Characterizing Students’ Epistemic Considerations
An Automated Computational Approach for Embedded Assessment Responses

Christina Krist\textsuperscript{a} krist@umich.edu and Joshua M. Rosenberg\textsuperscript{b} jrosen@msu.edu
\textsuperscript{a}Northwestern University, \textsuperscript{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 (Shen, 2014)
- Epistemic considerations = ideas that guide students’ meaningful construction, revision, and evaluation of explanatory models (Wertsch et al., advance online publication)

Identifying ECs in Use

- Characterizing students’ epistemic considerations through embedded assessment responses:

Findings

1. Initial Coding Scheme

<table>
<thead>
<tr>
<th>Code</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Generalized task goal or instructions:</td>
</tr>
<tr>
<td>2</td>
<td>Foregrounding: General:</td>
</tr>
<tr>
<td>3</td>
<td>Foregrounding: Communication:</td>
</tr>
<tr>
<td>4</td>
<td>Foregrounding: Sensitivity:</td>
</tr>
<tr>
<td>5</td>
<td>Foregrounding: Personal:</td>
</tr>
</tbody>
</table>

2. Cluster Analysis

- Identifying clusters of codes
- Assigning codes to clusters
- Analyzing the relationships between clusters

3. Revised Coding Scheme

- Refining the initial coding scheme
- Incorporating feedback from experts
- Improving the accuracy of the automated analysis

Method

- Our method applies a computational approach—commonly referred to as statistical natural language processing (NLP) or automated text analysis (Shen, 2013)—to the analysis of open-ended student responses using both supervised (Naive Bayes classification) and 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, we 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

- Shen (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???