Scaffolding Collaborative Critique and Sensemaking During Technology-Enhanced Science Explanations in Elementary School Classrooms

Elissa Sato, University of California at Berkeley, 4523 Tolman Hall, Berkeley CA 94720, elissa.sato@berkeley.edu
Vanessa Svihla, University of New Mexico, 1 University of New Mexico, Albuquerque, NM 87131, vsvihla@unm.edu

Abstract: This qualitative study extends a comparison study investigating critique and feedback in a technology-enhanced sixth-grade earth science curriculum unit. The unit was designed to elicit and scaffold collaborative sensemaking of criteria for explanations in science. Quantitative analysis suggested the activity had limited success in terms of prompting revision, but the qualitative analysis illustrated different dimensions of engagement in critique, demonstrating the rich potential of critique for collaborative sensemaking. Implications for instructional design are discussed.

Introduction

Generating explanations is central in science (National Research Council, 2007) and important in learning (Chi et al. 1994; NRC, 2007), yet non-trivial for students learning complex topics. Typical instruction does not support students in the practice of generating explanations, yet prior research has shown the value of scaffolding students to construct and evaluate explanations (e.g., McNeill, Lizotte, Krajcik, & Marx, 2006; Sandoval & Reiser, 2004). We situate the generation of explanations within the knowledge integration (KI) framework (Linn & Eylon, 2006), which identifies four processes: eliciting ideas, adding new ideas, distinguishing among ideas, and refining the repertoire of ideas through reflection. From this perspective, learning to generate quality explanations involves, in addition to synthesizing relevant domain knowledge, making sense of and applying criteria for what constitutes a good explanation. We conceptualize scaffolded critique of explanations as generative activities in that they support students in a) distinguishing among criteria in the context of specific explanations, and b) sorting through and refining the ideas captured by the targeted explanations. We investigate how critiquing explanations might impact students' ability to distinguish and refine their repertoire of ideas. We present two cases to highlight ways students engaged with a particular critique activity, examining how students make sense of commonly used, yet vague criteria for good scientific explanations.

Methodological Approach: Building on Prior Work

This paper presents findings that extend a prior comparison study investigating the source of feedback (i.e., from the teacher or a peer) and the quality of the explanation being critiqued (i.e., high or low KI score, cf. Linn, Lee, Tinker, Husic, & Chiu, 2006) and their impact on students' revised explanations. Quantitative analysis found a significant advantage for instructor feedback on gains after controlling for the quality of the original explanation. A comparison of instructor and peer feedback revealed that instructor feedback was more prompt-specific than peer feedback, although students still struggled to apply specific feedback during revision (for a more detailed discussion of the study findings, please refer to Sato and Linn, 2011 in preparation). Although this may suggest that the activity did not help students distinguish between their ideas, our video data suggest otherwise and indicate elements in activity design that have implications for instruction. In this qualitative analysis, we present two cases that illustrate the kinds of engagement during critique that were observed in the data corpus as a whole. Although their orientations were neither one-dimensional nor fully consistent throughout the critique activity, the orientations were generally representative of each dyad's engagement with the task. Due to space limitations, we primarily focus on one case.

Context

A sixth-grade technology-enhanced earth science curriculum unit, *Global Climate Change* (GCC), was developed using the Web-based Inquiry Science Environment (WISE, Linn, Davis, & Bell, 2004) based on the KI perspective (Linn & Eylon, 2006). Students worked in pairs throughout the unit. They generated an explanation for a phenomenon, then their ideas about criteria were elicited in a *first critique* of two sample explanations chosen from student responses in a previous implementation of the project (Table 1). We focus on this activity as it affords insight into how students make sense of the criteria presented in the task. Because students tend to focus on surface features rather than the underlying ideas, explanations were selected to represent these: one was stylistically sound in terms of spelling and grammar, but sparse in terms of science content; the other had imperfect spelling and grammar, but described the phenomenon in detail. Students critiqued the two in succession. Our goal was to implicitly prompt students to compare the two and help them distinguish between their existing ideas about what makes a good explanation. Students rated each explanation

for both surface (spelling and grammar) and science content (e.g., the response needs more evidence□the science ideas are wrong or vague□the science ideas could be described in more detail) from a list of criteria, then e□plained their choice for science content. □ur goal was to prompt students to discuss criteria commonly encountered during instruction□there was no correct choice per se among the list of science content criteria, nor were they mutually e□clusive. By asking students to choose the criterion they felt best captured the science content in the e□planation, we hoped to motivate students to engage in sensemaking of the criteria.

Table 1: Preselected E□planations Criti□ued by Students during the □irst Criti□ue

Prompt: What happens to global temperature in an environment with low albedo□What happens to solar radiation (SR) in the model in step 4.2 that supports your answer□	
Characteristics	*Preselected E□planation*
Good style (spelling□grammar), vague science content	It changed a lot. It went down then bounced.
Imperfect style (spelling□grammar), detailed science content	THe globle tempeture went down when the albedo was low, like for the ocean, or when the albeod reflected only 5 percent of the solar radiation. When the soler radiatoin was reflecting, it coud not change into heat energy it □ust went back to space.

Case 1: The Sense-Making Pair

Kostas and Ted studied the *peer feedbac*□version of the GCC unit. Though they began the unit in the bottom □uartile of their treatment group based on pretest measures, they were both highly motivated students who engaged with the unit and held e□tended discussions before deciding on responses to e□planation prompts. It was common for the two to raise □uestions and ob□ections if one did not share the other□s opinion or understanding of a particular activity. Although they did not generate detailed feedback, their discussions illustrated rich instances of both surface and content criteria sensemaking throughout criti□ue. Kostas and Ted negotiated whether accurate spelling and grammar can be □udged independently of science content during surface criti□ue of both e□planations. Kostas appeared to grapple with the distinction between stylistic clarity and its impact on the clarity of science content. In the transcript below, the pair has begun surface criti□ue of the first e□planation (□It changed a lot. It went down then bounced,□Table 1). Ted asserted his rating (□good: few errors□), and after an initial assent, Kostas had reversed his position.

T: What would you give it□ □ut of three□
□n a scale □of□□□
K: □□□ne□□□
T: □ne□((incredulously))
K: Yeah, this is like, so bad. ((sweeps finger across the e□planation))
T: □o, it□s not□
K: It□s not good English, you know□
Like, □It changed a lot. It □bounced and□□□□ ((reading out loud in a singsong))
T: □□That□s in□□□scientific content and clarity.□((pointing to ne□t criti□ue prompt for science content and clarity))
That□s what it means.
Here, □the science idea could be described in,□□some ideas are wrong or vague.□((reading some of the science content criteria□K leans in toward the monitor as T reads))
K: Like, what changed, though□((pointing to □It changed□portion of first e□planation))

Although Kostas did not specify his misgivings other than that they were not about spelling, he elaborated that the e□planation had □bad English□(lines 7-8). Ted countered that Kostas was referring to science content and clarity, not spelling and grammar (lines 9-12), and oriented Kostas to the ne□t criti□ue prompt and its list of criteria targeting science content and clarity. By asking Ted, □what changed□□(line 14) Kostas seemed to have been trying to identify what was bothering him on a more specific level. Given that the e□planation did not specify what □it□was that changed, this problem can be viewed simultaneously as an issue of vagueness in language and of lack of clarity in content. Here, Kostas was focusing on the vagueness of the language. In effect, Kostas may have been alluding to sophisticated ways in which the surface and content criteria can be considered to overlap. Although they eventually chose □good□as their response, the discussion re-emerged during the surface criti□ue of the second e□planation, indicating that making sense of the surface and content criteria was important for Kostas. Kostas and Ted continued to engage in sensemaking discussions during the science content criti□ue for both e□planations. In the e□ample below, they had formed initial impressions about whether each criterion was applicable to the particular e□planation before reaching the last criterion on the list,

□the science ideas could be described in more detail.□The ensuing discussion illustrates the rich conversations they had as they negotiated the meanings of the criteria:

K: This is... This is it. ((points to □needs more detail□criterion))
In more detail... □cause □the e□planation□□ust said, □It changed a lot. It went down then bounced.□
So it needs like, more support... evidence □another criterion□, and like, more detail.
T: Yeah, it□s more evidence.
It□s... □o, it doesn□t need more detail. □Cause, I mean... □it□has a L□T of detail.
Well, not that... well□ ((sighs))
It has like, a 4 out of 5 detail.
It said, □It went down, then bounced.□That□s kind of describing it in detail.
But it□s not describing it in like, proper grammar.
K: I think this is right. ((points to □needs more detail□))
T: I think□ no.
This is □need more detail.□I think it□s □needs to have evidence.□
K: I□m really confused between both of those. ((points to □needs more detail□and □needs to have evidence□criteria))
I don□t know. □inaudible□
I think what you chose is right.
T: I don□t think there□s, like, a wrong answer in this □uestion.

Although the target e□planation prompt e□plicitly asked for evidence from a previously e□plored model (Table 1), Kostas and Ted did not discuss whether the e□planation contained e□plicit references to the model. Instead, they discussed how to distinguish evidence from detail. Ted struggled to clarify why he believed that □needs to have evidence□ was the more appropriate criti□ue over □needs more detail (lines 19-21).□Ted□s concluding remark, □But it□s not describing it in like, proper grammar (line 23),□is notable in light of the fact that they had already evaluated the grammar and spelling of the e□planation to be □good.□It may suggest that he had some ideas about what counts as evidence in science e□planations as opposed to simply descriptive detail. His struggle is similar to Kostas□earlier attempt to distinguish between surface stylistic clarity and content clarity in that Ted also struggles to articulate why he believes one criterion is more applicable than the other (lines 19-21). In this case, however, Kostas and Ted did not pursue this line of thought to the same e□tent as they did previously for Kostas. After Kostas professed his uncertainty about what evidence and detail meant (line 27-28), Ted responded that any of the criteria may be acceptable answers (line 31). □nable to pin down what e□actly they felt was lacking in the e□planation, they eventually decided on □the science ideas are wrong or vague□ and wrote, □We think that the student needs to add more detail and support his answer because we can□t understand the answer, it is vague and needs to become clearer.□Whereas the first e□planation was intended to serve as an e□emplar for good style and vague science content, the second was intended to serve as an e□ample for poor style but detailed science content. As such, the second e□planation was rife with spelling and grammatical errors. It may therefore seem a straightforward task to score the spelling and grammar, perhaps even more so than the first e□planation due to the obvious errors. However, the confusion between stylistic and content criteria reemerges in their e□change. In the transcript below, Kostas and Ted have □ust finished identifying numerous spelling errors in the second e□planation and were about to choose a criterion for spelling and grammar. As before, they disagreed, but this time, Kostas used the first e□planation as a point of reference when □uestioning Ted□s assessment of □□ot so good.□This prompted Ted to repeat his earlier attempt to orient Kostas toward science content criteria to address his concerns with the e□planation.

K: □□□::h.□□
T: □□□::h.□□
T: □::::not good. ((selects □□ot so good□criterion with pointer))
K: □o, uh, it□s, □it□s good.□□((points to □Good□criterion with finger))
T: □□It has many spelling.□□
T: Has □□many□□spelling.
K: □□□hh.□□ □hh.
K : Yeah, but this like, so you□re saying the, the one before that has like one sentence, is □good... few spelling and □the□□□ ((points to first e□planation, then to □Good□criterion))
T: □□That□s□□for grammar and spelling. ((traces □grammar and spelling□in criti□ue prompt□K drums fingers on desk))
T: I like this, the idea that it was, the□ ((points to second e□planation))
K: Yeah, you□ oh, oh. ((nods, points to science content and clarity criteria list))
T: I like, for science clarity, □response,□uh, □needs evidence to e□plain,□so□ ((points at criterion))

And the::n, □ideas could be elaborated.□
I think... I think yeah.

In □uestioning Ted□s □not so good□surface criti□ue, Kostas drew attention to the brevity of the first e□planation relative to the second (lines 39-41). This indicates that Kostas was still struggling to distinguish between surface and content criteria. □rom this, his earlier dissatisfaction with the first e□planation for □not good English□ seemed to have remained unresolved. In response, as in the first case, Ted repeated his assertion that Kostas□ issue with the first e□planation pertained to science content and not spelling and grammar (lines 41-42). This time, Kostas nodded and pointed to the science content criteria list as if to accept and affirm Ted□s claim. □or Kostas, criti□uing two e□planations with contrasting dimensions seemed critical for making progress in teasing apart surface and content issues. Although Kostas and Ted engaged in sensemaking discussions about both surface and content criteria, their rich conversations and oral criti□ue were not fully captured in their written feedback. □ollowing the discussion, they wrote, □We chose this answer because we thought that his□her e□planation could describe his□her evidence in more detail.□However, their active engagement with the criti□ue tasks indicates that the criti□ue activity provided them with multiple opportunities for beginning to make sense of and distinguish among the criteria.

Case 2: The Task-Oriented Pair

Tammy and Giulia studied the *instructor feedbac□(lo□ □I)* version of the GCC unit. They began the unit at the top □uartile of the treatment group and were highly motivated students who remained on-task throughout. They typically approached prompts from what might be characterized as task completion-oriented remarks (e.g., □We could say□ □We could put□ □) when beginning to formulate their responses. Although they were able to identify issues during criti□ue, their sensemaking discussions were brief. During criti□ue, Tammy and Giulia discussed the criteria until they reached agreement. When choosing the criterion, one of the pair would immediately suggest a candidate, generally with no rationale and without subse□uent disagreement, as the following e□ample of content criti□ue illustrates:

T: □Score this response for SCIE□CE content and clarity.□
((Reads list of criteria out loud (□igure 3)))
T: I like the last one. □□The science ideas can be described in more detail.□□
G: Yeah. ((T checks this criterion))
T: □E□plain your choice for SCIE□CE C□□TE□T and CLARITY and give an e□ample from the response to help the student improve the e□planation.□

Tammy immediately declared that she liked the last criterion (line 3) but without stating why. Giulia agreed without providing rationale (line 4), and the pair moved on to the ne□t prompt (line 5). Similar episodes were observed for both criti□ues. Thus, if one student did not find cause to disagree with the other□s proposal, there was no further discussion, and the pair moved on to the ne□t task without elaborating on their decisions. There were instances that could have led the dyad to e□plore a criterion□s meaning. However, they did not pursue those opportunities, instead rapidly coming to agreement and moving on to the ne□t task. It is possible that they knew each other well and felt so aligned with each other that they were not motivated to e□plicate their reasoning. Another e□planation is that they did not engage in sensemaking beyond what was necessary to complete the task. They achieved consensus about which criterion to select, but it is unclear whether their rationales for their opinions actually aligned because neither of them e□plicated the basis for their respective opinions. Similar episodes during their criti□ue tasks were observed, suggesting that, while Tammy and Giulia both engaged with the criti□ue tasks and collaborated with each other, they missed several opportunities to delve more deeply into the criteria.

Cross-Case Comparison

□ur findings from the comparison study seemed to suggest that the activity of generating feedback for a peer had little impact. The case studies indicate that a lack of detail in the written responses does not *de facto* indicate a lack of sensemaking. The case studies highlight the range of engagement during a criti□ue activity (Table 2), and potential tradeoffs between sensemaking and efficient task completion. Although Kostas and Ted did not generate detailed feedback, their discussions illustrated rich instances of criteria sensemaking that are arguably at least one of the essential first steps students must take before they can engage productively in criti□ue and generate good feedback. In contrast, although Tammy and Giulia were similarly highly engaged with the tasks as a team, their discussions did not lead to rich sensemaking. □nlike Kostas and Ted, who engaged in cross-e□planation comparisons in their discussions, Tammy and Giulia seemed to approach each e□planation criti□ue as an isolated, unconnected task□this limited the activity□s potential to help the students develop integrated understanding. The dyad seldom leveraged opportunities to delve deeply into the criteria and try to understand

the nuanced distinctions to the same e□tent as Kostas and Ted. Meager responses belie the diversity of sensemaking about both science content and evaluative criteria students engaged in. This speaks to the findings of the comparison study: few students provided detailed, actionable feedback to their peers, yet the activity of providing criti□ue may have provoked sensemaking. The task was specifically designed to encourage students to struggle with commonly-encountered vague criteria□the degree to which students made sense of these criteria is tied to a tradeoff between efficient completion of isolated tasks, and more time-consuming negotiation and consensus building.

Table 2: Summary of Cases with Dimensions of Engagement

Case	*Students*	*Engagement as a Team*	*Criteria Sensema□ing*	*Tas□Completion □rientation*
Sensemaking Pair	Kostas and Ted	Yes	Yes	□o
Task-□riented Pair	Tammy and Giulia	Yes	□o	Yes

□m□l□c□t□ons

These findings provide insight into the affordances and constraints of peer criti□ue and identify promising avenues for scaffolding productive criti□ue activities in technology-enhanced instruction. Although students encounter abstract criteria for what makes a good science e□planation in instruction, findings indicate that students would benefit from opportunities to reflect on and distinguish between these criteria in the conte□t of specific e□planations. Providing students with more than one peer e□planation to serve as contrasting cases during criti□ue may increase the activity□s potential for helping students refine their criteria for e□planations. □orcing students to select among non-mutually e□clusive criteria prompted some students to engage in sensemaking of the criteria. The case study findings indicate the potential value of e□plicitly problematizing the criteria selection process to more effectively engage students in criteria sensemaking. However, the case studies also point to the need for designing criti□ue activities that prompt students to contrast e□planations and distinguish their ideas across multiple e□planations instead of considering each e□planation in isolation during criti□ue.

References

Chi, M. T. H., de Leeuw, N., Chiu, M., & LaVancher, C. (1994). Eliciting self-explanations improves understanding. *Cognitive Science*, *18*(3), 439-477.

Linn, M. C., Davis, E. A., & Bell, P. (Eds.). (2004). *Internet Environments for Science Education*. Mahwah, NJ: Lawrence Erlbaum Associates.

Linn, M. C., Lee, H.–S., Tinker, R., Husic, F., & Chiu, J. L. (2006). Teaching and assessing knowledge integration. *Science*, *313*, 1049-1050.

McNeill, K. L., Lizotte, D. J., Krajcik, J., & Marx, R. W. (2006). Supporting students' construction of scientific explanations by fading scaffolds in instructional materials. *Journal of the Learning Sciences, 15*(2), 153-191.

Sandoval, W. A., & Reiser, B. J. (2004). Explanation-driven inquiry: Integrating conceptual and epistemic scaffolds for scientific inquiry. *Science Education*, *88*, 345-372.

Sato, E., & Linn, M.C. (2011, April). Developing Criteria for Explanations in Science: Scaffolding Peer Critique and Feedback in Technology-Enhanced Instruction. Poster presented at the 2011 American Educational Research Association Annual Meeting, New Orleans, LA.

Svihla, V., Gerard, L., Ryoo, K., Sato, E., Visintainer, T., Swanson, H., Linn, M.C., Lee, H-S., Liu, O.L., & Dorsey, C. (2010, June). Energy across the Curriculum: Cumulative Learning Using Embedded Assessment Results. Paper presented at the International Conference of the Learning Sciences, Chicago, IL.

National Research Council. (2007). Taking science to school: learning and teaching science in grades K-8. Washington, DC: National Academy Press.