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

Katherine L. McNeill[1], Seth Corrigan[2], Jacqueline Barber[2], Megan Goss[2] & Amanda M. Knight[1]
[1] Boston College, Chestnut Hill, MA [2] Lawrence Hall of Science, Berkeley CA

Argumentation and Assessment

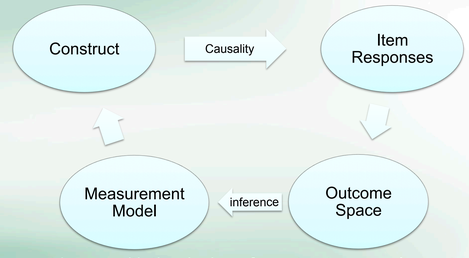
- Science proficiency includes students reading and writing complex informational text as well as engaging in written and oral argumentation (Pearson et. al., 2010).
- The educational importance of argumentation is reflected in recent science education research (e.g. McNeill et. al., 2006; Duschl, Schweingruber, & Shouse, 2007; Osborne, 2010; Sampson & Clark, 2008), the new K-12 Science Education Frameworks (NRC, 2011) 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 (Atkin, 2002).
- Large-scale assessments can impact teachers' pedagogical approaches to 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 in a complex 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; Wilson, 2009), in conjunction with elements of evidence-centered design (Mislevy, Almond & Lukas, 2004) to design our argumentation assessment instruments.
- The BAS is comprised of iterative steps that include four building blocks:



- 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 informs 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 (Mislevy, Almond, Lukas, 2004) to clarify specifications for the eventual items and ensure they result in valid inferences about student abilities.

Construct Maps

	Reception (Reading)			Expression (Writing/Talk)		
1. Claim: What is being argued?	R1	**What is being argued?** R1. Student identifies the claim of an argument.		WT1	**What is being argued?** WT1. Student constructs the claim of an argument.	
2. Justification: How is the claim supported?				WT2a	**How is the claim justified?** WT2a. Student constructs an argument that includes justifications for the claim other than relevant E & R (i.e. irrelevant or inaccurate data or science ideas, appeal to authority or expertise, etc.).	
	R2b	**What is the evidence?** R2b1. Student identifies the evidence of an argument.	**What is the reasoning?** R2b2. Student identifies the reasoning of an argument.	WT2b	**How is evidence used to justify the claim?** WT2b1. Student constructs a C-E argument that includes some irrelevant date and/ or inaccurate data as well as some relevant scientific evidence.	**How is reasoning used to justify the claim?** WT2b2. Student constructs a C-R argument that includes some irrelevant science ideas and/or inaccurate science ideas as well as some relevant reasoning.
	R2c	**What is the evidence AND reasoning?** R2c. Student identifies the evidence AND reasoning of an argument.		WT2c	**How are evidence and reasoning used to justify the claim?** WT2c. Student constructs a C-E-R argument that uses some irrelevant or inaccurate data and science ideas as well as some relevant E & R. This irrelevance is in both.	
	R2d			WT2d	**How are evidence, reasoning and rebuttal used to justify the claim?** WT2d. Student constructs a C-E-R argument that uses some irrelevant or inaccurate data and science ideas and some relevant E & R. In addition, the student provides a rebuttal that uses either or both relevant E & R, and may include irrelevant/inaccurate E & R.	
3. Relevant Justification: Is the justification relevant?	R3a	**What evidence is relevant?** R3a1. Student distinguishes relevant evidence from irrelevant data in a C-E-R argument.	**What reasoning is relevant?** R3a2. Student distinguishes relevant reasoning from irrelevant scientific ideas in a C-E-R argument.	WT3a	**How is relevant evidence used to justify the claim?** WT3a1. Student constructs a C-E or C-E-R argument that includes only relevant evidence.	**How is relevant reasoning used to justify the claim?** WT3a2. Student constructs a C-R or C-E-R argument that includes only relevant reasoning.
	R3b	**What evidence AND reasoning is relevant?** R3b. Student distinguishes relevant evidence from irrelevant data AND relevant reasoning from irrelevant scientific ideas in a C-E-R argument.		WT3b	**How is relevant evidence and reasoning used to justify the claim?** WT3b. Student constructs a C-E-R argument that includes only relevant evidence AND reasoning.	
4. Relevant Counter-Justification: Is the counter-justification relevant?	R4a	**Does the counter-evidence contradict the argument?** R4a1. Student identifies whether counter-evidence contradicts a target argument.	**Does the counter-reasoning contradict the argument?** R4a2. Student identifies whether counter-reasoning contradicts a target argument.			
	R4b	**What is the relevant counter-evidence?** R4b1. Student distinguishes relevant counter-evidence from irrelevant data.	**What is the relevant counter-reasoning?** R4b2. Student distinguishes relevant counter-reasoning from irrelevant science ideas.			
	R4c	**How is the counter-evidence relevant?** R4c1. Student evaluates a target C-E-R argument using relevant counter-evidence.	**How is the counter-reasoning relevant?** R4c2. Student evaluates a target C-E-R argument using relevant counter-reasoning.	WT4c	**Why is the counter-evidence less relevant?** WT4c1. Student constructs a C-E-R argument that includes an E-rebuttal that critiques the relevancy of the counter-evidence.	**Why is the counter-reasoning less relevant?** WT4c2. Student constructs a C-E-R argument that includes a R-rebuttal that critiques the relevancy of the counter-reasoning.
	R4d	**How is the counter-evidence AND counter-reasoning relevant?** R4d. Student evaluates a target C-E-R argument using counter-evidence AND counter-reasoning.		WT4d	**Why are both the counter-evidence AND counter-reasoning less relevant?** WT4d. Student constructs a C-E-R argument that includes an E-R-rebuttal that critiques the relevancy of the counter-evidence AND counter-reasoning. All of the E & R must be relevant.	
5. Comparing Arguments: How does the justification in argument 2 strengthen or weaken argument 1's claim?	R5a	**How does the evidence in argument 2 strengthen or weaken argument 1's claim?** R5a1. Student evaluates how the evidence in argument 2 strengthens or weakens argument 1's claim or vice versa.	**How does the reasoning in argument 2 strengthen or weaken argument 1's claim?** R5a2. Student evaluates how the reasoning in argument 2 strengthens or weakens argument 1's claim or vice versa.			
	R5b	**How does the evidence and reasoning in argument 2 strengthen or weaken argument 1's claim?** R5b. Student evaluates how the evidence and reasoning in argument 2 strengthens or weakens argument 1's claim or vice versa.				
6.Justifying Across Arguments: What is the justified decision after evaluating opposing arguments?	R/WT6a	**What is the justified decision?** RWT6a. Student evaluates 2 opposing C-E-R arguments through the construction of an argument that justifies with evidence, reasoning, AND a rebuttal that critiques either counter-evidence OR counter-reasoning (i.e. the student reads or listens to 2 opposing C-E-R arguments and then constructs their own C-E-R + critique of counter-E OR counter-R argument). All the E & R must be relevant.				
	R/WT6b	**What is the justified decision?** RWT6b. Student evaluates 2 opposing C-E-R arguments through the construction of an argument that justifies with evidence, reasoning, AND a rebuttal that critiques either counter-evidence AND counter-reasoning (i.e. the student reads or listens to 2 opposing C-E-R arguments and then constructs their own C-E-R + rebuttal that critiques counter-E AND counter-R argument). All the E & R must be relevant.				

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.

Location	Landform	Earthquake Depth		Earthquake Strength	Perceived Shaking
Japan (a)	Island	Shallow < 60 kilometers	32 kilometers	12	Cataclysmic
Japan (b)	Island		32 kilometers	11	Extreme
Alaska, USA	Continent	Intermediate 60-300 kilometers	62 kilometers	11	
Pakistan	Continent		83 kilometers	11	
Santiago del Estero, Argentina	Continent	Deep >300 kilometers	577 kilometers	10	Intense

Zach's Argument:

Earthquakes are stronger when they are shallow, occurring closer to the Earth's surface. The shallow earthquakes listed in the data table only had a depth of 32-kilometers. One of these shallow earthquakes had a strength of 12, which means "cataclysmic" shaking. The other had a strength of 11, which means "extreme" shaking. Both shallow earthquakes occurred in Japan, which is an island. The middle earthquakes started 62-kilometers and 83-kilometers below the earth's surface. These middle earthquakes both had a strength of 11, which was "extreme" shaking. However, the deepest earthquake had a depth of 576.8-kilometers, and it only had a strength of 10. This one had "intense" shaking. The middle and the deepest earthquakes started on continents. The reason the strongest earthquakes were stronger is because they both occurred closer to the Earth's surface. Deeper earthquakes are weaker because they are further from the Earth's surface.

22. (R3a) What additional data could Zach use as evidence to help support his argument

a. examples of other shallow earthquakes that were very strong
b. examples of other earthquakes that give their strength and depth
c. examples of other deep earthquakes that were very strong
d. examples of other shallow earthquakes that were very weak

29a. (R4c) 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 deep earthquakes that were very strong
d. examples of earthquakes that were strongest where they started

29b. (R4d) Why or how would this data weaken the argument of Scientist 2?

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