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; RC, 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., McEneill, Lizotte, Kračík, & Mar, 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.

Ethical considerations

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 not 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. Our 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 exclusive. 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 Critiqued by Students during the First Critique |

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Preselected E:planation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Good style (spelling grammar),</td>
<td>It changed a lot. It went down then bounced.</td>
</tr>
<tr>
<td>vague science content</td>
<td></td>
</tr>
<tr>
<td>Imperfect style (spelling grammar),</td>
<td>The global temperature went down when the albedo was low, like for the ocean, or when</td>
</tr>
<tr>
<td>detailed science content</td>
<td>the albedo reflected only 5 percent of the solar radiation. When the solar radiation was</td>
</tr>
<tr>
<td></td>
<td>reflecting, it could not change into heat energy it just went back to space.</td>
</tr>
</tbody>
</table>

Case 1: The Sense-Making Pair

Kostas and Ted studied the peer feedback version of the GCC unit. Though they began the unit in the bottom quartile of their treatment group based on pretest measures, they were both highly motivated students who engaged with the unit and held extended discussions before deciding on responses to e:planations. It was common for the two to raise questions and objections 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 critique. Kostas and Ted negotiated whether accurate spelling and grammar can be judged independently of science content during surface critique 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 critique 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 out of three?
K: Ne
T: Ne
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 next critique 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 next critique 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 critique 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 critique for both e:planations. In the example 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 explanation just 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.
Its... so, it doesn't need more detail. Cause, I mean... it has a LoT 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 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 question.

Although the target explanation prompt explicitly asked for evidence from a previously explored model (Table 1), Kostas and Ted did not discuss whether the explanation contained explicit 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 criterion 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 explanation to be good. It may suggest that he had some ideas about what counts as evidence in science explanations 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 extent 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). Unable to pin down what actually they felt was lacking in the explanation, 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 explanation was intended to serve as an exemplar for good style and vague science content, the second was intended to serve as an example for poor style but detailed science content. As such, the second explanation 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 explanation due to the obvious errors. However, the confusion between stylistic and content criteria reemerges in their change. In the transcript below, Kostas and Ted have just finished identifying numerous spelling errors in the second explanation and were about to choose a criterion for spelling and grammar. As before, they disagreed, but this time, Kostas used the first explanation as a point of reference when questioning Ted’s assessment of not so good. This prompted Ted to repeat his earlier attempt to orient Kostas toward science content criteria to address his concerns with the explanation.

K: 
T: 
K: o, uh, it's good. (points to Good criterion with finger)
T: It has many spelling. (points to needs more detail)
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... (points to first explanation, then to Good criterion)
T: That's for grammar and spelling. (traces grammar and spelling in criteria prompt K drums fingers on desk)
T: I like this, the idea that it was, the (points to second explanation)
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 explanation, so it (points at criterion)
And the::n, ideas could be elaborated. 
I think... I think yeah.

In questioning Ted’s not so good surface critique, Kostas drew attention to the brevity of the first explanation relative to the second (lines 39-41). This indicates that Kostas was still struggling to distinguish between surface and content criteria. From this, his earlier dissatisfaction with the first explanation 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 explanation 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. For Kostas, critiquing two explanations 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 critique were not fully captured in their written feedback. Allowing the discussion, they wrote: We chose this answer because we thought that his/her explanation could describe his/her evidence in more detail. However, their active engagement with the critique tasks indicates that the critique 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 feedback version of the GCC unit. They began the unit at the top quartile 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 criticism, their sensemaking discussions were brief. During criticism, 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 subsequent disagreement, as the following example of content critique illustrates:

T: Score this response for SCIENCE content and clarity. ((Reads list of criteria out loud (figure 3)))
T: I like the last one. The science ideas can be described in more detail. 
G: Yeah. (T checks this criterion)
T: Explain your choice for SCIENCE CONTENT and CLARITY and give an example from the response to help the student improve the explanation.

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 next prompt (line 5). Similar episodes were observed for both critiques. 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 next task without elaborating on their decisions. There were instances that could have led the dyad to explore a criterion’s meaning. However, they did not pursue those opportunities, instead rapidly coming to agreement and moving on to the next task. It is possible that they knew each other well and felt so aligned with each other that they were not motivated to explicate their reasoning. Another explanation 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 explicated the basis for their respective opinions. Similar episodes during their critique tasks were observed, suggesting that, while Tammy and Giulia both engaged with the critique tasks and collaborated with each other, they missed several opportunities to delve more deeply into the criteria.

Cross-Case Comparison
Our 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 critique 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 critique 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. Unlike Kostas and Ted, who engaged in cross-explanation comparisons in their discussions, Tammy and Giulia seemed to approach each explanation critique 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 extent 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 critique 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

<table>
<thead>
<tr>
<th>Case</th>
<th>Students</th>
<th>Engagement as a Team</th>
<th>Criteria Sensemaking</th>
<th>Task Completion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sensemaking Pair</td>
<td>Kostas and Ted</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Task-Oriented Pair</td>
<td>Tammy and Giulia</td>
<td>Yes</td>
<td></td>
<td>Yes</td>
</tr>
</tbody>
</table>

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

**References**