Designing Student Assessments for Understanding, Constructing and Critiquing Arguments in Science

Katherine L. McNeill¹, Seth Corrigan², Jacqueline Barber², Megan Goss² & Amanda M. Knight¹
¹ Boston College, Chestnut Hill, MA ² Lawrence Hall of Science, Berkeley CA

Argumentation and Assessment

- Science proficiency includes students reading and writing complex informational text as well as engaging in written and oral argumentation (Pearson et al., 2010).
- The educational importance of argumentation is reflected in recent science education research (e.g., McNeill et al., 2006; Duschl, Schweinberger, & Shouse, 2007, Osborne, 2010; Sampson & Clark, 2008), the new K-12 Science Education Frameworks (NRC, 2012) and in the English Language Arts common core in the standards for literacy in science and technical subjects (Common Core State Standards Initiative, 2010).
- The rise of standards and associated assessments over the past decades has played an increasing role in what teachers do in science classrooms (Alton, 2011).
- Large-scale assessments can impact teachers’ pedagogical practices: instruction through efforts to align with the types of items included on the test (Britton & Schneider, 2007). Furthermore, classroom assessments can provide important information about student learning, which can then be used by the teacher to modify instruction (Bell, 2007).
- Although there has been important research around measuring students’ argumentation in both writing and talk (Sampson & Clark, 2008), as a field we need to develop valid and reliable student assessment measures for argumentation that can distinguish between levels of proficiency (Osborne, 2010).

Multiple Modalities

- One unique aspect of this work is its focus on assessing argumentation across multiple modalities.
- We believe at its most complete, evidence-based argumentation requires that students be competent on the same set of abilities: 1) comprehending and critiquing texts that present scientific arguments, 2) engaging in oral argumentation, and 3) writing arguments.
- In our project, we are designing measurement tools to assess students’ facility with the tasks of understanding, constructing, and critiquing arguments in the context of these three modalities – reading, writing and talking.

Assessment Development

- We are using the BEAR Assessment System (BAS), developed by the UC Berkeley BEAR Center (Wilson, 2005), in conjunction with elements of evidence-centered design (Miyake, Almond & Lukas, 2004) to design our argumentation assessment instruments.
- The BAS is comprised of iterative steps that include the following building blocks:

  1. **Construct Maps**
     - What is being argued?
     - What is evidence and reasoning used to justify the claim?
     - What is the counter-evidence and reasoning used to disprove the claim?
     - How is relevant evidence and reasoning used to justify the claim?

  2. **Response Model(s)**
     - How are evidence, reasoning, and rebuttal used to support the claim?

  3. **Argument Models**
     - What is the relevant evidence and reasoning?
     - How is relevant evidence and reasoning used to justify the claim?

  4. **Modeling Maps**
     - How are evidence, reasoning, and rebuttal used to support the claim?

  5. **Comparison Models**
     - What is the relevant evidence and reasoning?
     - How are evidence, reasoning, and rebuttal used to support the claim?


- Our development began with the design of construct maps. A construct map is a theoretical model of cognition that extends from high to low, illustrating qualitatively distinct groups of respondents and responses to items (Wilson, 2005).
- Our initial construct maps were developed based on an understanding of expert disciplinary knowledge and practices as well as research on student learning in the domain that informed potentially more and less sophisticated levels of the construct (Wilson, 2009).
- The next building block of assessment design focuses on how the theoretical construct can be measured with specific items or tasks designed to span one or more levels of the continuum. We rely on evidence-centered design (Miyake, Almond & Lukas, 2004) to clarify specifications for the eventual items and ensure they result in valid inferences about student abilities.

### Example Construct Map

#### Claim
- A student identifies the claim of an argument.

#### Evidence
- A student specifies evidence supporting the claim.

#### Reasoning
- A student specifies reasoning supporting the claim.

#### Rebuttal
- A student specifies rebuttal that critiques counter-claim.

Sample Reading Items

Carla and Zach are in Mr. Thomson's class. They live in Washington D.C. and felt their first earthquake in 2011. Carla knows that earthquakes can cause a lot of damage. She has a friend in California who saw an earthquake destroy many buildings in her city. Mr. Thomson asked the class: Why are some earthquakes stronger than others?

Zach wonders whether earthquakes are stronger when they occur closer to the Earth’s surface. Zach found the data in the table below.

<table>
<thead>
<tr>
<th>Location</th>
<th>Landform</th>
<th>Earthquake Depth (km)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Japan (a)</td>
<td>Island</td>
<td>&lt;5 kilometers</td>
</tr>
<tr>
<td>Japan (b)</td>
<td>Continent</td>
<td>&gt;5 kilometers</td>
</tr>
<tr>
<td>Alaska, USA</td>
<td>Continent</td>
<td>60-300 kilometers</td>
</tr>
<tr>
<td>Fiji</td>
<td>Continent</td>
<td>&gt;300 kilometers</td>
</tr>
<tr>
<td>Santiago de Chile, Argentina</td>
<td>Continent</td>
<td>577 kilometers</td>
</tr>
</tbody>
</table>

Jodie finds that the depth of an earthquake is often an important factor in determining its strength.

22. (R3a) What additional data could Zach use as evidence to help support his argument?
   a. examples of other shallow earthquakes that were very weak
   b. examples of other deep earthquakes that were very weak
   c. examples of deep earthquakes that gave their strength and depth
   d. examples of other shallow earthquakes that were very weak

29a. (R4d) If someone disagreed with Zach’s argument, what evidence might he/she use to make their case?
   a. examples of shallow earthquakes that were very strong
   b. examples of deep earthquakes that were very weak
   c. examples of other deep earthquakes that were very weak
   d. examples of other shallow earthquakes that were very weak

This research is funded by the Carnegie Corporation of NY grant B 8780.