. 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. 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."

## 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. The findings of this paper provide insights on both standards alignment and meeting local needs, which will be of significant interest to NARST members in Strand 4 (Science Teaching - Middle & High School), where we are submitting this proposal. This is also relevant to Strand 8 (In-service Science Teacher Education) and Strand 10 (Curriculum, Evaluation, & Assessment).

**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.