DESCRIBE THE MAJOR FINDINGS RESULTING FROM THESE ACTIVITIES.

Research Activity 1: Develop and Validate an Assessment Framework

We used the *ConQuest* software (Wu, Adams, & Wilson, 1997) to conduct a Rasch analysis based on the Partial Credit Model. ConQuest provides an ability estimate for each student and an item difficulty estimate for each item. Both ability and item estimates are calibrated to be on the same logit scale with values ranging from -4.0 to 4.0. The higher the logit value, the more able the student and the more difficult the item.

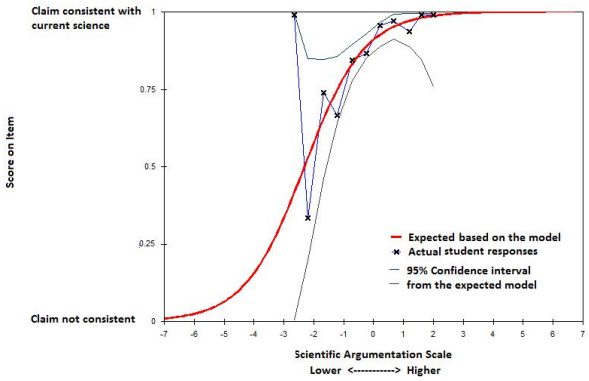
Item fit.

Table 1 shows item fit statistics in mean square values. The acceptable range for item fit to the Rasch Partial Credit Model is between 0.70 and 1.30 (Bond & Fox, 2007). There were no misfit items based on infit and outfit statistics. *According to these results, students' responses to all four types of items could be interpreted on the overall scientific argumentation scale.* Figure 1 shows how well students' actual responses to the Life item set fit the Rasch Partial Credit Model. In all figures, the x-axis indicates students' scientific argumentation abilities from low (-7.0) to high (7.0). The y-axis represents students' scores on the item. The Rasch Partial Credit Model represents a monotonically increasing relationship between student ability and student score on the item. That is, students are more likely to receive higher scores on the item as their underlying scientific argumentation abilities increase. Students' responses to justifications, uncertainty, and conditions of rebuttal items in the Life item set closely map onto the model lines. In the claim item, this monotonically increasing relationship holds except the very low ability students who picked the scientifically correct claim

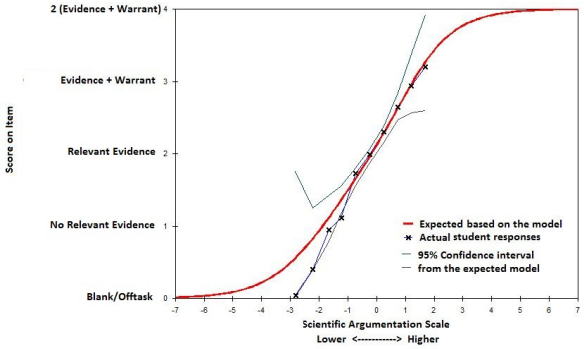
	Item	Infit		Outfit	
Items	difficulty	mean square	error	mean square	error
Claims					
Pinatubo	-0.57	1.03	0.07	1.03	0.07
T2050	0.87	0.97	0.08	0.95	0.08
Ocean	1.16	1.03	0.09	1.10	0.09
Galaxy	-1.15	1.07	0.08	1.09	0.08
Life	-2.24	0.98	0.11	0.93	0.11
Spectra	0.20	1.00	0.08	1.00	0.08
mean item difficulty =	- 0.29				
Justifications					
Pinatubo	0.23	0.95	0.06	0.94	0.06
T2050	0.65	0.93	0.05	0.91	0.05
Ocean	0.10	0.94	0.04	0.94	0.04
Galaxy	0.01	0.97	0.05	0.97	0.05
Life	-0.30	0.95	0.04	0.95	0.04
Spectra	0.73	0.94	0.04	0.93	0.04

mean item difficulty =	0.24				
(c) Uncertainty qualifiers					
Pinatubo	-1.42	0.96	0.06	0.96	0.06
T2050	0.24	1.08	0.05	1.13	0.05
Ocean	-1.00	0.99	0.05	0.99	0.05
Galaxy	-1.38	1.08	0.06	1.18	0.06
Life	-1.29	0.97	0.06	0.97	0.06
Spectra	-0.07	1.13	0.05	1.16	0.05
mean item difficulty =	- 0.82				
(d) Conditions of rebuttals					
Pinatubo	0.89	1.04	0.05	1.05	0.05
T2050	0.88	0.95	0.06	0.95	0.06
Ocean	1.10	0.98	0.05	0.97	0.05
Galaxy	0.57	1.05	0.04	1.04	0.04
Life	0.72	1.04	0.04	1.06	0.04
Spectra	1.07	0.97	0.06	0.98	0.06
mean item difficulty =	0.87				

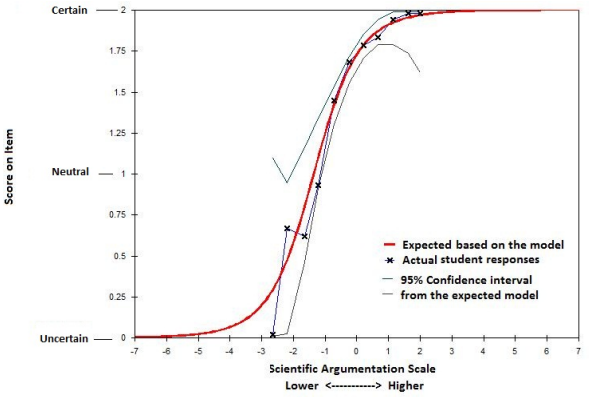
Table 1. Rasch Partial Credit Model Analysis Results.



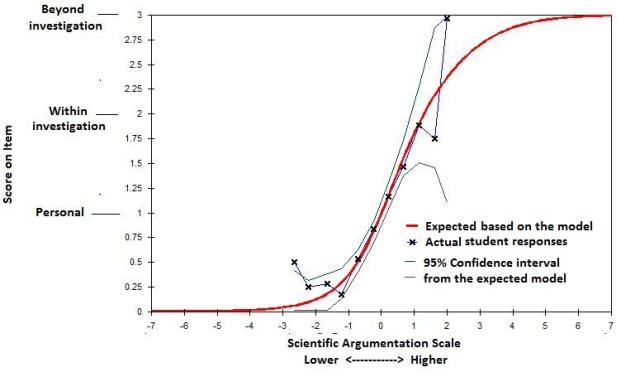
Life-Claim



Life-Justification



Life-Uncertainty



Life-Conditions of rebuttal

Figure 1. Item characteristic curves for the Life item set.

Rasch Scale for the Scientific Argumentation Construct

We examined how difficult each item was on the scientific argumentation scale. Table 1 shows that the easiest item on the scale was the claim item in the Life item set with the item difficulty value of -2.24. This means that students whose scientific argumentaiton ability were at -2.24 had a 50% chance of answering this item correctly. The most difficult item was the claim item in the Ocean item set with the item difficulty value of 1.16. We then compared average item difficulty values across claim, justification, uncertainty, and conditions of rebuttal items. The easiest item group was uncertainty items, followed by claim item group. The most difficult item group was conditions of rebuttal. Justification items were placed between claims and conditions of rebuttal. See Table 1. *These results indicate that the order of the required ability on the scientific argumentation scale was uncertainty → claim → justification → conditions of rebuttal, instead of the hypothesized order of claim → justificaiton → uncertainty → conditions of rebuttal.*

Figure 2 shows how items and students distributed on the scientific argumentation scale expressed in logit values from -4.0 to +4.0. On the left side, the distribution of students according to their scientific argumentation ability is shown. *The higher on the scale, the more able students are on the scientific argumentation construct.* On the right side, item thresholds of all scores in claim, justification, uncertainty, and conditions of rebuttal items are shown. An item threshold is defined as students with the matching ability would have a 50% chance of receiving a score j as compared to receiving a score $j - 1$.

The locations of these score bands across four types of items indicate that justification items covered the widest range of the scientific argumentation ability scale between -3.60 to +3.80. Conditions of rebuttal items covered the range of -1.35 to +3.10. The range covered by claim items was smaller than those covered by justification and conditions of rebuttal items but slightly larger than the range covered by the uncertainty items. Both uncertainty and claim items covered the middle ability range of the scientific argumentation scale.

The score band of making single warrants was located at the similar range to that of considering conditions of rebuttal within investigation. The score band of making two or more warrants was located at a similar range to that of conditions of rebuttal beyond investigation. *These findings suggest that students who could make single warrants were more likely to consider conditions of rebuttal within investigation. Students who could make multiple warrants were more likely to consider conditions of rebuttal beyond investigation, indicating that students need to make multiple warrants based on multiple evidence pieces in order to consider limitations of the investigations imposed by current science, inquiry method, or other factors.*

The scientific argumentation scale shown in Figure 2, had the person separation reliability of 0.77 and the item separation reliability of 1.00.

Research Activity 2: Characterize Uncertainty

Unlike scientific uncertainty, in which the limitations reflect the current status of scientific knowledge and investigation methods commonly adopted by the community of scientists, student uncertainty can also reflect the student's assessment of his own status of knowledge, ability, and skill. Table 2 shows four types of student uncertainty identified in students' argumentation items. We used the numerical scores associated with these four types in the analysis of students' overall scientific argumentation ability.

Our preliminary results indicate that student uncertainty in scientific arguments appears to be characteristically different from scientists' uncertainty in the sense that the latter addresses limitations of collective knowledge, understanding, and tools of science while the former primarily addresses self-concepts before it is transitioned to scientific uncertainty.

Source of Uncertainty	Uncertainty source	Description of categories
No Information (Score 0)	• No response	• Did not respond to the related uncertainty item but answered the linked claim and explanation items.
	• Simple off-task responses	• Wrote "I do not know" or similar answers • Provided off-task answers
	• Restatement	• Restated the scientific claim made in the claim item. • Restated the uncertainty rating.
Personal (Score 1)	• Question	• Did/did not understand the question.
	• General knowledge/ability	• Did/did not possess general knowledge or ability necessary in solving the question. • Did/did not learn the topic (without mentioning the specific topic) • Can/cannot explain/estimate
	• Lack of specific knowledge/ability	• Did not know specific scientific knowledge needed in the item set.
	• Difficulty with data	• Did not make sense of data provided in the item..
	• Authority	• Mentioned teacher, textbook, and other authoritative sources.
Scientific-Within investigation (Score 2)	• Specific knowledge	• Referred to/elaborated a particular piece of scientific knowledge directly related to the item.
	• Specific data	• Referred to a particular piece of scientific data provided in the item.
Scientific-Beyond investigation (Score 3)	• Data/investigation	• Recognized the limitation of data provided in the item and suggested a need for additional data. • Mentioned that not all factors are considered.
	• Phenomenon	• Elaborated why the scientific phenomenon addressed in the item is uncertain.
	• Current science	• Mentioned that current scientific knowledge or data collection tools are limited to address the scientific phenomenon in the item.

Table 2. Students' uncertainty types

Research Activity 3: Document Student Learning

We are in the process of scoring students' responses to the pre- and post-tests and embedded assessments for the "Modeling Earth's Climate" investigation and the "Is there life in space?" investigation. We are using scoring rubrics similar to what we used in the early-year test. We will follow students' learning and argumentation skills across four time points, early year, before, during, and after the High-Adventure Science investigations, to develop students learning trajectories. We will be looking for answers to research questions about students learning gains as well as looking at evidence for change in student's scientific reasoning skills. It is our goal to focus on helping students' reason about frontier science, what is known and unknown, what are the limitations of the data and models, and what scientists grapple with when exploring open-ended questions. We will write a research paper using the research and the data analysis. We will also revise the curriculum and assessment as a result of the results from data analysis.

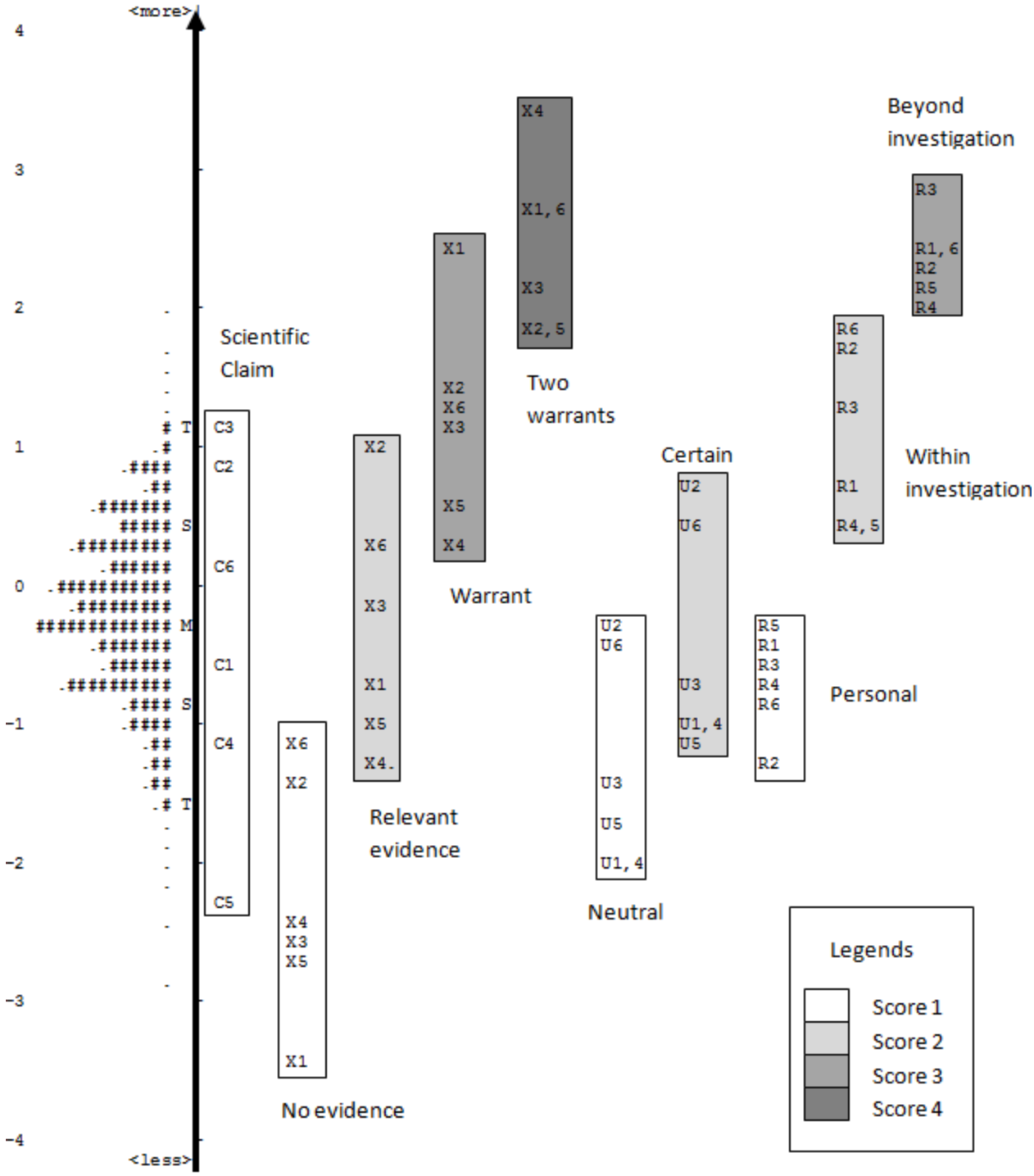


Figure 2. Wright Map for the scientific argumentation construct. Note. "C" = Claim; "J" = Justification; "U" = Uncertainty; and "R" = Conditions of rebuttal; "1" Pinatubo Item Set; "2" T2050 Item Set; "3" Ocean Item Set; "4" Galaxy Item Set; "5" Life Item Set; "6" Spectra Item Set; "#" represents 7 students.

References

Bond, T. G., & Fox, C. M. (2007). Applying the Rasch model. Mahwah, NJ: Lawrence Erlbaum Associates.