Collaborating with Teachers to Generate ML-Based Feedback: Contextualizing and Developing Meaningful and Relevant Feedback

Selin Akgün, Kevin Haudek, Lora Kaldaras, Joseph Krajcik

Problem / Objective:

Next Generation Science Standards (NGSS) highlights the importance of supporting students in explaining natural phenomena, solving real-life problems, and building and applying knowledge in K-12 science learning settings (NGSS Lead States, 2013). Accordingly, the nature of student learning is reflected in and aligns with the idea of a learning progression (LP), since it represents a developmental view of science learning. LPs are also aids to design and use assessments in science instruction and can act as an instructional tool and provide guidance on what supports students need to reach higher levels of understanding. Therefore, our overarching goal in this project is to develop an AI-based feedback system designed for open- ended formative assessments in high-school science based on a validated, NGSS-aligned LP. This specific paper focuses on one of the sub-goals of the project: designing formative feedback by collaborating with teachers on student work as reflected in student progression along with the levels of previously developed and validated NGSS-aligned LP. Therefore, this paper addresses the questions of:

a) what principles we use to create personalized feedback statements on students' performance aligning with their LP levels on a particular assessment item,

b) how we collaborate with teachers to get their constructive feedback to revise and make more compelling and relevant feedback statements to be delivered by AI (aligning with our principles) for the students' work.

Considering the process of doing and learning science is multi-modal as students use verbal and non-verbal forms of expressions through drawings, modeling, writings and other modalities to explain phenomena, we build on a validated NGSS-aligned multi-modal LP reflecting diverse ways of modeling and explaining electrostatic phenomena (Kaldaras et al., 2021). More specifically, we focus on generating feedback statements considering students' scientific models and explanations to justify their understanding on Coulomb's law and charge transfer. Then, we collaborated with high school science teachers to get their perspective on proposed feedback statements to make the statements more relevant, meaningful and applicable for students.

Theoretical Underpinnings:

In generating feedback statements for students' electroscope models, we extracted several critical principles that guide us to create meaningful, culturally relevant, and quality feedback. These principles align with the theoretical and empirical underpinnings of feedback generation from existing literature (e.g., Hattie & Timperley, 2007; Lyon, 2023; Nicol & Macfarlane-Dick, 2007; Penuel & Shepard, 2016; Qadir et al., 2020). We adopt seven core principles in generating feedback statements for students' models and explanations. Based on these principles, we argue that feedback should be a) constructive and compatible with students' prior knowledge, b) comprehensible and interpretable for students, c) actionable and specific, d) aligning with learning goals and LP, e) encouraging of student reflection, self-adjustment and improvement, f) encouraging of students' motivational beliefs and self-esteem, and g) responsive to students' cultural, linguistic and academic resources.

In addition to adopting these principles, our partnership and collaboration with the high school science teachers were another the key aspect of the quality and relevant feedback generation process. As teachers reviewed students' models and justifications to an assessment item about charge transfer in an electroscope, they shared their interpretations of the model, to what extent students might grasp the phenomena of Coulomb's law and charge transfer, what ideas students might be struggling with and what potential feedback they would provide for students' responses. Teachers also carefully reviewed the feedback statements we created for a range of typical student models and shared their suggestions to make it more relevant for students considering the seven principles we center in our work. Teachers' participation and involvement as critical experts also helped us to support culturally, linguistically and academically diverse students in building their models and providing justifications for them which was essential for working towards equitable science assessment (Grapin & Lee, 2022).

Methods:

Building on the process and NGSS-aligned LP outlined in Paper 1 in this set, we developed feedback statements to support students in revising their models and justifications on the electroscope item (see Figure 1) as part of the Interactions physical sciences curriculum. To capture the 3D nature of the electroscope item, we evaluate student responses based on their ability to develop a causal model explaining the difference between scenario A and B using Coulomb's law and charge transfer.

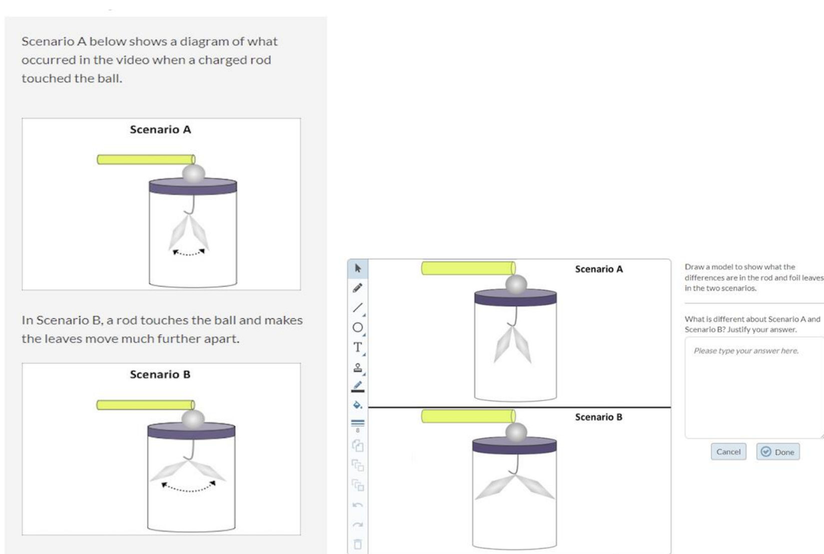

Figure 1. *The electroscope assessment item that we shared with teachers during the interview*

In our feedback generation process for students' models in Scenario A and B, we pay specific attention to reflect on the seven core principles into personalized feedback statements considering the differences in the student work from different LP levels.

To validate and enrich our feedback statements from the teacher's perspective and expertise, we interviewed three public high school science teachers in the Midwest, US. Asha, Peter and Jake (pseudonyms) teach Physics and Biology classes at the high school level for more than last ten years each. They also taught the *Interactions* curriculum and administered this particular electroscope item to their students multiple times over the last three years. Therefore, they were familiar with the phenomena and content, as well as expected quality responses for the model and justifications for this item. Since they are familiar with the curricular context and hold critical insights about their students' cultural and linguistic backgrounds, and academic needs, successes and challenges in building models and providing justifications to them, we interviewed them to incorporate their expertise and experiences into the feedback statements. Their suggestions for better feedback helped us support our work on connecting human and AI collaboration and intelligence in creating assessment and feedback tools for secondary science classrooms.

We conducted two semi-structured individual interviews (60 mins each) with each of the three teachers via Zoom. Our interview protocol was comprehensive in a sense that reminding teachers a) what LPs are and their importance to support meaningful science learning experiences for students, b) the electroscope item (Figure 1) within the expected models and justifications of Scenario A and B, and c) our rubric for how we assess and scored students models based on the LP levels. After these important reminders and highlighting the importance of their participation for the feedback generation process, we have an in-depth conversation within three core parts of our interview protocol:

Part 1 – Introducing student models and our feedback statements for them: In this part of the interview, we showed 4 different electroscope models that students created. These models represent different LP levels (from 0 to 2 based on the quality of responses). After showing the models, we asked teachers a) what they see and notice in the student model, b) what ideas are visible or missing in the student model and c) what type of feedback to a student they would provide for the particular model. In that way, we were able to capture teachers' thinking on how they see and interpret student models and what critical feedback they would provide for larger group of students with models like those.

Following, we showed proposed feedback statements we created for those 4 models which align with explicated in the LP levels and based on the seven principles of quality feedback that we adopt for this study. This part was one of the most critical parts of the interview as we get to hear teachers' opinions and suggestions for our feedback statements. To get constructive feedback on each statement, we asked 7 main questions that would be reflective of our 7 principles (P1 to P7). (see Figure 2). For example, to be able to hear teachers' suggestions for our

Principle of quality and relevant feedback	**Interview protocol items to get teachers' feedback on how responsive our feedback statements to our principles**
P1. Feedback should be constructive that is compatible with students' prior knowledge	How do you think this feedback builds on what students' have learned throughout the Unit 1?
P2. Feedback should be comprehensible, and interpretable	In what ways do you think your students can make sense of and interpret this statement to revise their model?
P3. Feedback should be actionable and specific	How do you think your students would use this feedback to improve their model?
P4. Feedback should be aligned with learning goals and progressions	How do you think this feedback aligns with the rubric we use for this item? *(Tell the level of the student model and ask how the feedback would align with LP statement for the particular level student is assigned.)*
P5. Feedback should encourage student reflection, self-adjustment and improvement	How do you think this feedback would encourage students to reflect on what they already know, and what they still need to learn to improve their models?
P6. Feedback should encourage students' motivational beliefs and self-esteem	Do you think this feedback would make students feel confident or capable of learning further about the phenomena to improve their models? Why/ Why not?
P7. Feedback should draw connections and be responsive to students' cultural/linguistic resources (such as feedback uses language that clear and meaningful to students, use multimodal resource when they can do it; provide ways/examples to make the phenomena more culturally relevant to students' daily lives and funds of knowledge)	a. What would be the benefit for a student who comes from a diverse cultural background, such as an EBL student, to get this feedback? b. What challenges would a student who comes from a diverse cultural background, such as an EBL student, have when receiving this feedback? c. How might you change this feedback so that students from diverse cultural background, such as EBL students, would find it more useful? Why did you suggest that change? d. Considering all the things we just talked about, do you have any other suggested revisions for the statement?

Figure 2. *Core principles of quality feedback and the interview questions that are responsive to the principles*

feedback statement considering the first principle (P1) - feedback being constructive and compatible with students' prior knowledge - we asked the question: How do you think this feedback builds on what students' have learned throughout the Unit 1?

Part 2 – Introducing student justifications for these models: In this following part, we showed students' justification statements for those 4 models that we showed in Part 1 of the interview Through these justification statements, students provide further explanations to their models and thinking about the Coulomb's law and charge transfer. In this part, we asked questions such as: what do you notice about the student's model and justification together? does the justification change how you view the model and your feedback for the model? If so, how? what kind of feedback would you give to this student so they can improve their model and justification? Through these questions, we highlight teachers' rationale of evaluating the models and justification together and how important to see them together to provide more comprehensive and relevant feedback statements for students to work on.

Part 3 – Introducing student justifications without models: After having a critical conversation about the 4 student models and the justifications for them, we finally introduced 4 additional student responses that doesn't include any model but only provide written justification. These responses are cases in which students provide rich justification statements, but they don't include any (or many) features in their models. Therefore, we only provide the justification statements and not show the associated models in this part. Our goal was to hear teachers' perspective on what kind of feedback they would provide for larger group of students that struggle with developing models but indicate levels of understanding in text. After showing each justification statement, we asked teachers: what do they notice about the student response, what ideas do they think this student is using or missing, and most importantly what kind of feedback they would give to this student to improve their justifications (as well as their models).

Findings:

For this paper, we mainly focus on the analysis of the first part of our interviews. There are two important reasons for that. First, in the first part of the interview, we converse with teachers on a) how they perceive and interprets students' models, b) what type of feedback they provide for those students' models, and c) most importantly what constructive feedback they can provide for our feedback statements that are created for those student models. Second, at the time of the interview, we had proposed feedback statements developed for student models based on the LP, but not the justification. Therefore, we wanted to focus teachers' reviews on the completed proposed statements.

In our analysis for the first part of the teacher interviews (see Figure 3), we conducted thematic analysis and identified two main categories to guide our analysis process: 1. teachers' reflection on student models and their own feedback for these models; 2. teachers' constructive suggestions for our feedback statements for those models. To delve into first theme, we identified a number of codes that would help us to make sense of the themes of teachers' interpretation of the student model and how teachers generate their own feedback with students.

	PART 1: SHOWING STUDENT MODELS + TEACHER FEEDBACK FOR OUR FEEDBACK STATEMENTS								
	REFLECTING ON STUDENT MODELS AND PROPOSING FEEDBACK		**REFLECTING ON OUR FEEDBACK STATEMENT WITHIN THE PRINCIPLES**						
	Interpretation of the student model	Teacher feedback for the student model	P1	P2	P3	P4	P5	P6	P7
Asha	• Struggle to demonstrate the charges • Lack of understanding of attraction and repulsion • Rudimentary use of Coulombs law • Missing representation of protons, neutrons and electrons • Struggle to represent the relationship between charge and force	• Providing directions to make them show charge • Make them use arrows to show how charge travels from the rod to the leaves • Make them show the attraction and repulsion of charges • Make sure to let students experience phenomena and come up with predictions	• Need for more participatory and collective approach (where students can also evaluate their peers' models) • Providing more in-depth facilitation and scaffolding to get better explanations (asking more why questions in feedback)	• Good level of clarity for students to figure out and interpret the feedback	• More specificity needed on directions for how charges travel in both leaved and the idea of force (arrows) • More specificity needed on how leaves positively charged and push each other apart (attraction/ repulsion) • Actionable enough to revise and show the charge transfer and direction	• Good level of consistency with feedback and rubric (agreement with T and us)	• Good level of encouragement for reflection and self-adjustment • Organizational suggestions to support students with diverse reading and comprehension level (e.g., EBL/special ed) • Using bullet points and clear labels	• Supporting interaction and engagement with the model and apply students' thinking • Supporting students' motivational beliefs and self-esteem • Multiple opportunities for revising and improving the model	• Encourage using the translators • Challenge for international student who are not familiar with NGSS type of instruction • Providing an organized table or rubric for them to state what the expectations are to build a better model and justification (EBL/special ed, ALL students)

Peter	• Confusion about the representation of positive charges in leaves) • Indication of understanding on the distance between leaves due to the amount of charges • Confusion about the types of charges (like charges) in the rod and leaves • Confusion about the concept of repulsion	• Make their reasoning visible on representing the positive charges in Scenario A and B (Does scenario B means more charge) • Make them show/explain what charge transfers from the rod to leaves • Make them show/explain where the charge on leaves comes from (to make sure the clarify that rod being negative and leaves being positive)	• Feedback is helpful for further clarifications • More prompts needed on how the charges are distributed on, and how it affects how the leaves are repelled	• Good level of clarity for students to figure out and interpret the feedback	• More specificity and personalized questions/prompts are needed to be responsive of the student model • Asking number of how/why questions to make students reasoning visible (such as: how does your model tell there's more charge on the leaves in scenario B than there is in A?)	• Good level of consistency with feedback and rubric (agreement with T and us)	• Good level of encouragement for reflection and self-adjustment	• More support needed to make them more complete explanations	• Challenge of making students to come up with explanations by highlighting their reasoning (going beyond stating facts) • Pushing student to come up with justification (using more why/how prompts)
Jake	• Indication of charged rods • No indication of how charges attract or repel • Struggle with the concept of	• Make students show the travel of the charges • Make students indicate the amount of charges	• Feedback can encourage them to keep revising and building	Good level of clarity for all students to comprehend and	• Actionable enough in terms of making changes to show the charge transfer and direction	• Good level of consistency with feedback and rubric (agreem	• Good level of encouragement for reflection and self-adjustment	• Feedback supports sense of accomplishment • Encouraging students to show the	• Making them connect with their personal experiences as they learn to justify/explain

	charge and charge transfer • Struggle with the idea of force • No mention of the movement of the leaves (charge transfer)	• Make students indicate the direction of the leaves	their knowledge • Feedback encourages them to see the relationship between parts of electroscope	interpret the feedback on charge transfer in different parts of electroscope		ent with T and us)	• Good connection with the previous activities and conversations they had (magnets, van de Graff generator) • Feedback allows students to see the connection between prior phenomena	differences between the scenarios • Encouraging students to show how the charge of the rod affects and travels each part of the electroscope	• More organized and simplistic version • Bringing real world connections and using analogies

Figure 3. *Summary version of the interview analysis to represent the analytical categories, themes, and codes*

To delve into second category, our seven principles of the quality feedback served as the analytical tool for the analysis process. We positioned the seven principles as the analytical themes and generated a number of codes for each of those themes that would be reflective of teachers' constructive feedback for our feedback statements. In Figure 3, we share the shortened and summarized version of interview analysis that is representative of the coding process.

For example, prior to interview, Asha shared with us that she has number of emergent bilingual students (EBS), immigrant students, and students with special needs in her science classrooms. Therefore, we were able to view her constructive suggestions for our feedback statements focused on principle 7, which relate to how to incorporate equitable teaching practices with equitable assessment practices. Asha shared her experiences and suggestion on how to be more *responsive to students' cultural and linguistic resources* considering their diverse backgrounds and needs in feedback generation process:

> *"I just received a young man from Colombia, and he speaks no English. If the feedback is simplistic enough in the statement, they can use the translator, and they could get a good enough sense of what it means. I also noticed with my English language learners that; they can nicely give me facts. When it comes to explanation, I say that you got to give me a reasoning… Therefore, the first challenge that would happen for them is that they are not coming from a culture that uses NGSS. So, everything is very rigid… They can give me facts, but when it comes to explanation, it's a struggle for them to give me a reasoning. So, when you say justify, that's something they're not used to doing. Then, for the special education students, they're not used to justifying phenomena. They just want to have a straight answer. So, this year, when I focused on justifying, I ask them to tell me exactly "why", which is a very difficult road for them. We want to focus on how things work, but not give them any kind of justification for it... So, maybe having some type of table, a rubric that are available to them that's simply stating very shortly, your model must include this and that and then you list them. For example, your model should have a direction, or your model should indicate charges (positive or negative) … You know, listing things especially for the special education and the ELLs... Make it more organized and simplified. Just make it very simple."*

Here, Asha discusses how challenging it is for her immigrant students who are not familiar with NGSS type of instruction to experience and explain phenomena through online drawing tools. Therefore, she points out the need for *simple and specific* statements that make students a) use effectively the online drawing and modeling tools, b) use translators for words student might not be familiar with, and c) use organizational tables or rubrics that provide bullet points and clear labels, so students can refer to each point as they revise their models and justifications. In addition, Asha suggested to keep considering how to make this model revision process more participatory and collaborative for students. She argued that instead of students solely work on the provided feedback individually, they should also work in small groups where they also provide peer feedback and interact with their peers as they revise their models and justifications.

From that perspective, Asha's suggestions focused more on how assessment and feedback is a core part of equitable teaching and teaching practice. Her suggestions underscore how important to consider cultural, linguistic and academic resources that students have while generating ML-based feedback in the effort of keeping our assessments humanized and equitable.

As another core finding, Peter and Jake provide critical and constructive feedback for us on how to make students think more specifically about the Coulomb's law and charge transfer (i.e. the targeted DCI) in their models and justifications, especially referring to principle 3, making feedback actionable and specific. Similar to Asha, they both mentioned students struggle to come up with comprehensive justification statements including reasoning. Therefore, in their feedback, we see the pattern of pushing students to come up with justifications that go beyond restating facts for their models. To do so, they suggest using more personalized how and why question prompts based on students' models. Some of their suggestions include making students show and explain further a) what charges transfer from rod to leaves, b) where the charges on leaves comes from, c) how the charges are distributed on the system, d) how the distribution affects how the leaves are repelled, and e) how does their model tells us there's more charge on the leaves in scenario B than there is in A. Therefore, the main trend from Peter and Jake's suggestions is prompting for more specific clarifications to push students to come up with more detailed and comprehensive models and justifications.

For example, in one of our interviews with Peter, we showed him a student model (which is representative of LP level 1) and asked him to interpret the student model (see Figure 4), come up with a feedback statement for it, and what suggestions he would have for our feedback statement.

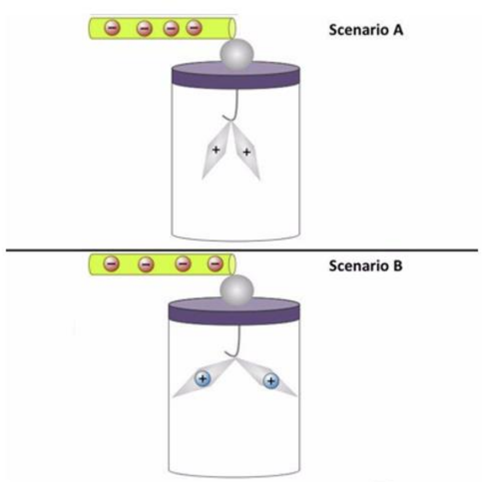

Figure 4. *Student's electroscope model that we shared with Peter during the interview*

After taking a close look on the model, Peter's initial interpretation of the model was "I guess somehow their model implies that there's more positive charge on the leaves than there was in A and somehow they made it blue or a different color than the other, but they have the same amount of negatives on the rod, so they know there's some charge there, but they're having some issues."

After having some conversation about his interpretation of the model, we then shared a proposed feedback statement and then prompted him for a review of the proposed statement:

Proposed feedback statement: Your model mostly shows a complete mechanism of how the charged rod causes the leaves to move apart in the electroscope for both scenarios. Your model also shows that like charges repel. As you revise, how would you modify your model to show a more complete picture of how the charge is distributed through the system and how it affects how far the leaves will be repealed?

Peter's critique and suggestions for our feedback statement: *I would first ask them what's the difference in the two scenarios? Why are these bottom ones (protons) blue and the top ones gray or whatever they are and try to get them to articulate that. That means, you know, somehow, they need to tell us there's more charge. Okay, well, what might be a better way to show me more charge? Then, the other question I would ask them, where does the charge on the leaves of the electroscope come from? Well, it came from the rod and the rod is negative. What charge would transfer from the rod to the leaves? And is what you have consistent with what you just said? To add to that, we would want to say where specifically does the charges on the leaves come from which will help them maybe clarify that rod being negative and the leaves being positive? That's a big piece of it.*

Based on Peter's response, we can see that he really looks for more specific explanations about how student represents the charges in their models across the different scenarios, and what are their in-depth reasoning on the travel of the charge through the questions of: where does and what charge on leaves come from the rod to leaves. Even though there is a good amount of intersection between our and Peter's feedback statements, it is evident that Peter was looking more specifically into the specific details of the student model. Since we created these feedback statements that should relevance for a much larger group of students at a similar LP level (level 1), we tend to frame our statements in a more general way that would still capture the content of students' model and support them to revise their models to reach the expected LP level. For example, for this model, student's use of different shapes and colors to represent the positive charges on the leaves is a unique case, which is uncommon in other student models in this LP level. Therefore, Peter's suggestion brought the critical point to our attention in relation to what extent we aim to provide meaningful personalized feedback with the collaboration of ML systems, while scaling to reach hundreds of students.

Conclusion and Contribution:

This study demonstrates how an automatic feedback system could be effective to build capacity in the field for the design, implementation, and use of formative assessment and feedback in NGSS aligned secondary science classrooms. Especially by bringing in teachers' voice, experiences and expertise, our study supports the efforts of how to combine AI and human endeavor to develop personalized, real-time, automated feedback that aligns with validated NGSS LP. The offered feedback principles and revised feedback statements act as a tool for teachers to help monitor and shift their instruction to assist development of student proficiency aligned to LPs, and support student sensemaking on 3D performance assessments.

According to our interview analysis, we highlight several key findings which all three teachers agreed on. First, the teachers prioritized providing feedback in a more dynamic, dialogical and interactive form. In other words, their suggested feedback provided a number of clarifying questions specifically about students' models. Second, the teachers acknowledged that our proposed feedback statements align with the coding rubric based on NGSS-aligned LP levels. Finally, the teachers highlighted the importance of integrating the core principles we identified into feedback statements. For example, all three teachers agreed that our proposed feedback statements nicely incorporate and align with the principles of 2, 4, 5 and 6 (see Figure 3). They expressed that proposed statements provide a good level of clarity for students to figure out and interpret the feedback. They also agreed that proposed feedback statements have a good potential to motivate students and boost their self-esteem as they keep revising their models and justifications over the unit. Finally, all three teachers suggest that our proposed feedback statements can encourage students to reflect on their previous models and improve them considering the provided personalized feedback.

In addition, teachers specifically keep referring to the importance of incorporating principles of 1, 3, and 7 to the statements. To do so, they provided suggestions on how we can connect our feedback statements to principles in a more contextualized and meaningful way. This was especially true for principles 3 and 7. Such challenges highlight critical considerations as AI-based agents interacting with students becomes more common in classrooms (Shin, Haudek & Krajcik, 2025). For example, teachers shared their perspectives on how to attend to the third principle of *"Feedback should be actionable and specific"* by offering revisions to include more specific prompts and follow-up questions to figure out what the student model really tells us and how feedback can encourage students to take a further step in explaining the mechanism of how the phenomena happens. In addition, considering classroom contexts and what works better for students with diverse backgrounds (related to principle 7), the teachers offered the following revisions to reconsider in feedback statements: 1) bringing simplicity in the wording of the statements, 2) making sure to support emergent bilingual students to make their reasoning visible in explanations, and 3) providing prompts for students on how they can justify phenomena and not just get the straight answer/fact (especially for students with special needs).

Overall, such participatory design and the exploration of using AI as a partner in assessment evolution will significantly contribute to the teaching and learning of science with innovative technologies. Our approach will inform other researchers at NARST who are interested in applying generative AI to support students' learning and teacher practices.

References:

Grapin, S. E., & Lee, O. (2022). WIDA English language development standards framework, 2020 edition: Key shifts and emerging tensions. TESOL quarterly, 56(2), 827-839.

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of educational research*, *77*(1), 81-112.

Kaldaras, L., Akaeze, H., & Krajcik, J. (2021). Developing and validating Next Generation Science Standards-aligned learning progression to track three-dimensional learning of electrical interactions in high school physical science. *Journal of Research in Science Teaching*, *58*(4), 589–618. https://doi.org/10.1002/tea.21672

NGSS Lead States. (2013). *Next generation science standards: For states, by states*. National Academies Press.

Nicol, D. J., & Macfarlane-Dick, D. (2006). Formative assessment and self-regulated learning: A model and seven principles of good feedback practice. *Studies in higher education*, *31*(2), 199-218.

Penuel, W. R., & Shepard, L. A. (2016). Social models of learning and assessment. *The Wiley handbook of cognition and assessment: Frameworks, methodologies, and applications*, 146-173.

Qadir, J., Taha, A. E. M., Yau, K. L. A., Ponciano, J., Hussain, S., Al-Fuqaha, A., & Imran, M. A. (2020). Leveraging the force of formative assessment & feedback for effective engineering education.

Shin, N., Haudek, K., & Krajcik, J. (2025). *The Potential of Using AI to Improve Student Learning in STEM: Now and in the Future*. Community for Advancing Discovery Research in Education (CADRE). Education Development Center, Inc.