Characterizing Students' Epistemic Considerations

An Automated Computational Approach for Embedded Assessment Responses

Christina Krist[a] ckrist@umd.edu and Joshua M. Rosenberg[b] jrosen@msu.edu
[a]Northwestern University, [b]Michigan State University

Why Epistemic Considerations? (ECs)

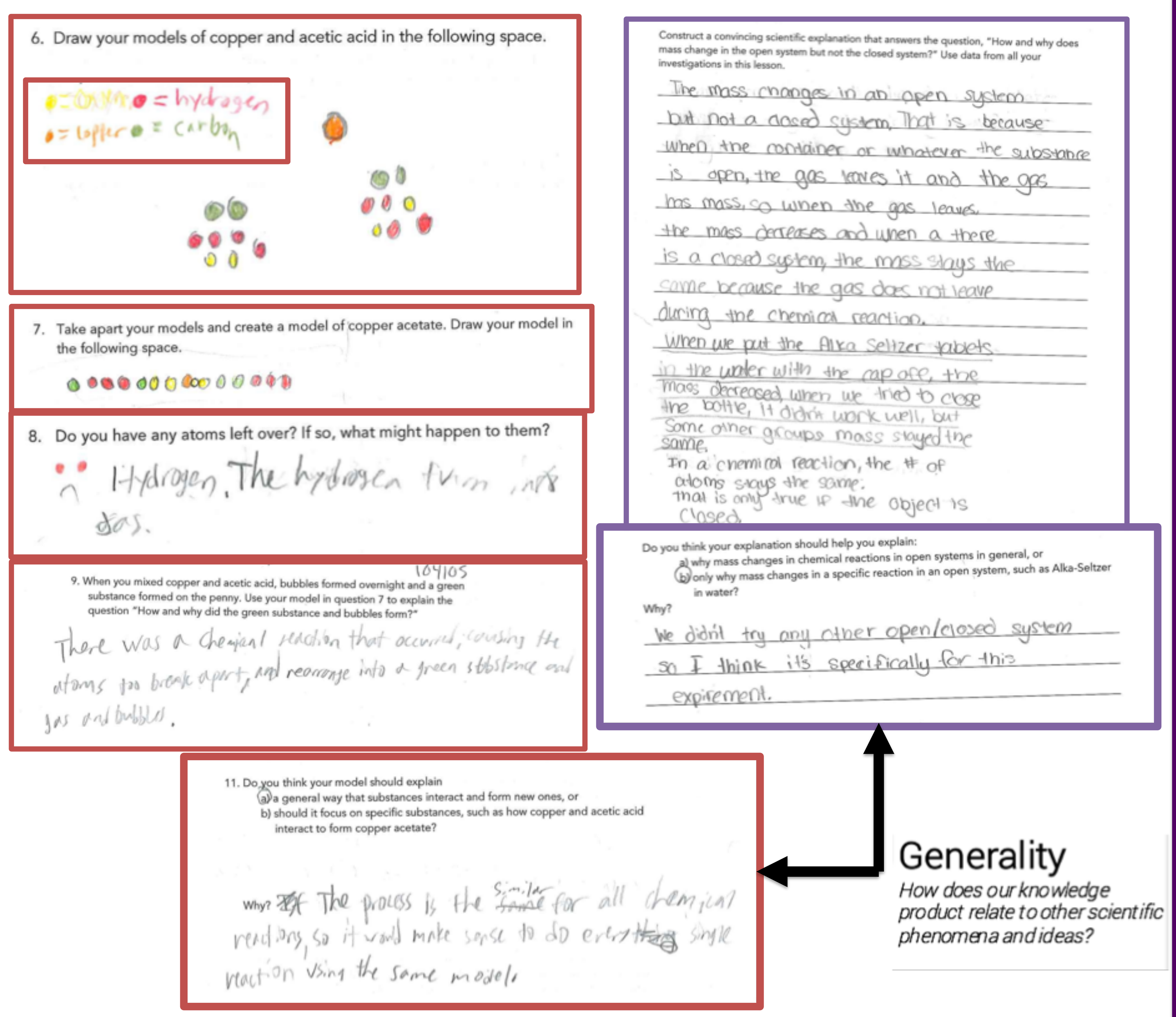
- Current reforms in science education emphasize scientific practices as the means by which students develop and use scientific ideas
- Meaningful engagement in scientific practices requires that students learn how to engage with others to productively build ideas using disciplinary criteria (e.g., Engle & Conant, 2002)
- Also requires student learn *how and why* those disciplinary criteria are effective for accomplishing disciplinary knowledge-building goals (Manz 2014)
- Epistemic considerations = ideas that guide students' meaningful construction, revision, and evaluation of explanatory models (Berland et al., advance online publication)

Identifying ECs in Use

- Characterizing students' epistemic considerations through embedded assessment responses:

IQWST 7P L8.3

IQWST 7P L12.2

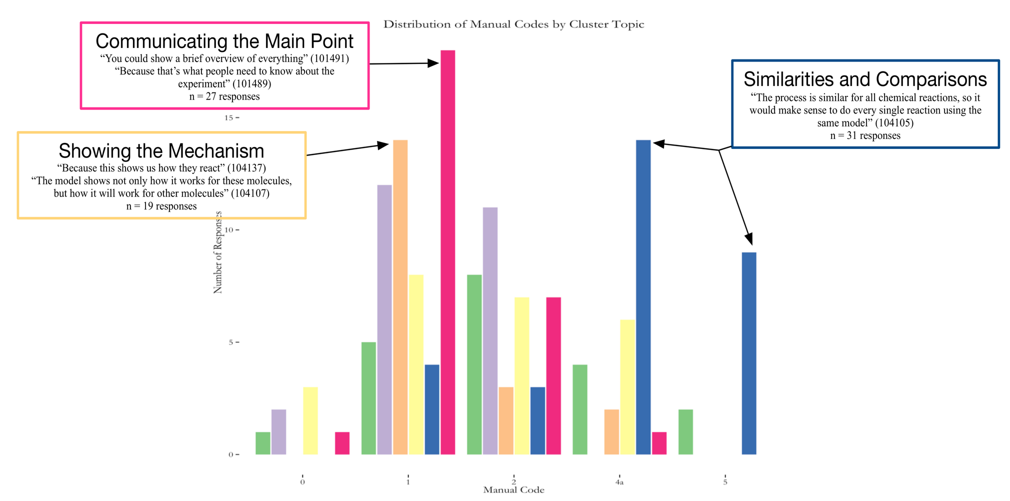


Findings

1. Initial Coding Scheme

Code	Description of Code	Sample Responses
1	One level (G or S): No rationale	"S; Because that's the whole point of the model" (101309)
2	One level (G or S): With rationale	"S; That can explain the atoms [rearrangement] better" (101536)
3*	Analogical mapping	*Responses were rare in dataset*
4a	Level-crossing: S to G	"G; This should help with all open systems in general because we know that if this happens with other reactants, the atoms would still leave in an open system" (104148)
4b*	Level-crossing: G to S	*Responses were rare in dataset*
5	Level-crossing: Boundary conditions of G and/or S	"S; Because the question is asking you about only 1 specific thing and if I talked about all the chemical reactions in general it would not make sense because in different chemical reactions different things happen, like bubble, smell" (101323)

2. Cluster Analysis



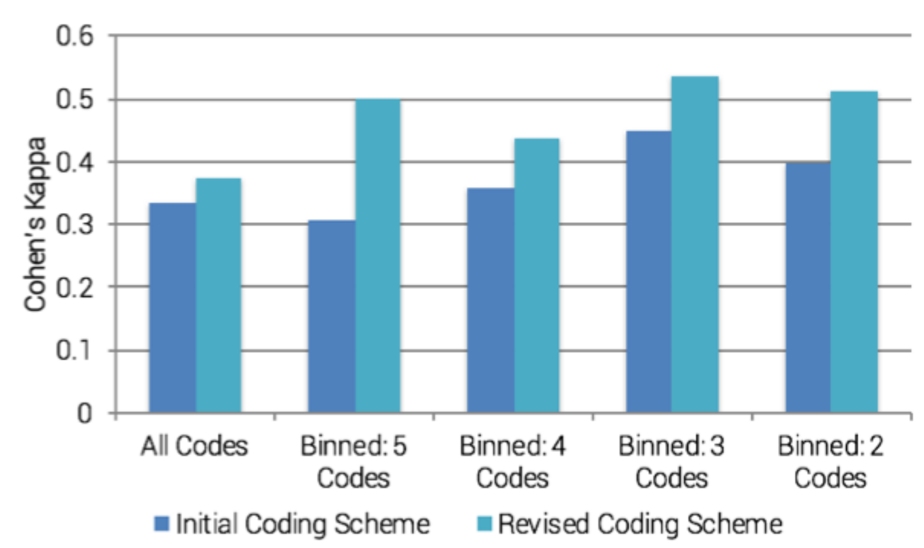
Initial Coding Scheme Manual Coding Reliability: .847 (n = 128)

Revised Coding Scheme Manual Coding Reliability: .896 (n = 24)

3. Revised Coding Scheme

Code	Description of Code	Sample Responses
1.1	Foregrounding literal task goal or instructions: "That's what we did"	"S; Because that's the whole point of the model" (101309) "S; Because we are focusing on the chemical reaction that happened on the copper without touching the vinegar" (104157)
2.1	Foregrounding Communication: G or S is better because of criteria related to accuracy, detail, or clarity of communication	"S; Because it helps describe the reactions better" (101536) "S; It gives more reasoning and evidence" "S; It will explain more of what happened"
2.2	Foregrounding Communication: Emphasis is on the audience's understanding, thinking, or learning	"S; So whenever someone looks at it they can see and tell what happened and with what it happened with" "S; Because if you go in this way, it proves your point about the two substances"
3	Foregrounding Mechanism: G or S better shows how and why something is happening	"G; It should explain how the chemicals create a chemical reaction and form a new substance" "S; My opinion on the multiple choice is B because then you learn about the factors that mainly cause the substance to react"
4.0	Foregrounding Generality: Recognizing the goal of the activity was to know/understand some general idea	"G; Cause we wanted to no [sic] how new substances were made"
4.1	Foregrounding Generality: Does not apply as generally as A says; OR, is doing more than what B says	"S; Not all atoms react the same way" "G; It was supposed to show why the mixture formed, not how the specifics were created"
4.2	Foregrounding Generality: G or S is better able to apply, generate, and/or predict other situations	"Both; A for the fact of how chemical reactions happen, but B for how this reaction happens" "G; Because it will help with other problems in the future and is a great baseline for other models to come. It does ask for a specific model, but this model will be one of many, and I must have a base to start on" "S; When I explain why mass changes in a specific reaction in an open system such as Alka-Seltzer in water so then we can compare the evidence to mass changing to other [reactions]"

4. Accuracy of Automated Analysis



Criteria for interpreting Cohen's Kappa:

- *<= 0: poor*
- *.01 - .20: slight*
- *.21 - .40: fair*
- *.41 - .60: moderate*
- *.61 - .80: substantial*
- *.81 - .1: almost perfect*

(Landis & Koch, 1977)

Method

- Our method applies a computational approach—commonly referred to as statistical natural language processing (NLP) or automated text analysis (Sherin, 2013)—to the analysis of open-ended student responses using both *supervised* (Naive Bayes classification) *unsupervised* (hierarchical cluster analysis) strategies
- First, we coded 175 responses to develop an initial coding scheme and establish its reliability
- Next, we "trained" a Naive Bayes classifier to then "test" how reliably the computational approach performed
- To improve the reliability of the computational approach, used an unsupervised approach to cluster responses and improve the coding frame
- Finally, we trained and tested a Naive Bayes classifier with the revised coding frame

Discussion

- Sherin (2013) highlights usefulness of computational methods on qualitative data for an additional metric of *reliability*
- We demonstrate how taking advantage of computer's pattern-finding "skills" can assist in *conceptual development*
- Ideas-in-Progress for continued improvement:
    - Define features by hand, based on common words
    - By-grams (catch phrases rather than individual words)
    - Use a parts-of-speech tagger
    - Use another classifier (e.g., support vector machine)
    - More data (both more responses and more of a given student's response)
    - Other suggestions???