Promoting Scientific Argumentation with Computational Models

Amy Pallant[1]

Hee-Sun Lee[2]

Sarah Pryputniewicz[1]

[1] The Concord Consortium, Presenter

[2] University of California, Santa Cruz

April 9, 2013. 1:00 - 2:30 pm. Strand 12: Educational Technology Related Paper Set

Pallant, A., Lee, H. –S., & Pryputniewicz, S. (2013). Promoting scientific argumentation with computational models. Paper presented at the annual meeting of the National Association for Research in Science Teaching, Rio Grande, Puerto Rico.

Acknowledgements
This material is based upon work supported by the National Science Foundation under grant No. 0929774. Any opinions, findings, conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation. The authors gratefully acknowledge students and teachers who participated in this study.

Introduction

Scientific argumentation involves both scientific reasoning to draw inferences from initially available information (Holyoak & Morrison, 2005) and critical thinking to sort out evidence for making conditional claims (Yeh, 2001). Toulmin (1958) characterized the structure of written and dialogic arguments in six elements: claim, data, warrant, backing, modal qualifier, and conditions of rebuttal. Current science education literature on scientific argumentation has focused mostly on scientific reasoning dealing with claim, data, warrants, and backing (Sampson & Clark, 2008). In our study, we investigated not only scientific reasoning but also critical thinking shown in students' choice of uncertainty modal qualifiers and rationale for conditions of rebuttal in their rhetorical scientific arguments.

Since students are not naturally inclined to consider uncertainty and their rationale for uncertainty in their scientific arguments (Chin & Osborne, 2010), we developed online curriculum modules where consideration of uncertainty and uncertainty rationale are scientifically necessary and pedagogically scaffolded with prompts. For curriculum contexts, we chose three from the 125 important questions currently pursued by scientists (Kennedy & Norman, 2005) such as "How hot will the greenhouse world be?", "Is there-or was there-life elsewhere in the solar system?", and "Is there enough fresh water?" Since these questions cannot be investigated hands-on, we used scientists' data and personal accounts and a series of computational models for students to explore. In supporting students' scientific argumentation, we created a four-step prompts: multiple-choice claims, open-ended justifications to support claims, 5 point Likert-scale uncertainty ratings from not very certain to very certain, and open-ended explanations for their uncertainty ratings. We used these prompts in assessing students' scientific argumentation abilities as well as in curricular activities where students are engaged with scientific inquiry based on complex data sets and computational models.

In this paper, we report on our recent study where we investigated the following research questions:

- How do students' scientific argumentation performances change before and after three HAS modules? How consistent are student performance changes across teachers? How are students' performance changes correlated with their gender, technological experience, and ELL (English Language Learner) status?
- How does students' scientific argumentation progress throughout the year across teachers?

Curriculum Design

As part of an NSF-funded curriculum project called High Adventure Science (HAS), we developed three online curriculum modules related to climate change, fresh water sustainability, and life on other planets. These modules were designed based on the following design principles.

Design Principles

Principle 1: Use open-ended, authentic, frontier science topics to frame the modules. The use of contexts authentic to current practices is a powerful way to increase student motivation, engagement, and learning. Authentic science is not always accessible to secondary students or able to be linked to learning goals at this level due to lack of students' knowledge and experience, as well as uncertainty involved in current science. We have, however, identified three topics that are topical and important, of great research interest, comprehensible to the target students, and linked to grade-appropriate learning goals. These topics are climate change, fresh water sustainability, and life on other planets.

Principle 2: Acquaint students with working scientists, their research, and their use of computer models. We connect scientists' work with students' learning by giving students opportunities to understand current research and the nature of science. We use multimedia introductions to scientists whose research involves computer models similar to, but vastly more complex than, the computational models used in the modules. By personalizing science, we combat the stereotypes of scientists by showing diverse scientists who work in teams and who can make an intellectual connection to students through their common use of computational models. By using social media, we tap into student familiarity with these tools and provide a way to keep the modules alive and current.

Principle 3: Use model-based experimentation as the primary means for students to acquire content. Computational models are ideal for exploring geosciences and human impact. HAS models simulate the evolution of a system and are based on mathematical algorithms that approximate fundamental physical laws. Much as scientists do, students can experiment with models by controlling the parameters, the starting conditions, and conditions during a run. The models have vivid graphics and run quickly, so that students can experiment and gain insights about the system by carefully observing the evolution of the system. Students can learn the content and the process of science by experimenting with the models, they can gain insights about contemporary science and scientists in the activities, and they can see the cause and effect in a system because the behavior of these models emerge from basic science-based rules. They can make predictions and over many runs, evaluate the probability of their predictions, thereby exploring issues of uncertainty inherent in predicting the future. Virtual environments that students can actively explore with tools and models are valuable for both motivation and content acquisition. It is also important that students take an active role in trying different parameters, arrangements, and initial conditions to run experiments and see the results of their selections.

Principle 4. Support scientific reasoning and argumentation. Engaging students in scientific argumentation deepens science concept learning, altering student views of science, and supporting student decision-making. Research on scientific argumentation has grown substantially in the last ten years. One aspect that has been overlooked, however, is how students treat uncertainty in formulating their arguments. Uncertainty can play two roles when students construct an argument. One type of uncertainty represents students' confidence in their own knowledge and ability. The other type is inherent in scientific inquiry due to measurement errors, lack of conclusive theories or models, and limitations associated with current equipment and technologies. As argumentation is a central scientific practice in the discourse of science, student argumentation will give students insight into how scientists construct knowledge.

Three Curriculum Modules

Module 1: What will Earth's climate be in the future? In this module, students explore past climate changes and learn how mechanisms for positive and negative feedback can affect global temperature. They think about how scientists use this information to make climate change predictions. Students learn about where there is certainty in the climate data and where there is uncertainty with regard to predicting what will happen. This module pays special attention to helping students think about the presented evidence and how to evaluate the conclusions scientists can draw from the evidence. Students explore data from NASA and the Vostok ice cores and look at trends over different time scales. They begin to explore the limitations of conclusions drawn from the data. Then students interact with models to learn about how radiation interacts with Earth's surface and atmosphere, the relationship between ocean surface temperature and carbon dioxide sequestration, the relationship between atmospheric carbon dioxide levels and the amount of water vapor, and, in the final model, the relationship between all three (carbon dioxide, ocean surface temperature, and water vapor). Additionally, students explore albedo, changing the amount of ice and cloud cover in their models to examine how different surfaces provide negative and positive feedbacks to the temperature increases resulting from increased levels of greenhouse gases. Finally, students explore how all the variables interact with each other to produce global temperature effects.

Module 2: Is there life in space? The main focus of this unit is student exploration of planet-hunting methods using a dynamic model that simulates a single planet orbiting a star. The uncertainty questions focus on data interpretation and being able to detect faint to moderate signals in noisy data. Students are introduced to the transit method and the radial velocity method of planet-hunting. The transit method involves interpreting light intensity data from a star in an attempt to observe a periodic drop in brightness. Students explore factors such as planet size, the angle of orbit with respect to the observer, and the precision of the light-sensing instrument on scientist's ability to detect planets via the transit method. Students are also introduced to the radial velocity, or wobble, method of detecting planets. This method involves interpreting the shift in the apparent wavelengths of light coming from a star; as the planet moves around the star, it exerts a gravitational pull, resulting in a star wobble. Students use models to explore the effects of planetary mass on a star's motion, changes in wavelengths of light as related to star motion, and how the angle of orbit influences a scientist's ability to detect a shift in the wavelength. Finally, the module explores conditions for habitability. Students look at properties of five different star types and the zone of habitability around each star. Students end the modules with a focus on how telescopes can be used to analyze light from a star to look at planetary atmospheres and how this information might reveal clues about which planets are more likely to be habitable.

Module 3: Will there be enough freshwater resources for Earth's growing population? The main focus of this module is to have students explore Earth's freshwater resources: where they can be found, how we use them, and why we must think about sustainable use as Earth's population increases. The module ultimately explores why human and ecological needs should be balanced and how freshwater resource issues vary around the world. Students begin by exploring parts of the water cycle: groundwater flow and recharge, evapotranspiration, and precipitation. With the model, students are able follow water through the water cycle. Students evaluate how the supply and demand for fresh water differs around the world. Students then explore the movement of water though the ground; models show how water moves through substances of different permeability. Students use models to explore how aquifers are created. The models enable students to investigate how the level of the water table affects the water level in streams and ponds. Students experiment with creating different subsurface layer configurations to look at the formation of water tables and aquifers. Finally, students focus on the relationship between groundwater recharge, related to permeability and porosity, and the rate at which water is pumped out for human use. Students are introduced to some ways in which humans have disrupted the water cycle and are challenged to suggest solutions to a freshwater availability problem.

Methods

Instrument Design

In order to assess students' scientific argumentation ability, we used nine scientific argumentation item sets in the early year and the end year tests: three addressing climate module content, three addressing water module content, and three addressing space module content. For pre-post tests for each module, the three scientific argumentation item sets that appeared in the early year and the end-of-year tests were included. In addition, additional multiple-choice claim items were included in the module specific pre-post tests. See Table 1 for item content throughout the year.

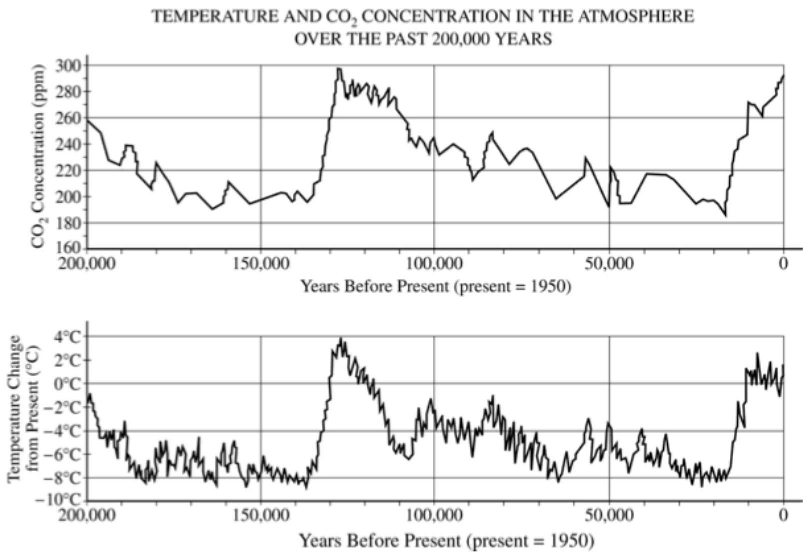
Table 1. Test Content

Items	Early year test	Climate pre-post tests	Water pre-post tests	Space pre-post tests	End year test
• Albedo item set	X	X			X
• T2050 item set	X	X			X
• Ocean item set	X	X			X
• Galaxy item set	X		X		X
• Life item set	X		X		X
• Planet item set	X		X		X
• City water item set	X			X	X
• Well item set	X			X	X
• Sediment item set	X			X	X
• Additional items	--	7 claim + 1 explanation items	6 claim items	5 claim + 1 explanation items	--
Total score	118	51	45	48	118
No. of students who took the test(s)	993	406	380	245	473
No. of teachers who administered the test	12	9	9	7	9

Scientific Argumentation Item Set Example

We selected nine science contexts that were currently investigated by scientists and constructed a scientific argumentation item set for each of nine current science contexts. Each scientific argumentation item set consists of making scientific claims (claim), explaining scientific claims based on evidence (justification), expressing the level of certainty about explanations for the claims (uncertainty), and describing their source of uncertainty (conditions of rebuttal). For claims, either multiple-choice or short-answer item format was used. For justifications, we provided data in graphs, tables, or written statements and asked students to “Explain your answer” in an open-ended format. Then, students were asked to rate their certainty on a five point Likert scale from “1” being not certain at all to “5” being very certain. Students were then asked to explain their ratings. A scientific argumentation item set called T2050 item set is shown in Figure 1.

Figure 1. T2050 Scientific argumentation item set



The graphs show the variation of carbon dioxide concentration and air temperature in Antarctic ice cores over 200,000 years (left side) before 1950 (right side). The upper graph shows carbon dioxide concentration in parts per million (ppm). The lower graph shows the change in air temperature. (Source: 2006 Environmental Science AP Exam)

The CO2 concentration in the year 2000 was measured at 370 ppm. Scientific models predict that atmospheric CO2 will increase to 500 ppm in the year 2050. Based on the trends in the graphs, how much will the air temperature change between 2000 and 2050?

[CLAIM] Will the temperature be higher or lower in 2050?

- higher
- lower
- no change

How many degrees will the temperature change? ______

[EXPLANATION] Explain how you made your prediction.

[UNCERTAINTY] How certain are you about your prediction for the air temperature in 2050?

(1) not certain at all
(2)
(3)
(4)
(5) very certain

[UNCERTAINTY RATIONALE] Explain what influenced your uncertainty.

Data Coding

We coded scientific argumentation item sets according to the following rubrics.

- Multiple-choice claim items:

  - (score 1) congruent with current scientific claim
  - (score 0) incongruent
- Open-ended justification items:
  - (score 4) two or more theory-justified links between evidence and claim
  - (score 3) one theory-justified link between evidence and claim
  - (score 2) relevant pieces of evidence without theory-based justifications
  - (score 1) irrelevant pieces of evidence, scientifically-incorrect justifications, and non-normative ideas
  - (score 0) off-task/ blank
- Certainty rating items on a 5-point Likert scale
  - 1 to 5 according to the number students selected on the Likert scale
- Open-ended uncertainty rationale items:
  - (score 3) scientific uncertainty beyond investigation
  - (score 2) scientific uncertainty within investigation
  - (score 1) personal uncertainty
  - (score 0) no information

Maximum possible scores were 118 for the annual pre-posttests, 51 for climate module pre-post tests, 45 for water module pre-post tests, and 48 for space module pre-post tests.

Data Collection

Table 2 shows when 12 teachers implemented the early year and the end of the year tests as well as three HAS modules. All 12 teachers administered the early year test between September and October. Teacher 2 had the second cohort and administered the early year test in January for that cohort. Nine teachers administered the end of the year test. Seven of them did it in May and June. Two teachers, T3 and T6, administered the end year test in January which was after their second HAS module was implemented. Three teachers implemented one HAS module during the school year while six teachers implemented two HAS modules. Three teachers implemented three HAS modules. The three HAS modules were implemented at different times during the school year because teachers chose implementation times according to their teaching schedules. The nidyke implementation sequence is shown in the second column of Table 2.

Table 2. Module and Assessment Implementation Schedule in Year 3

Teacher Code	Module Sequence	No. of Modules	Early year	Water	Climate	Space	End year
T1	WCS	3	Oct	Oct	Jan	Apr	May
T2a	WS	2	Sep	Sep		Apr	May
T2b	WC	2	Jan	Jan	May		May
T3	CS	2	Oct		Oct	Dec	Jan
T4	SW	2	Oct	Jun		Jan	June
T5	WC	2	Mar	Mar	May		
T6	WC	2	Sep	Nov	Jan		Jan
T7	C	1	Sep		Oct		

T8	SW	2	Sep	Mar		Oct	May
T9	WCS	3	Oct	Jan	Mar	Apr	May
T10	SCW	3	Sep	Apr	Feb	Oct	May
T11a	WC	2	Sep	Sep	Dec		
T11b	C	1	Sep		Apr		May
T12	S	1	Sep			Nov	

Note. C=Climate module, W=Water module, S=Space module

Data Analysis

To compare whether and how much students changed in their scientific argumentation before and after a HAS module, we created a total test score as well as sub scores for claim, explanation, uncertainty rating, and uncertainty rationale items. We then applied repeated measures ANCOVA. The dependent variable was the total scientific argumentation test score and the independent variable was the teacher. Students' gender (male vs. female), technology experience (used technology for learning vs. not used), and ELL status (English as first language vs. second) were entered as covariates.

We used repeated measures ANOVA to examine students' scientific argumentation trajectories for each of the three science topics addressed in the three HAS modules. For the climate trajectories, we used students' scores on the three climate scientific argumentation item sets that appeared in the early-year, before and after the climate module, and the end-year tests. The three scientific argumentation item sets for each topic were taken by students four times over the year. For the water trajectories, we used students' scores on the three water scientific argumentation item sets. For the space trajectories, we used students' scores on the three space scientific argumentation item sets. The maximum possible scores were 39 for the three water item sets, 39 for the three space item sets, and 40 for the three climate item sets. To examine whether there was a systemic difference across teachers, we used the teacher as an independent variable in the repeated measures ANOVA.

Results

Finding 1: Students significantly improved their scientific argumentation ability before and after all three HAS modules. The improvement occurred in all four elements of scientific argumentation, i.e. claim, explanation, uncertainty rating, and uncertainty rationale.

Using identical pre-post tests, we assessed students' scientific argumentation ability before and after the implementation of the Climate, Water, and Space modules. The pretests were taken before the respective modules and the posttests were taken just after the modules were finished. The pretest and the posttest of each module consisted of claim, explanation, uncertainty rating, and certainty rationale items. During the 2011-2012 school year, students of nine teachers completed the Climate pre-post tests, those of nine teachers completed the Water pre-post tests, and those of seven teachers completed the Space pre-post tests. Table 3 shows descriptive statistics for student performances on four argumentation elements separately as well as combined. Student performance changes from pre to posttests are shown in Effect Size defined as Cohen's d (the mean difference from pre to posttest divided by the pooled standard deviations).

Table 3. Student Improvement Before and After HAS Modules

	No. of items	Maximum allowed score	Pretest Mean (SD)	Posttest Mean (SD)	Effect Size (d)***
(a) Climate Module (N= 448 students from nine teachers					
Claim	10	11	5.8 (2.1)	6.6 (2.2)	0.37 SD
Explanation	4	16	5.1 (2.2)	6.1 (2.5)	0.43 SD
Uncertainty rating	3	15	9.5 (2.5)	10.9 (2.6)	0.55 SD
Uncertainty rationale	3	9	2.4 (1.3)	2.7 (1.5)	0.22 SD
Total	20	51	22.8 (5.9)	26.3 (6.5)	0.56 SD
(b) Water Module (N= 409 students from nine teachers)					
Claim	9	9	4.7 (1.8)	5.7 (1.7)	0.57 SD
Explanation	3	12	4.5 (1.9)	5.8 (1.8)	0.70 SD
Uncertainty rating	3	15	10.9 (2.6)	12.3 (2.3)	0.57 SD
Uncertainty rationale	3	9	2.7 (1.7)	3.3 (1.9)	0.33 SD
Total	18	45	22.8 (5.9)	27.1 (5.6)	0.75 SD
(c) Space Module (N=270 students from seven teachers					
Claim	8	8	4.1 (1.5)	5.1 (1.7)	0.63 SD
Explanation	4	16	5.9 (1.8)	7.2 (2.5)	0.60 SD
Uncertainty rating	3	15	10.3 (2.2)	11.9 (2.5)	0.68 SD
Uncertainty rationale	3	9	3.2 (1.6)	3.7 (1.7)	0.30 SD
Total	18	48	23.6 (4.5)	27.9 (6.2)	0.81 SD

Note. SD = Standard Deviation
Effect Size = Cohen's d = Mean difference between pre and posttests divided by the pooled standard deviation of pre and posttests.
***: All pre-post changes listed in the table are statistically significant at the p<.001 level.

As shown in Table 3, students significantly improved their performance on all four elements of scientific argumentation in all three modules. When combining all elements, students' improvement became 0.64 SD for the Climate module, 0.77 SD for the Water module, and 0.85 SD for the Space module. Among the four scientific argumentation elements, the most improved were students' uncertainty rating and explanations while the least improved was the certainty rationale. These results indicate that the HAS curriculum modules supported students' content acquisition as shown in the improvement in scientific claims, scientific reasoning as shown in the improvement in explanations, and consideration of limitations of given evidence as shown in the improvement in certainty rationale. These results also indicate that (1) there was a lot of room for further improvement, and (2) scaffolding should be added to the curriculum modules to further assist students' development of scientific argumentation, in particular on how to consider and explain uncertainty associated with scientific investigations.

Finding 2. The amount of student improvement before and after the HAS modules did not differ in terms of student characteristics but differed across teachers.

Students' gains in scientific argumentation before and after HAS modules were statistically significant for all three HAS modules. See Table 4. The improvement was not significantly dependent upon students' gender, technology experience, and ELL status as there were no significant interaction effects of Time with Gender and Technology. This means that students improved regardless of their gender, technology experience, and ELL status for all three HAS modules.

Table 4. ANCOVA Results on Students' Scientific Argumentation across Three HAS Modules

Module	Climate	Water	Space
(a) Within subjects effects			
Time	11.55***	9.72**	7.76**
Time x Gender	0.004	0.03	0.83
Time x English	0.59	0.35	1.14
Time x Technology	0.50	0.29	0.76
Time x Teacher	8.66***	4.69***	13.87***
(b) Between subjects effects			
Teacher	20.04***	11.08***	13.47***
Gender	2.96	2.17	0.03
English	1.87	1.30	6.62*
Technology	1.18	17.29***	0.93

There was a significant teacher effect indicating that scientific argumentation abilities were different from teacher to teacher. This was expected as students were not randomly drawn from the student population. After controlling for variations due to teacher and students' gender, ELL status, and technology experience, there was a significant interaction effect between TIME and Teacher. That is, students' improvement was significantly different across teachers. This can be better illustrated by comparing Cohen's d values (Effect Sizes) across teachers for each HAS module. See Table 5. For the Climate module, the effect sizes varied from -0.14 SD to 1.72 SD. For the Water module, the effect sizes varied from 0.44 SD to 3.07 SD. For the Space module, the effect sizes varied from -0.09 SD to 2.15 SD. Among 25 module implementations, only two module implementations showed no significant changes: T6's Climate module and T3's Space module. Coincidentally, T3's Space module was implemented in December, right before the winter break, and T6's Climate module was implemented in January, right after the winter break. It might be possible that students were not giving their best efforts to take tests or learn modules.

Table 5. Student Gains in Scientific Argumentation across Teachers

(a) Climate Module

Teacher	n	Pre Mean	Post Mean	Effect Size (SD)
T1	9	22.89	31.33	1.72 SD***
T2	101	20.47	23.71	0.61 SD***
T3	21	18.52	20.86	0.49 SD**
T5	26	28.42	33.81	1.04 SD***
T6	12	24.25	23.58	-0.14 SD

T7	105	22.67	28.81	1.17 SD***
T9	56	21.04	22.59	0.24 SD*
T10	21	26.76	29.62	0.48 SD**
T11	55	26.95	28.35	0.33 SD*
Total	406	22.89	26.47	0.58 SD***

(b) Water Module

Teacher	N	Pre Mean	Post Mean	Effect Size (SD)
T1	8	23.25	31.63	3.07 SD***
T2	135	19.78	25.99	1.07 SD***
T4	66	25.85	28.21	0.44 SD**
T5	31	25.48	29.94	1.08 SD ***
T6	13	24.00	27.54	1.00 SD***
T8	35	23.80	26.49	0.62 SD**
T9	47	21.11	24.49	0.56 SD**
T10	24	24.83	28.33	0.75 SD***
T11	21	28.48	31.00	0.61 SD**
Total	380	22.85	27.15	0.75 SD***

(c) Space Module

Teacher	N	Pre Mean	Post Mean	Effect Size (SD)
T1	9	24.1	32.8	2.06 SD***
T2	27	24.0	32.5	2.15 SD***
T3	35	22.0	21.6	-0.09 SD
T4	68	24.8	30.6	1.22 SD***
T8	28	23.2	27.4	0.94 SD***
T9	52	22.0	24.1	0.42 SD**
T10	26	25.4	30.7	1.11 SD***
Total	245	23.6	27.9	0.81 SD***

Finding 3: Students significantly improved their scientific argumentation abilities over the year.

Early in the 2011-2012 school year, around September and October, 11 teachers administered the annual scientific argumentation pretest to 993 students. Toward the end of the 2011-2012 school year, around May and June, 9 teachers administered the same annual scientific argumentation posttest to 473 students. Among these students, 406 students took both pre and posttests. For the analysis, we took the 379 students who responded to both tests. The annual scientific argumentation test was identical and consisted of nine scientific argumentation item sets: three sets addressing climate topics, three sets addressing water topics, and three sets addressing space topics. Each item set had a maximum score of 13 and thus the whole test had a maximum score of 118. Table 6 shows mean values for the pretest and the posttest, along with student gains in standard deviation units (Effect Size, Cohen's d). Students of all nine teachers gained statistically significantly from the pretest to the posttest with an average effect size of 1.01 SD, a large impact.

We applied repeated measures ANCOVA to examine how student gains in scientific argumentation were correlated with student characteristic variables such as gender, technology experience, and ELL status and how student gains differ across teachers. According to ANCOVA results, students' argumentation performances did not significantly differ by gender, $F(1, 367) = 1.77$, $p = 0.18$, and ELL status, $F(1, 367)=1.21$, $p=0.27$. Nor were there significant performance differences by gender and ELL status in student gains in scientific argumentation. However, there was a significant technology experience effect, $(1, 367) = 5.21$, $p < .05$. Independent samples t-tests indicate that students' argumentation score was not significantly different in the pretest by their technology experience, $t(378)=1.19$, $p=0.23$, but became significantly higher for students with technology experience in the posttest, $t(378)=4.25$, $p<.001$, by an ES of 0.44 SD. In fact, the amount of student gains were significantly different between the two groups, $F(1, 367) = 9.88$, $p < .01$. These indicate that students with technology experience gained scientific argumentation abilities to a greater extent between early and end of the year than those without.

There was a significant main teacher effect, $F(8, 367) = 10.21$, $p < .001$, indicating students' scientific argumentation abilities were significantly different across teachers. See Table 6 for differences in pretest means and posttest means across teachers. In addition, the amount of yearly gains were significantly different across teachers, $F(8, 367) = 8.52$, $p < .001$, ranging from as small as 0.23 SD to as large as 3.06 SD.

Table 6. Descriptive Statistics for Early-Year and End-Year Scientific Argumentation Tests across Teachers

Teacher	n	Pre Mean	Post Mean	D (Effect Size) (Unit: SD)
T1	9	50.11	82.56	3.06***
T2	137	45.74	63.20	1.45***
T3	35	49.49	52.37	0.23*
T4	41	56.85	68.00	0.84***
T6	10	45.00	57.20	1.22***
T8	40	55.55	66.88	1.13***
T9	49	48.43	58.67	0.76**
T10	24	55.29	72.04	1.36***
T11	34	60.41	68.41	0.98***
Total	379	50.68	63.85	1.01***

Finding 4: Students' scientific argumentation trajectories indicated improvement over time. The largest improvement coincided with the implementation of HAS modules. Students retained or even further improved their scientific argumentation after HAS modules were finished.

Figure 2 shows mean plots for students' scores on the three Climate scientific argumentation item sets across four time points. On average, students significantly improved over time, $F(3,636) = 76.33$, $p<.001$. The largest improvement coincided with the Climate module implementation time. Students improved their scientific argumentation between the beginning of the school year and before the Climate module probably because students learned climate topics or related science prior to using the module. Students also well retained their scientific argumentation abilities after the Climate module was finished. There was a significant teacher effect, $F(6,212) = 10.69$, $p<.001$, indicating students were not sampled from the homogeneous student population. The student trajectories were significantly correlated with the teacher variable, $F(18, 636)=4.46$, $p<.001$. Figure 2 indicates this interaction effect, as the trajectories were not identical across teachers. In particular, T6 and T9 departed the pattern of small improvement -> larger improvement -> maintain. The T6 trajectory showed a statistically non-significant decline before and after the module. The T9 trajectory shows a decline between after the module and at the end of the year. The T1 trajectory showed the largest improvement before and after the Climate module.

Figure 3 shows student trajectories for the Water module. The general pattern was that students improved between the beginning of the year and prior to the Water module, followed by a grater improvement before and after the Water module. Then, students' scientific argumentation scores slightly declined. The student improvement over time was statistically significant, $F(3, 741) = 83.42$, $p<.001$. There was a significant teacher effect as in the Climate module, $F(6, 247) = 6.76$, $p<.001$. Figure 3 shows statistically significant variations across teachers in student trajectories, $F(18, 741)=2.35$, $p < .01$. The most distinguished trajectory was T6's as there was a noticeable drop between after the module was finished and the end of the school year. Again, the T1 trajectory shows the largest improvement before and after the module.

Figure 4 shows student trajectories for the Space module. The general pattern of overall improvement was similar to the general pattern found with the Water module. The improvement was statistically significant over time, $(3, 546)= 115.72$, $p<.001$. There was a significant teacher effect, $F(6, 182) = 10.16$, $p<.001$. The student trajectories were significantly different across teachers, $F(18, 546) =5.68$, $p<.001$. In particular, T3 shows almost no changes across time points. As discussed earlier, T3 implemented the Space module just before the winter break when students might not take the Space module seriously and administered the end year test after the winter break when students might not remember what they learned before the winter break. The T1 trajectory indicates the largest improvement before and after the Space module, followed by a further improvement towards the end of the year.

Figure 2. Climate Student Trajectories by Teacher

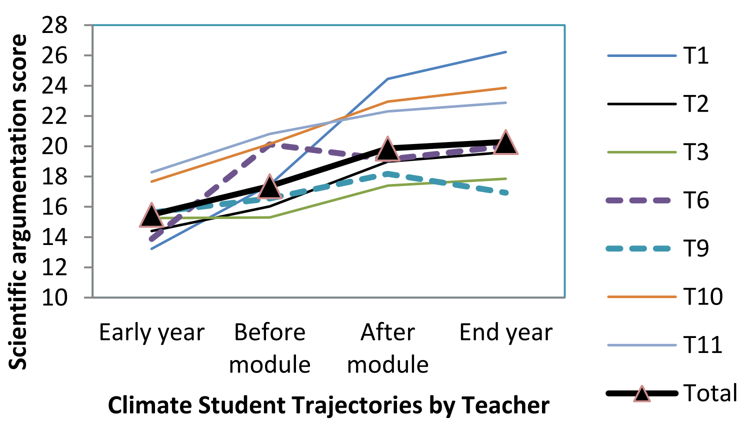


Figure 3. Water Student Trajectories by Teacher

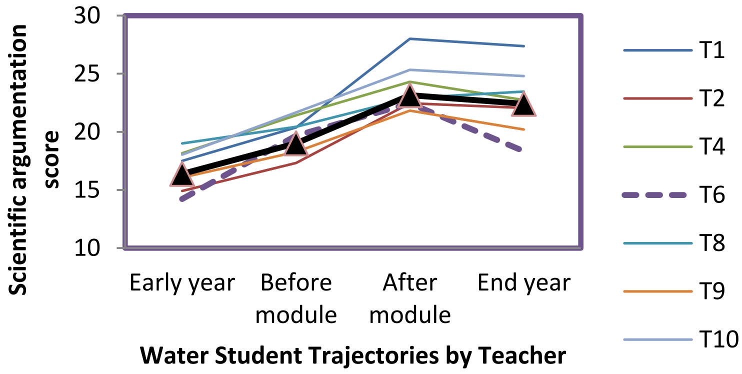
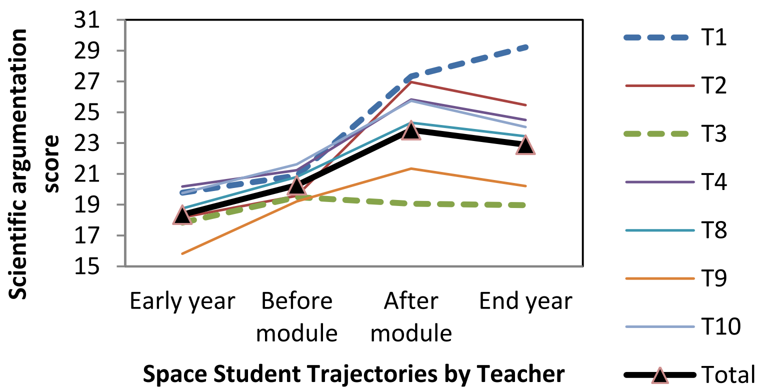


Figure 4. Space Student Trajectories by Teacher



Conclusion

We designed three online inquiry-based curriculum modules addressing authentic science topics present day scientists are investigating. We promoted scientific argumentation tasks throughout the modules as part of module activities. We assessed students' scientific argumentation performances over the course of one school year and found that students' scientific argumentation performances increased significantly between the beginning and the end of the school year, regardless of students' language status, gender, and technology use for learning. We also found that the improvement in students' scientific argumentation performances varied from teacher to teacher. These results indicate that students can improve their scientific argumentation abilities through computational model-based inquiry-based curriculum modules dealing with uncertain current science. We will further present how students progressed during the modules with computational models using student examples to support these findings and discuss implications for future curriculum design and research on scientific argumentation.

References

Chin, C., & Osborne, J. (2010). Students' questions and discursive interaction: Their impact on argumentation during collaborative group discussions in science. *Journal of Research in Science Teaching, 47*(7), 883-908.

Holyoak, K. J., & Morrison, R. G. (2005). Thinking and reasoning: A reader's guide. In K. J. Holyoak & R. G. Morrison (Eds.), *The Cambridge handbook of thinking and reasoning*. New York: Cambridge University Press.

Kennedy, D. & Norman, C. (2005). 125 Questions: What don't we know? Science, sciencemag.org/sciext/125th/.

Sampson, V., & Clark, D. B. (2008). Assessment of the ways students generate arguments in science education: Current perspectives and recommendations for future directions. *Science Education, 92*, 447-472.

Yeh, S. S. (2001). Tests worth teaching to: Constructing state-mandated tests that emphasize critical thinking. *Educational Researcher, 30*(9), 12-17.