Assessing Scientific Practices: Issues and Challenges Drawn from the Example of Argumentation

J. Bryan Henderson

ASU
Mary Lou Fulton
Teachers College
ARIZONA STATE UNIVERSITY

Jonathan F. Osborne

Stanford
GRADUATE SCHOOL OF
EDUCATION

Learning Science

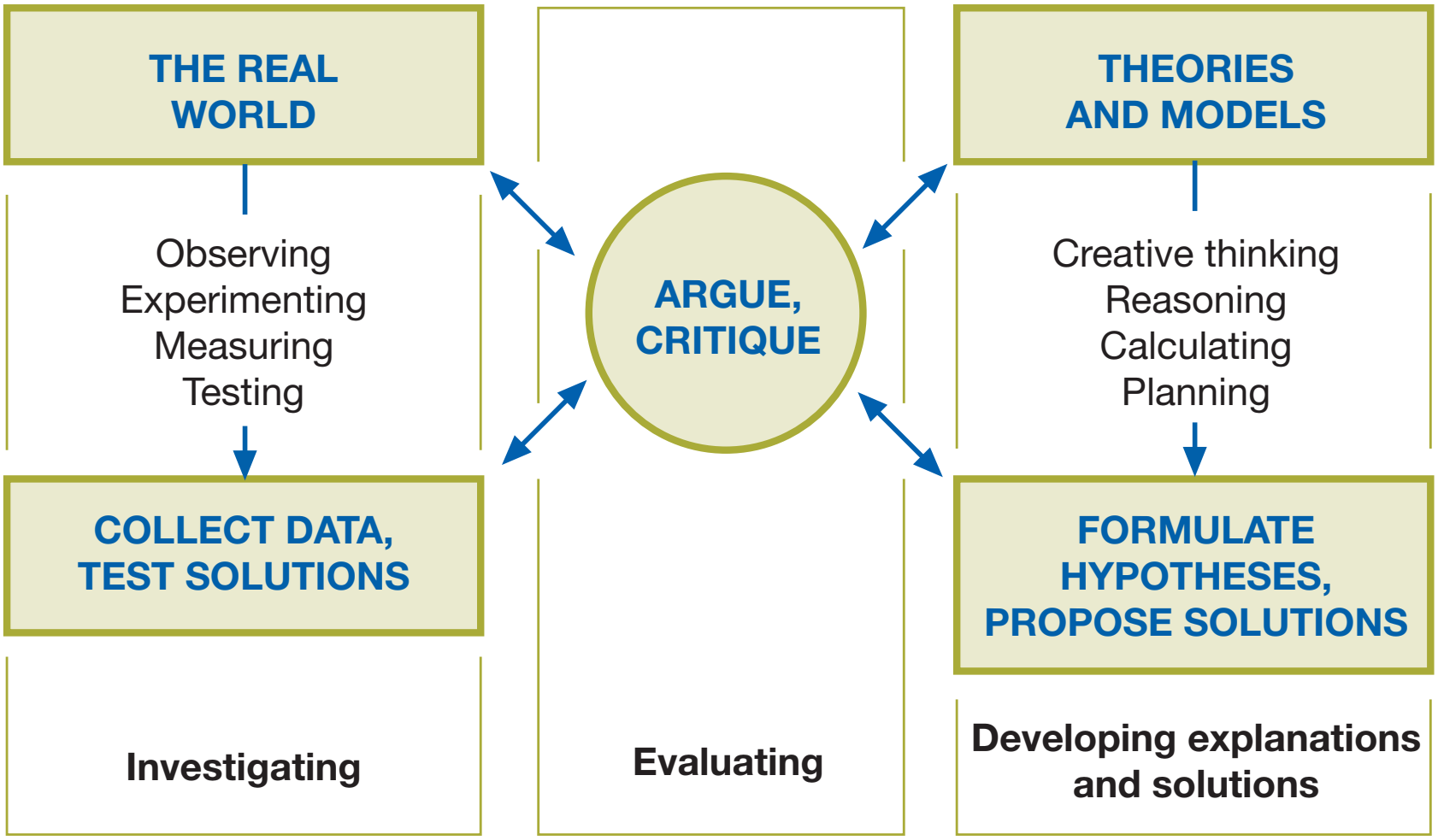
Construction

AND

Critique

Knowing why you are wrong matters as much as knowing why you are right!

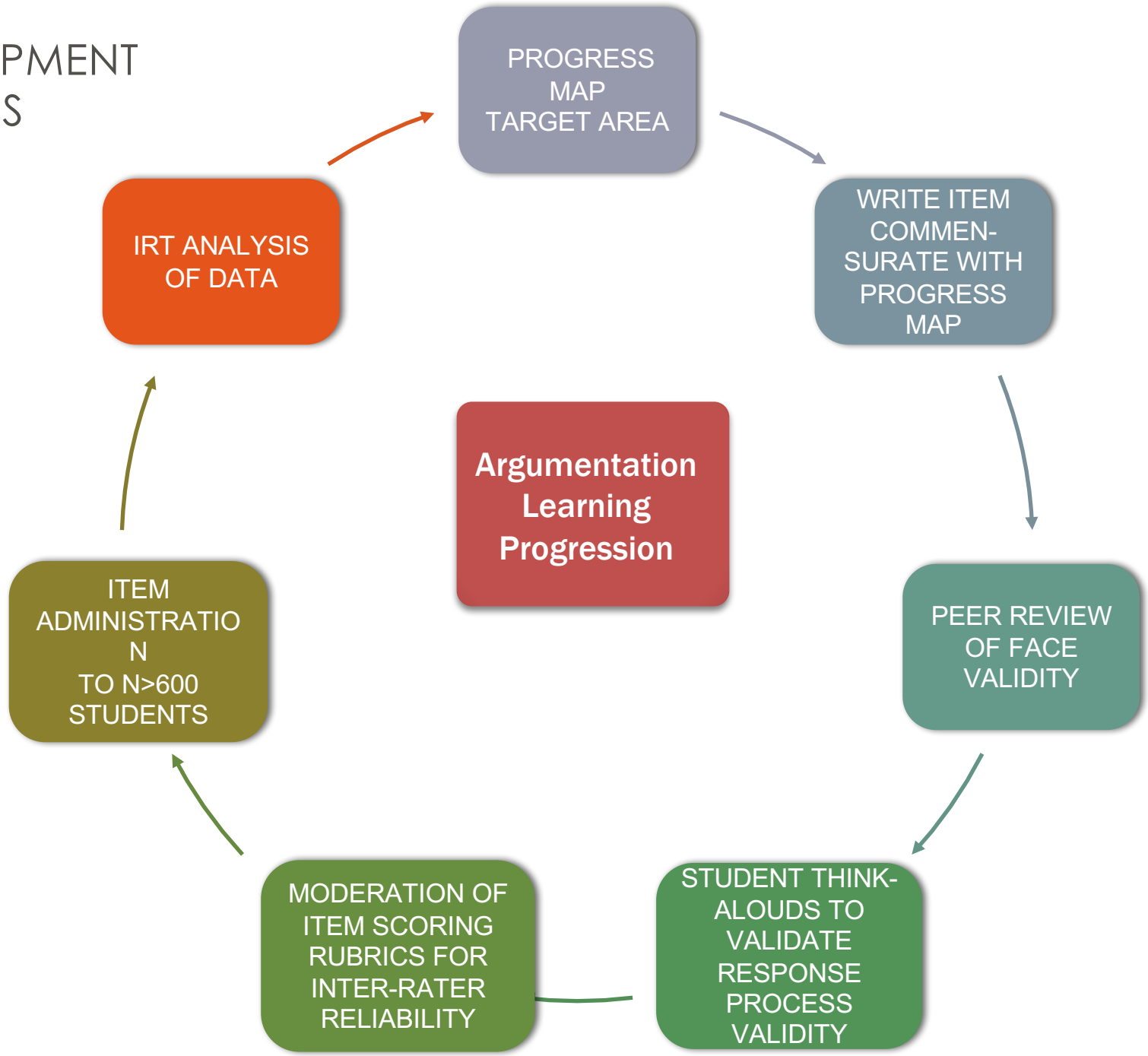
The Role of Argument in Science



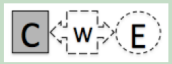
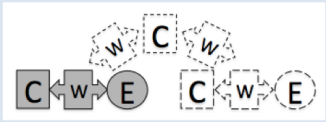
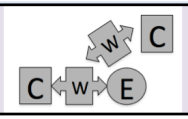
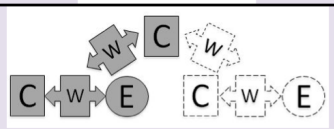
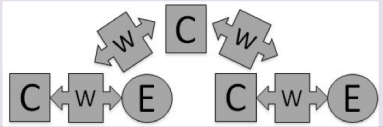
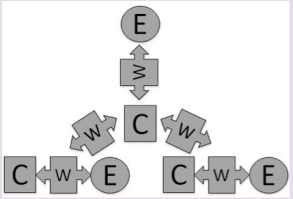
What does it mean to assess the practice of argumentation?

What might it mean to progress in the practice of argumentation?

ITEM DEVELOPMENT PROCESS

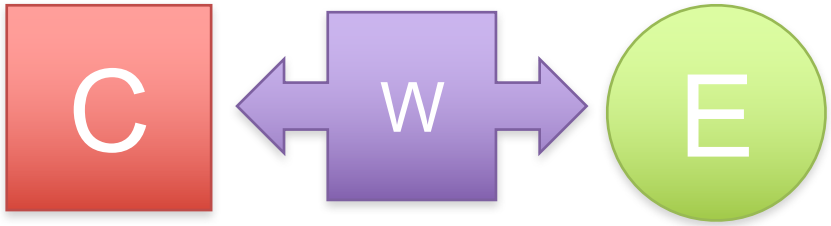


A LEARNING PROGRESSION FOR ARGUMENTATION

Level	Constructing	Critiquing	Description	Representation of elements
0			No evidence of facility with argumentation.	
0a	**Stating a claim**		Student states a relevant claim.	
0b		***Identifying* a claim**	Student identifies another person's claim.	
0c	**Providing evidence supporting a claim**		Student supports a claim with a piece of evidence.	
1a	**Constructing a warrant that links claim and evidence**		Student constructs an explicit warrant that links their claim to evidence.	
1b		**Identifying a warrant**	Student identifies the warrant provided by another person.	
1c	**Constructing a complete argument.**		Student makes a claim, selects evidence that supports that claim, and constructs a synthesis between the claim and the warrant.	
1d		**Providing an alternative counter argument**	Student offers a counterargument as a way of rebutting another person's claim.	
2a		**Providing a counter-critique**	Student critiques another's argument. Fully explicates the claim that the argument is flawed and *justification* for why that argument is flawed.	
2b	**Constructing a one-sided comparative argument**		Student makes an evaluative judgment about the merits of two competing arguments and makes an explicit argument for the value of *one* argument. No warrant for why the other argument is weaker.	
2d	**Providing a two-sided comparative argument**		Student makes an evaluative judgement about two competing arguments and makes an explicit argument (claim + justification) for why one argument is stronger and why the other is weaker (claim + justification).	
2e	**Constructing a counter claim with justification**		This progress level marks the top anchor of our progress map. Student explicitly compares and contrasts two competing arguments, and also constructs a new argument in which they can explicitly justify why it is superior to each of the previous arguments.	

THE TOULMIN ARGUMENTATION MODEL

CLAIM **EVIDENCE**

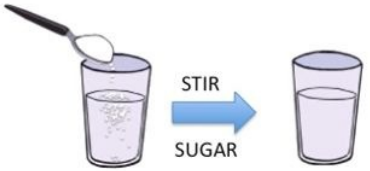


WARRANT

I agree with Mary because the weight is the same and the sugar would have nowhere to go.

Two students pour sugar grains into a glass of hot water. They make three observations:

1. Once the sugar is poured into the water, it is stirred. After stirring, the sugar can no longer be seen.



2. Also after stirring, each student tastes the water. They both agree that the water tastes sweet.

3. The weight of the water glass and the sugar before it was added to the water is the same as the weight of the water glass after the sugar was stirred in.

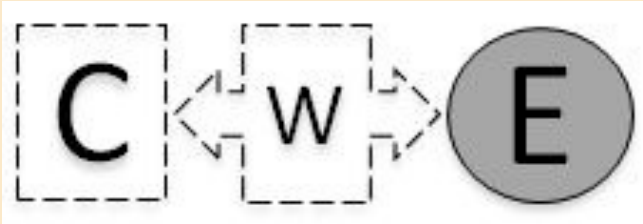
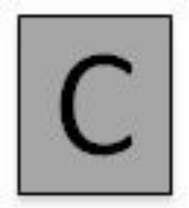
Their teacher asks if they think sugar remains in the water.

Laura says: *I think the sugar is gone.*

Mary says: *I think the sugar is still there.*

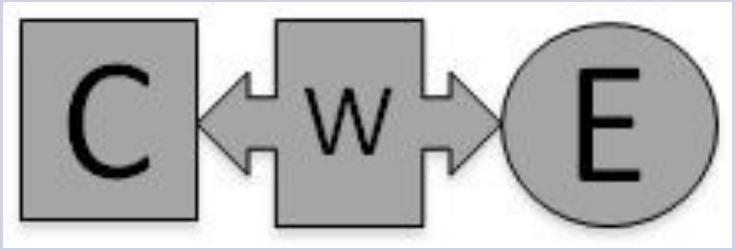
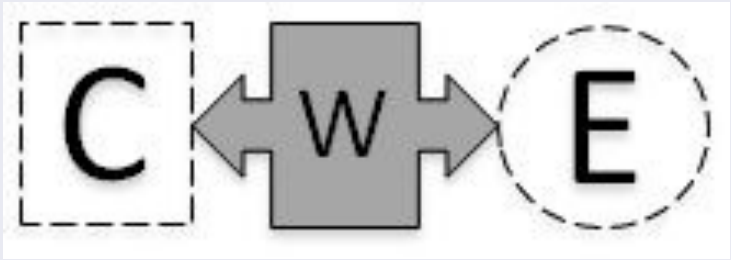
LEVEL 0: Claim and evidence are the fundamental building blocks of argument.

PROGRESS LEVEL 0

DESCRIPTION	INTRINSIC COGNITIVE LOAD
Stating/identifying an explicit and relevant piece of EVIDENCE	
Stating/identifying an explicit and relevant CLAIM	
No explicit understanding of CLAIMS and/or EVIDENCE	

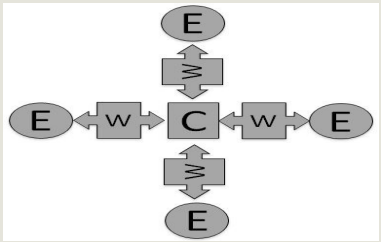
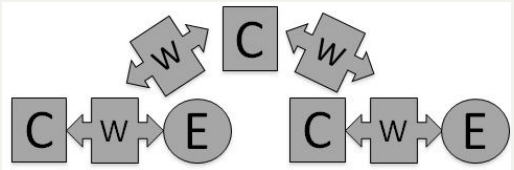
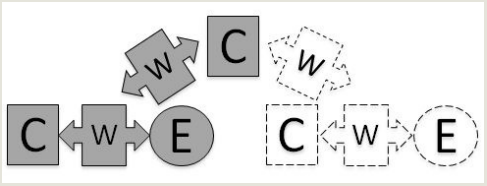
LEVEL 1: Claim and evidence must now be coordinated with an explicit warrant.

PROGRESS LEVEL 1

DESCRIPTION	INTRINSIC COGNITIVE LOAD
Constructing/Critiquing an explicit and relevant ARGUMENT or REBUTTAL	
Stating/identifying an explicit and relevant WARRANT	

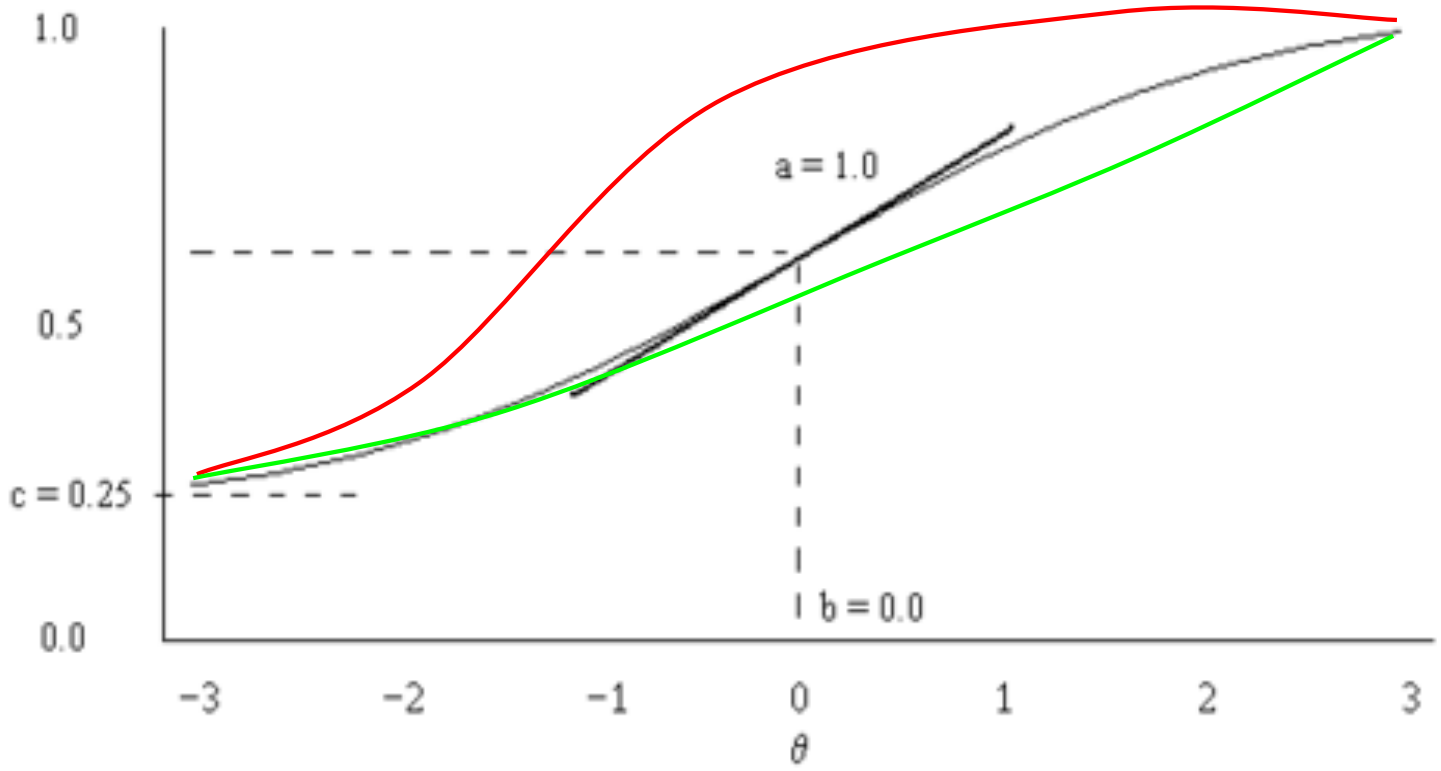
LEVEL 2+: TWO OR MORE EXPLICIT WARRANTS REQUIRED TO COORDINATE INCREASING NUMBERS OF CLAIMS AND EVIDENCE.

PROGRESS LEVEL 2

DESCRIPTION	INTRINSIC COGNITIVE LOAD
Comparing relative SIGNIFICANCE OF MULTIPLE PIECES OF EVIDENCE	
Constructing/ Critiquing a <u>TWO</u>-sided COMPARATIVE ARGUMENT	
Constructing/Critiquing a <u>ONE</u>-sided COMPARATIVE ARGUMENT	

The *Item Response Function* (IRF)

The IRF gives the probability that a person with a given ability level (theta) will answer correctly.

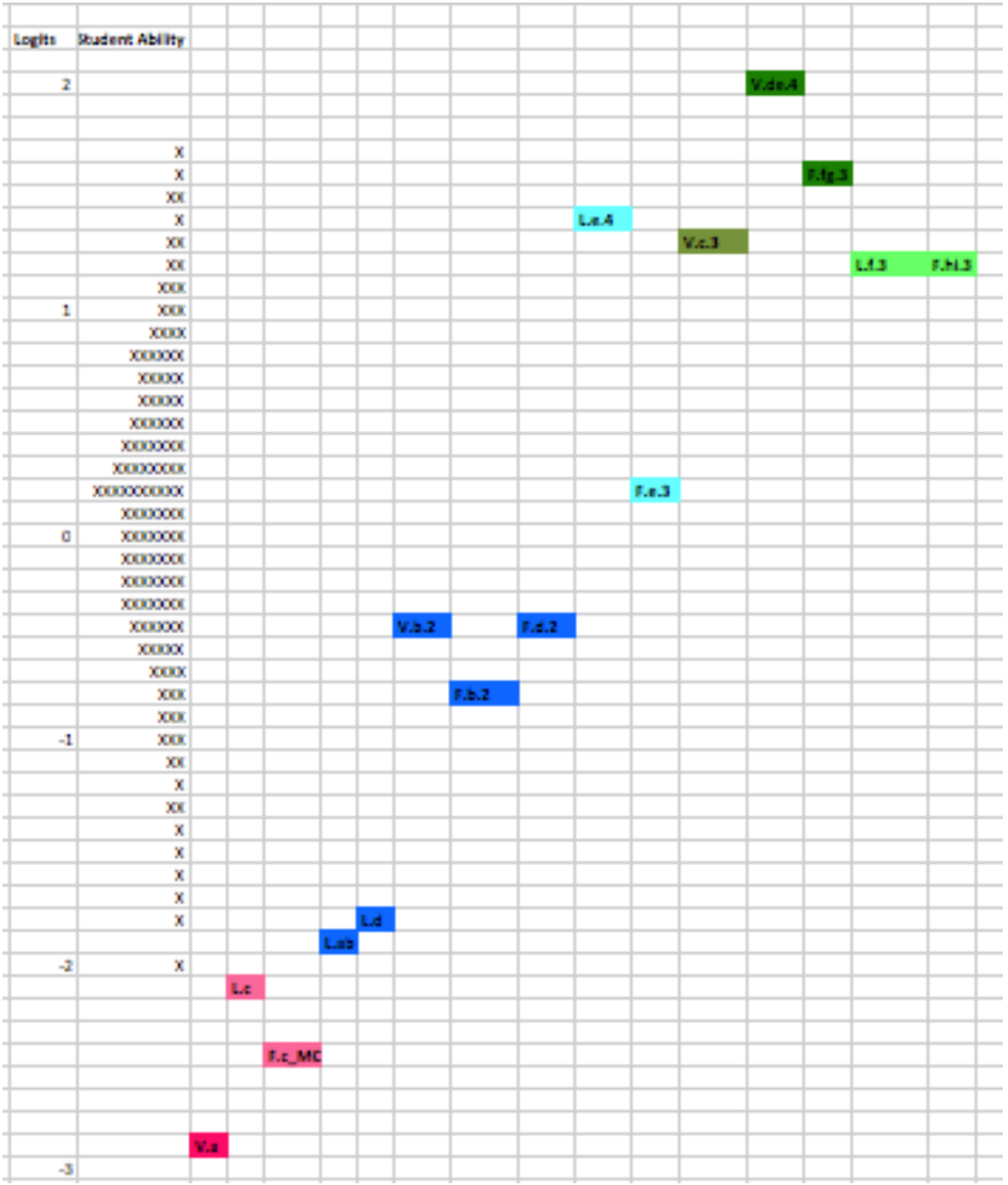


b = difficulty parameter (location along x-axis of max slope)

a = discrimination parameter (max slope)

c = guessing parameter (success by chance - y-intercept)

ITEM RESPONSE THEORY ANALYSIS



Harder

Easier

LEVEL 2

LEVEL 1

LEVEL 0

ITEM
RESPONSE
THEORY
ANALYSIS

Average difficulty of items testing scientific argumentation levels

Argumentation level	Average item difficulty
Level 0	-1.39 (Range: -2.59 – 0.31)
Level 1	0.46 (Range: -0.62 – 1.04)
Level 2	1.10 (Range: 0.91 – 1.56)

Findings

- Empirically-supported progress map for argumentation (Osborne, Henderson, MacPherson, Szu, & Wild, 2016).
- Greater student propensity to argue in the affirmative – CRITIQUE IS CHALLENGING!
- More students were able to critique with PROMPTING.

Challenges and Issues

- Given the absence of a language to define or assess progression, teachers will fall back on the familiar content-based objectives.
- This work offers a model of how student competency within a science domain progresses and ways in which it can be assessed.

Challenges and Issues

- However, critical considerations remain for researchers moving forward with attempts to develop sufficient notions of progression with each of the NGSS practices. These include:
  - To what extent is our argumentation learning progression generalizable across different domains?
  - How will teachers interpret the NGSS shift from knowledge-based to practice-based assessment? How can teachers be supported to use these assessments for the sake of student learning, as opposed to merely accountability?
  - Are open-ended assessments amenable to large-scale accountability testing?

Promise for automated scoring?

Predicted score

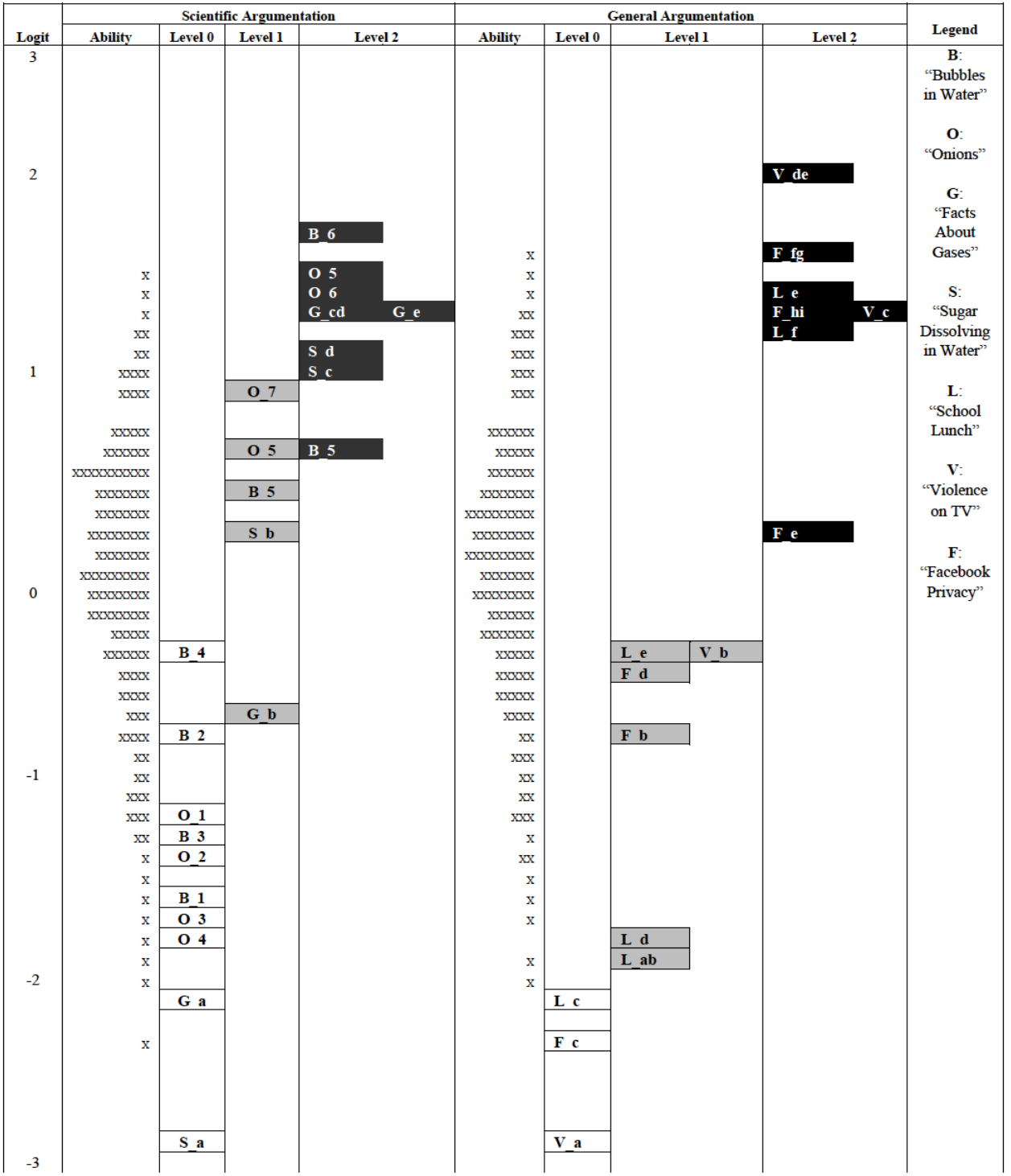
Actual score

Act \ Pred	1	2
1	20	17
2	6	214

Cohen's kappa = 0.60

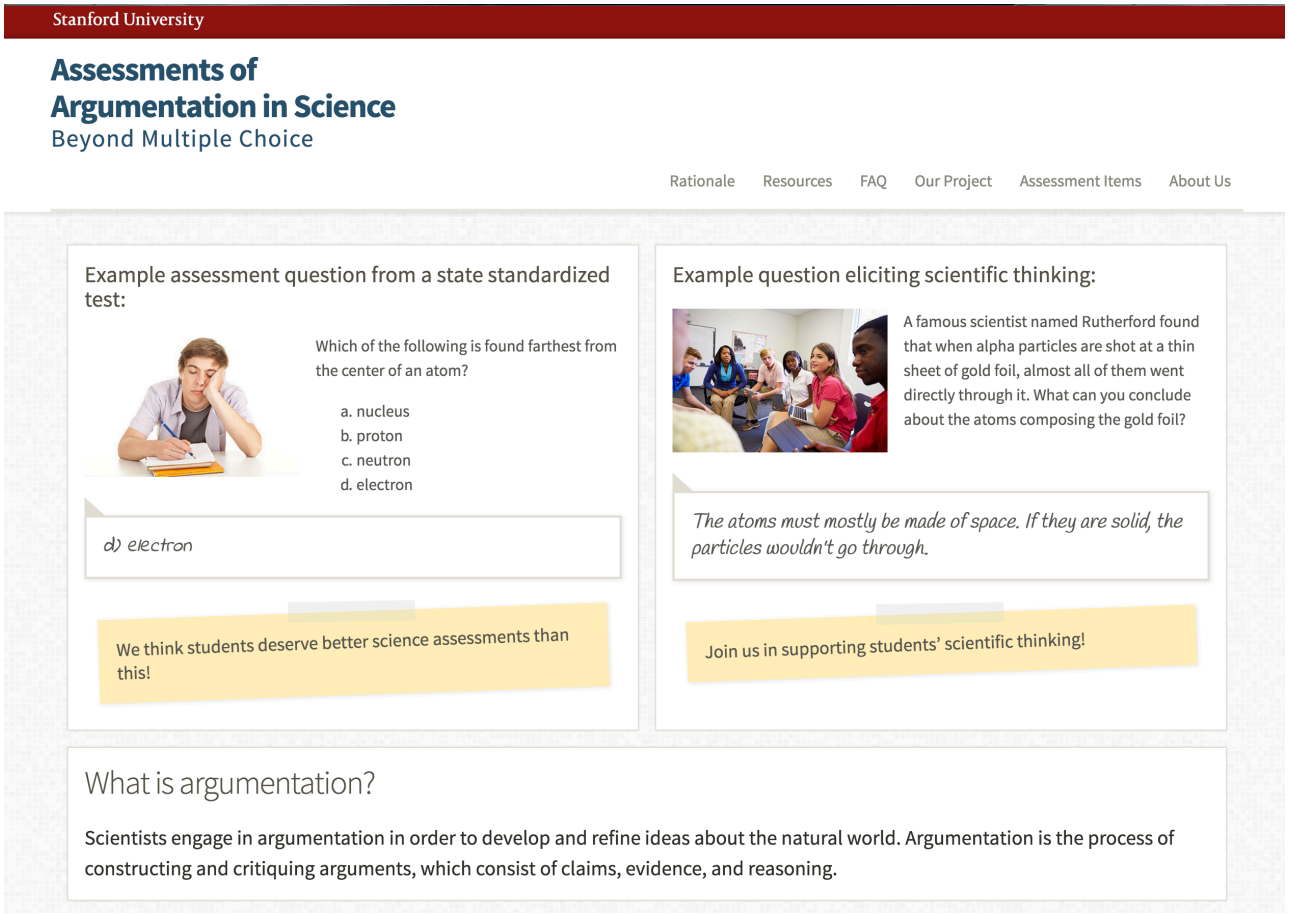
23 out of 257 open-ended responses scored I NCORRECTLY (less than 9% error rate)

Domain generality?

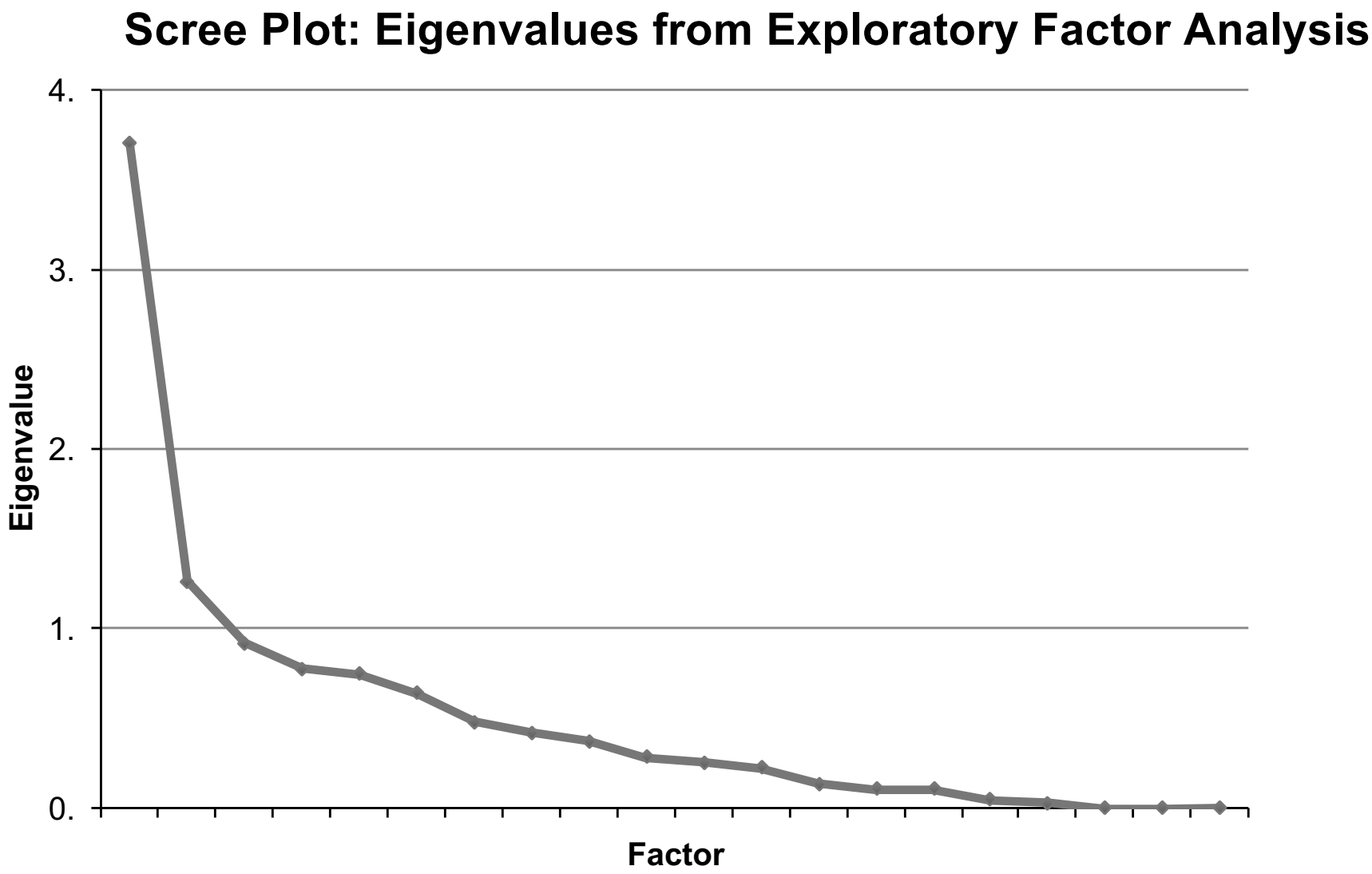


Sharing our Assessment Tasks

scientificargumentation.stanford.edu



ASSESSING UNIDIMENSIONALITY



Yao & Wilson, 2013