

# Scaffolding argumentation about water quality: a mixed-method study in a rural middle school

Brian R. Belland · Jianguye Gu · Sara Armbrust · Brant Cook

Published online: 5 March 2015

© Association for Educational Communications and Technology 2015

**Abstract** A common way for students to develop scientific argumentation abilities is through argumentation about socioscientific issues, defined as scientific problems with social, ethical, and moral aspects. Computer-based scaffolding can support students in this process. In this mixed method study, we examined the use and impact of computer based scaffolding to support middle school students' creation of evidence-based arguments during a 3-week problem-based learning unit focused on the water quality of a local river. We found a significant and substantial impact on the argument evaluation ability of lower-achieving students, and preliminary evidence of an impact on argument evaluation ability among low-SES students. We also found that students used the various available support—computer-based scaffolding, teacher scaffolding, and groupmate support—in different ways to counter differing challenges. We then formulated changes to the scaffolds on the basis of research results.

**Keywords** Computer-based scaffolding · Socioscientific issues · Middle school · Argumentation · Water quality

## Introduction

A critical process in science is argumentation—the manner by which scientists communicate investigations, findings, and interpretations (Abi-El-Mona and Abd-El-Khalick 2011; Osborne 2010). But argumentation is often absent from science classrooms, which reduces the chances that students will learn this skill (Jonassen and Kim 2010; Kuhn 2010). Providing scaffolding is one way to increase students' argumentation skill as they engage with locally relevant problems (Belland 2010; Klosterman et al. 2011). In this paper, we

---

B. R. Belland (✉) · J. Gu · S. Armbrust · B. Cook  
Department of Instructional Technology and Learning Sciences, Utah State University, 2830 Old Main Hill, Logan, UT 84322, USA  
e-mail: brian.belland@usu.edu

use a mixed method approach to explore the influence of computer-based scaffolds on students' abilities to create and evaluate arguments, as well as how and why students use the scaffolds.

### **The next generation science standards and authentic scientific problems**

The Next Generation Science Standards stress the need for K-12 students to address and argue about authentic, ill-structured scientific problems in science class (Achieve 2013; Wilson 2013). Authentic scientific problems have many equally valid solutions and many ways of arriving at a solution, and require complex cognitive processes (Chinn and Malhotra 2002; Giere 1990). To weigh solutions, students need to engage in argumentation (Jonassen and Kim 2010), but they often struggle doing so, in part due to limited opportunities to argue in school (Driver et al. 2000). To bring authentic scientific problems and the opportunity to argue about science to K-12 science instruction, one can employ problem-based learning, socioscientific issues, and scaffolding in science (Jonassen and Kim 2010; Wilson 2013).

#### **Problem-based learning**

In problem-based learning (PBL), small groups of students define an ill-structured problem, determine and find needed information, develop a solution, and justify the solution with evidence (Belland et al. 2008; Hmelo-Silver 2004). PBL requires that students direct their learning (Loyens et al. 2008), and teachers facilitate (Hmelo-Silver and Barrows 2006).

#### *Socioscientific issues*

Because K-12 students often struggle to grasp the idea of scientists arguing with each other, it is important to present authentic problems about which there are many publicized arguments among scientists and other citizens. Such problems include socioscientific issues (SSIs)—ill-structured problems that are addressed through consideration of scientific, social, and ethical concerns (Kolstø 2001; Sadler and Donnelly 2006). The perspective of the problem solver (stakeholder position) is central to consideration of potential solutions (Rose and Barton 2012). For example, when considering whether a factory should be allowed if it will release pollution into the sea, students needed to weigh potential drawbacks (e.g., habitat degradation) against job creation (Tal and Kedmi 2006). Unemployed individuals would bring a different perspective to the problem than people concerned with fish populations.

#### *Argumentation*

Central to PBL and SSIs is argumentation—backing claims with evidence by way of premises (Perelman and Olbrechts-Tyteca 1958). Argumentation is how scientific ideas are (a) proposed, supported, and refined, and (b) evaluated (Ford 2012; Osborne 2010). Arguments are considered valid based on the degree to which the audience (a) agrees with the claims and does not object to the premises (Perelman and Olbrechts-Tyteca 1958) and (b) does not find an opposing argument more compelling (Kuhn 1991). Middle school

(Hogan and Maglienti 2001; Yoon 2011) and high school (Walker and Zeidler 2007) students often construct and evaluate arguments based on personal beliefs and social norms. This results from little experience developing arguments (Abi-El-Mona and Abd-El-Khalick 2011; Jonassen and Kim 2010), poor understanding of epistemological norms (Kuhn 1991; Weinstock et al. 2004), and my-side bias (Kuhn 1991; Stanovich and West 2008).

### *Argumentation and epistemological criteria*

Arguments are evaluated with reference to epistemological criteria and exploration of alternative explanations (Hogan and Maglienti 2001; Weinstock et al. 2004). Epistemological criteria are beliefs about the nature of knowledge, how statements and evidence can be verified, and how claims can be supported (Schraw 2001; Weinstock et al. 2004). Epistemological criteria can vary by education level (Hogan and Maglienti 2001; Kuhn 1991). For example, professional scientists indicated that an argument needed to be coherent and open to verification, while university students and high school teachers did not (Abi-El-Mona and Abd-El-Khalick 2011).

### *Scaffolding*

Scaffolding was defined originally as contingent support provided by a teacher that enabled students to solve problems that are beyond the students' unassisted capabilities (Wood et al. 1976). Computer-based scaffolding is developed based on an analysis of expected student difficulties (Belland 2014), can be used to supplement teacher scaffolding (Saye and Brush 2002), and is often designed to simplify and highlight complexity in learning tasks (Reiser 2004). It can do this by providing motivational support, questioning, modeling, highlighting important task elements, and providing feedback (Belland 2014; Wood et al. 1976).

The influence of scaffolding strategies (e.g., promoting students' discussion of problems and evaluation of such) can vary by achievement level (Cuevas et al. 2002; Rivard 2004; So et al. 2010; White and Frederiksen 1998) and socio-economic status (Cuevas et al. 2005; Lynch et al. 2005). Furthermore, recent research indicates that students can use the same computer-based scaffolding in different ways depending on their individual goals (Belland and Drake 2013).

### *Scaffolding argumentation*

Efforts to scaffold argumentation about socioscientific issues have largely focused on three areas in which students often struggle: creating suitable claims based on presented evidence (Cho and Jonassen 2002), evaluating the quality of evidence (Nicolaidou et al. 2011), and engaging in the entire argumentation process (Belland et al. 2008; Buckland and Chinn 2010; Clark and Sampson 2007).

Helping students characterize the nature of the relationships among the underlying variables in a problem, and think about how such relationships relate to the problem is one way to scaffold students' creation of suitable claims based on presented evidence. For example, in a unit focused on water quality, students should be encouraged to think about how turbidity and dissolved oxygen levels inter-relate and inform each other, and are indications of the health of a river. But a claim that dissolved oxygen is lower than the

standards does not form the basis of a suitable argument in response to a driving question *What should be done to optimize the water quality in X river?* Such scaffolding often employs graphical interfaces such as concept maps (e.g., *Belvedere*), with which students can construct links between argument elements (Scheuer et al. 2010) or models of the underlying elements of the problem (Zhang et al. 2006). Such scaffolds have led to increased argumentation ability. For example, *Belvedere* helped university students produce higher quality claims and evidence than control students (Cho and Jonassen 2002). Many such systems come with preformed argument elements, such that using them during PBL is not feasible (Scheuer et al. 2010).

Students often struggle to evaluate the quality of evidence. One way to scaffold students' evaluation of the credibility of information is to prompt students to consider such factors as the funding source, publication type, and the use of experimental and control groups when evaluating sources (Nicolaidou et al. 2011). This approach led to a significant gain in credibility assessment ability among high school students (Nicolaidou et al. 2011).

When focused on the entire argumentation process, scaffolding often supports students' creation of claims, evaluation of the credibility of and use of evidence, and ability to put claims and evidence together, as well as evaluate and respond to others' arguments. For example, in an online unit on genetically modified food, questions invited high school students to consider validity of claims, objectivity, and ethics (Walker and Zeidler 2007). However, these scaffolds did not have a strong influence on students' arguments (Walker and Zeidler 2007). In another example, *Sensemaker* allows students to arrange preformed evidence pieces and claims (Bell 1997). This led students to use premises in their arguments "70 % of the time" (Bell 1997, p. 6).

## This study

Using context-specific scaffolding such as *Sensemaker* was not an option in this study, as students needed to construct arguments with evidence that they collected themselves about a local water quality problem. Furthermore, *Belvedere* has evolved to support arguments in mostly fairly well-defined domains (Scheuer et al. 2010). This study was in Year 2 of a 5-year design-based research project in which we develop a generic scaffold designed to support middle school students' construction of evidence-based arguments while solving locally relevant SSI problems.

## Research questions

1. How does argumentation scaffolding influence argument evaluation ability, and does the influence vary based on prior science achievement or socioeconomic status?
2. How does argumentation scaffolding influence group argument quality?
3. What do middle school students use for support as they create evidence-based arguments, and why?
4. How can argumentation scaffolds be redesigned to provide stronger support?

## Method

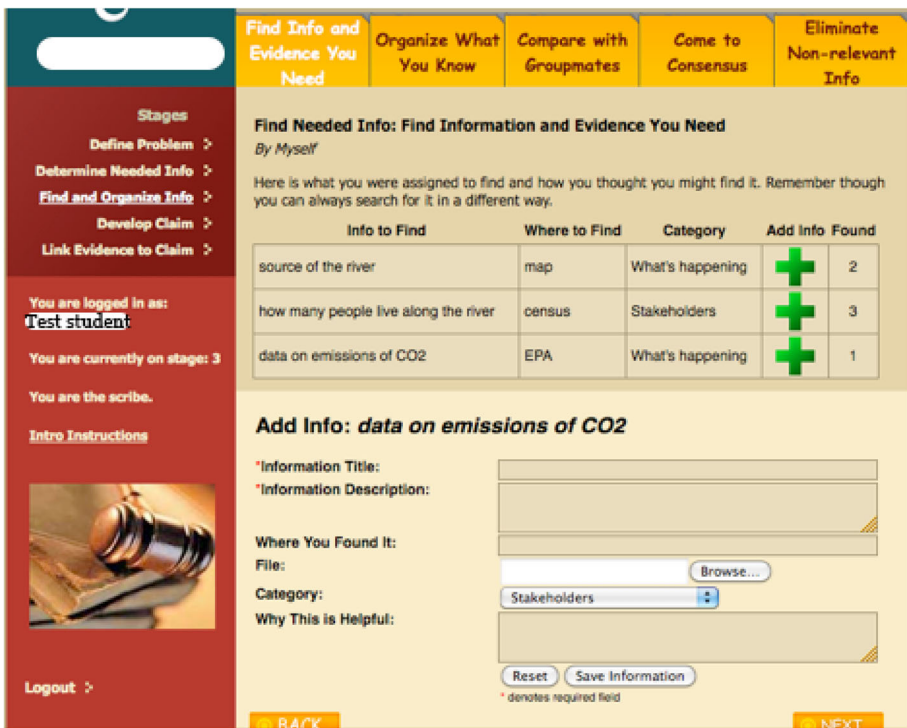
### Setting and participants

This study took place in a middle school (41 % free or reduced lunch) located in a small, rural community in the Intermountain West. One teacher and sixty-nine 7th grade students from three intact classes participated. The teacher, Mr. Thomas (note: all names have been changed) taught middle school science for 22 years, but had never used inquiry-based approaches.

Students engaged in a 3-week problem-based learning unit focused on the water quality of the local river—the Dale River. Students worked in groups of 3–4, and each group argued from a unique stakeholder position. There were 6–7 groups per period. Stakeholder positions included environmentalists, the Environmental Protection Agency, recreationists, farmers from Monroe, Monroe state government, common citizens of Greenville, and farmers from Madison (the next state south). For example, students who represented farmers from Monroe needed to consider what farmers need (e.g., water good enough for irrigation and for animals to drink).

During this PBL unit, students:

- Tested water quality (e.g., phosphates) at various locations along the river
- Compared data to standards and historical data and determined sources of increases in pollution



The screenshot shows the 'Connection Log' interface. At the top, there are five yellow tabs: 'Find Info and Evidence You Need', 'Organize What You Know', 'Compare with Groupmates', 'Come to Consensus', and 'Eliminate Non-relevant Info'. The first tab is active.

On the left, a red sidebar contains a 'Stages' menu with options: 'Define Problem', 'Determine Needed Info', 'Find and Organize Info' (highlighted), 'Develop Claim', and 'Link Evidence to Claim'. Below this, it shows 'You are logged in as: Test student', 'You are currently on stage: 3', and 'You are the scribe.' There is also an 'Intro Instructions' section with an image of a gavel and a 'Logout' button.

The main content area is titled 'Find Needed Info: Find Information and Evidence You Need' and 'By Myself'. It contains the text: 'Here is what you were assigned to find and how you thought you might find it. Remember though you can always search for it in a different way.'

Info to Find	Where to Find	Category	Add Info	Found
source of the river	map	What's happening	+	2
how many people live along the river	census	Stakeholders	+	3
data on emissions of CO2	EPA	What's happening	+	1

Below the table, there is a section titled 'Add Info: data on emissions of CO2' with a form containing the following fields:

- \*Information Title:
- \*Information Description:
- Where You Found It:
- File:
- Category:
- Why This is Helpful:

At the bottom of the form are 'Reset' and 'Save Information' buttons. A note indicates '\* denotes required field'. Navigation buttons 'BACK' and 'NEXT' are at the bottom of the interface.

Fig. 1 One group member logged in at stage 3 step 1 of the *Connection Log*

- Recommended and presented evidence-based solutions from their stakeholder perspectives

We randomly assigned two class sections to use computer-based scaffolds (described below). Control students completed the same activities as the experimental students, but without the support of computer-based scaffolds.

Mr. Thomas led both experimental and control students during the unit. Research group members were present to collect data and answer technical questions related to the computer-based scaffold used by the experimental group (the *Connection Log*), but were also present in the control section; they provided some support related to unit content, and such support was equal between experimental and control conditions. However, Mr. Thomas provided the vast majority of non-technical, one-to-one support.

## Materials

The *Connection Log* is a generic scaffold designed to support middle school students' creation of evidence-based arguments during PBL units (see Fig. 1). In 2010–2011 we revised the *Connection Log* based on expert reviews and one-to-one evaluations with target students.

### *Theory behind design*

Two big challenges middle school students face in argumentation are knowing argumentation norms and process (Kuhn 1991; Weinstock et al. 2004) and engaging in productive group work (Vellom and Anderson 1999). Argumentation norms and process are forms of cultural knowledge (Driver et al. 2000), which cannot be transmitted through didactic instruction, but rather can be assimilated through social interaction (Luria 1976; Vygotsky 1978). Thus, we scaffolded students' (a) understanding of the argumentation process through a process map and question prompts, (b) engagement in the work coordination and consensus-building process, and (c) anticipation of criteria used to judge their arguments (Quintana et al. 2003). Also, we encouraged students to apply the norms and process they were learning to evaluate claims and evidence produced by their groupmates.

### *How it works*

The scaffolds consist of five stages in the process of building an evidence-based argument (Belland et al. 2008):

1. Define the problem
2. Determine needed information
3. Find and organize needed information
4. Develop claim
5. Link evidence to claim

The *Connection Log* provides support both for students' inquiry process and their argumentation. Within each stage are 2–4 steps (e.g., responding to questions). In the first 1–2 steps of each stage, students respond individually to question prompts. For example, step 1 of the *Link evidence to claim* stage prompted students to support claims that they had been assigned:

To add evidence to support a given claim, click the add sign next to the claim and follow the instructions below. If after you click the Plus button you do not see enough evidence to support the claim, please click the Determine Needed Info link on the left, enter new information to find, and then find it.

In later steps, students read what their groupmates wrote and come to consensus. For example, in step 3 of the *Link evidence to claim* stage, students can see how groupmates supported the group claims, and can rearrange the argument to make it more coherent.

### *Existing empirical evidence*

In an earlier study, lower-achieving experimental students performed significantly better in argument evaluation ( $ES = 0.61$ ) than lower-achieving control students (Belland et al. 2011). In another study, average-achieving experimental students performed significantly better in argument evaluation ( $ES = 0.62$ ) than average-achieving control students (Belland 2010). Furthermore, students used the scaffolds in various ways in response to their challenges (Belland 2010; Belland et al. 2011). These prior studies were conducted in the context of a seventh grade PBL unit on the Human Genome Project. The current study focuses on a PBL unit on water quality. Thus, it is important to establish if the *Connection Log* supports student learning in the context of a unit of a different subject—environmental science versus genetics.

### Variables

Scaffold condition (*Connection Log* vs. control), students' prior science achievement level (lower-achieving versus higher-achieving), and SES (low SES versus average to higher SES) were used as independent variables. Gain scores of students' argument evaluation ability scores from pre to post test were the dependent variable for each student. Group argument quality scores (one score assigned to each group) were the dependent variable for each group.

### Data collection instruments

#### *Argument evaluation pre and posttest*

All students took an argument evaluation ability pre and posttest, which was adapted with permission from that used by Glassner, Weinstock, and Neuman (2005). These tests varied but were found to be of equivalent difficulty when piloted among 7th grade students. In previous studies among 7th grade students, the instrument showed good internal consistency (coefficient alphas from 0.78 to 0.82; Belland 2010; Belland et al. 2011). The pretest and posttest both consist of four sets of questions; each set has six questions. In each set, individual students need to read a scenario in which a claim and two supporting statements are provided. Students needed to rate how well each statement supported the claim from among from three choices: “doesn't help”, “helps a little”, and “helps a lot.” For example, students need to rate how well the statements “skateboarding is a fun activity for kids to do after school” and “50 % of the kids in this town skateboard, and skateboarding after school at a skateboard park would keep them out of trouble” can support the claim “the city should build a skateboard park.”

### *Group argument quality rating*

We videotaped all groups' presentations of their solutions. Two trained raters rated the argumentation quality of each presentation separately and came to consensus. The rubric was used to assign a numerical score for argument quality. Each group could earn up to one point in each of the following categories: interpretation of pertinent data, provision of a problem solution, reasoning with refutable evidence, integration with their stakeholder perspective, and interconnection among groupmates' segments of the presentation.

### *Videotaped interactions*

In the larger data set, we selected three groups per period based on typical case sampling (Patton 2002), covering both the experimental and control groups. From these nine groups, we selected four based on typical case sampling for this study, which represent at least one typical group per class period. We videotaped the groups during the entire unit, using a microphone to capture all dialogue. All dialogue was transcribed verbatim.

### *Prompted interviews*

We conducted 30-min interviews with the small groups that we videotaped. In each interview a unique 20-min video prompted students' recollection of how they used the *Connection Log* and why. Video clips were selected by viewing video of the entire unit and selecting representative episodes in which students talked about evidence with the teacher, their groupmates, or by themselves. Questions focused on (a) what students were doing and why in covered episodes, (b) how students perceived that they were supported during the unit, and (c) what students thought it means to prove something.

### *Student responses to scaffold prompts*

We retrieved responses to scaffold prompts from students in the experimental condition.

### *Log data*

We retrieved log data that indicated when students visited various pages of the *Connection Log*, how long they stayed there, and what page they arrived from.

### *Prior science scores*

We collected students' scores from science quizzes, projects, and homework from school records and used the average thereof as students' science scores. Previous studies indicated that the impact of the *Connection Log* varied among students with different achievement levels (Belland 2010; Belland et al. 2011). We did a mean-split to divide students into groups: students whose science scores were higher than 92.6 were deemed higher-achieving, while students with scores lower than 92.6 were considered lower-achieving.

### *Free and reduced lunch status*

Students who received free or reduced lunch were considered low SES, while those who paid for lunch were deemed average to higher SES. Free and reduced lunch status is not a



perfect measure of SES, as it relies on self-report of income to principals, and students can receive free and reduced lunch if their family income is up to 185 % of the Health and Human Services poverty guidelines (Cruse and Powers 2006). The HHS guideline for a family of 4 in 2012 was \$26,510 (U.S. Department of Health & Human Services 2012), meaning children in such a family could receive free or reduced lunch if their family income was up to \$49,043.50. According to IRB guidelines, the principal could not match exact family income to student data, but could do so with free and reduced lunch data. The average income for 4-member households in the Intermountain west was \$65,912.50 (Census 2014), thus using free and reduced lunch status allows for a reasonable classification of SES levels.

## Procedures

Before unit start, students took the pretest. Then, they participated in the unit 50 min per day. On Day 1, a guest speaker introduced the history of the Dale River. On Day 2, the teacher introduced the project by explaining the central problem and water quality concepts (e.g., turbidity, beneficial uses). On Day 3, each group was assigned a stakeholder position and began to discuss their stakeholder's characteristics. On Day 4, students conducted several water tests at three different spots along the Dale River. Then from Day 5 to Day 14, students worked in small groups to (a) analyze water quality data, (b) find additional needed information, (c) develop a solution from their stakeholder perspective, and (d) link evidence to their problem solution. For part of Day 5, experimental students learned how to use the *Connection Log* through demonstration and practice. On Day 5 to Day 14, experimental students completed tasks with the help of the *Connection Log* and teacher scaffolding. On these same days, control students did so with the help of only teacher scaffolding. At unit end, each group explained its solution to the county commissioner. The day after, all students took the posttest. Then, we interviewed three groups from each class section. These same groups were videotaped during the entire unit.

## Analysis strategies

### *How does argumentation scaffolding influence argument evaluation ability?*

We followed a two-stage analysis strategy. First, we conducted exploratory data analysis by calculating descriptive statistics for the entire sample and for subgroups according to prior science achievement and SES (Gelman 2004).

There was no significant difference in pretest scores between class sections,  $F(2, 60) = 3.24, p = 0.798$ . Because students were assigned to treatment conditions as intact class sections, we used nested ANOVA to determine if there was a nested effect. There was no significant nested effect of scaffolds condition on argument evaluation ability,  $F(1, 56) = 0.14, p = 0.71$ . Thus, we used two-way ANOVA to analyze how scaffolds influence students' argument evaluation ability among student subgroups. In the first ANOVA analysis, we compared gain scores, using prior science achievement and scaffold condition as independent variables. For the second ANOVA analysis, we used SES and scaffold condition as independent variables.

### *How does argumentation scaffolding influence group argument quality?*

Because the sample size for group argument quality was low (one score assigned to each group), we used a two-tailed *t* test to compare argument quality between the control and experimental groups.

### *How can argumentation scaffolds be redesigned to provide stronger support? and What do middle school students use for support as they create evidence-based arguments, and how?*

We followed the iterative process of qualitative research: data collection, data reduction, data display, and conclusion-drawing/verification (Miles and Huberman 1984). Data reduction involved the use of a coding scheme characterized by 20 top-level nodes (e.g., action in response to scaffold prompt, trustworthiness of information). Analysis was guided by the symbolic interactionism framework, which acknowledges that individuals use tools in manners that respond to their needs and the extent to which past and present interactions with the tools and similar tools met needs (Blumer 1969). Validity was addressed through triangulation, prolonged engagement in the field, searching for disconfirming evidence, thick, rich description, and weekly peer debriefing throughout the analysis process (Creswell and Miller 2000).

## Results

### How does argumentation scaffolding influence argument evaluation ability?

The Cronbach's alpha for the pretest and posttest was 0.84 and 0.88, respectively, which represents satisfactory internal consistency (Nunnally 1978). Since we assigned intact classrooms, we first examined gain scores by class period to search for potential outliers using the 1.5 X interquartile range criterion (Sheskin 2011). After removing two outliers, we calculated descriptive statistics for each subgroup (see Table 1).

**Table 1** Gain scores from argument evaluation pretest to posttest by subgroup

Condition	Sub group	M	SD	<i>N</i>
Control group	Lower-achieving	-1.67	2.88	6
	Higher-achieving	2.67	3.18	15
	Low SES	1.25	3.55	12
	Average to high SES	1.67	3.94	9
	Total	1.43	3.63	21
Experimental group	Lower-achieving	1.21	3.26	14
	Higher-achieving	1.36	3.19	22
	Low SES	2.71	3.91	14
	Average to high SES	0.41	2.26	22
	Total	1.31	3.17	36
Total	Lower-achieving	0.35	3.36	20
	Higher-achieving	1.89	3.20	37
	Low SES	2.04	3.75	26
	Average to high SES	0.77	2.84	31
	Total	1.35	3.31	57

Gain scores ranged from -6 to +9

The gain scores were normally distributed in the control and experimental groups. Control and experimental students improved on the posttest; the average gain score in the control group ( $M = 1.43$ ) was slightly higher than the average gain score in experimental group ( $M = 1.31$ ).

To explore trends among subgroups, we examined gain scores among students with different prior science achievement levels and socioeconomic status (see Table 1). In the experimental condition, lower-achieving students gained less ( $M = 1.21$ ) than higher-achieving students ( $M = 1.36$ ),  $ES = 1.43$ . Low SES students gained more ( $M = 2.71$ ) than average to high SES students ( $M = 0.41$ ),  $ES = 0.72$ .

In the control condition, lower-achieving students had a negative average gain score ( $M = -1.67$ ), while higher-achieving students had positive gain scores ( $M = 2.67$ ),  $ES = 1.43$ . The average to high SES students also had higher gain scores ( $M = 1.67$ ) than low SES students ( $M = 1.25$ ),  $ES = 0.11$ . These results show a potential differential influence of computer-based scaffolds among subgroups. Thus, we conducted follow up inferential data analysis.

We first used students' prior science achievement and scaffold condition as independent variables to compare gain scores between the control and experimental groups. A two-way ANOVA showed a significant main effect of prior science achievement,  $F(1, 53) = 5.69$ ,  $p < 0.05$ , and a significant interaction effect between scaffold condition and prior science achievement,  $F(1, 53) = 4.96$ ,  $p < 0.05$ . The main effect of scaffold condition was not statistically significant,  $F(1, 53) = 0.70$ ,  $p = 0.41$  (See Table 2 for ANOVA source table). Lower-achieving experimental students gained more from pre to posttest than lower-achieving control students ( $ES = 0.93$ ), while higher-achieving experimental students gained less from pre to posttest than their counterparts in control group ( $ES = 0.41$ ) (see Fig. 2). This provides evidence that lower-achieving students benefitted from the *Connection Log*, while higher-achieving students did not.

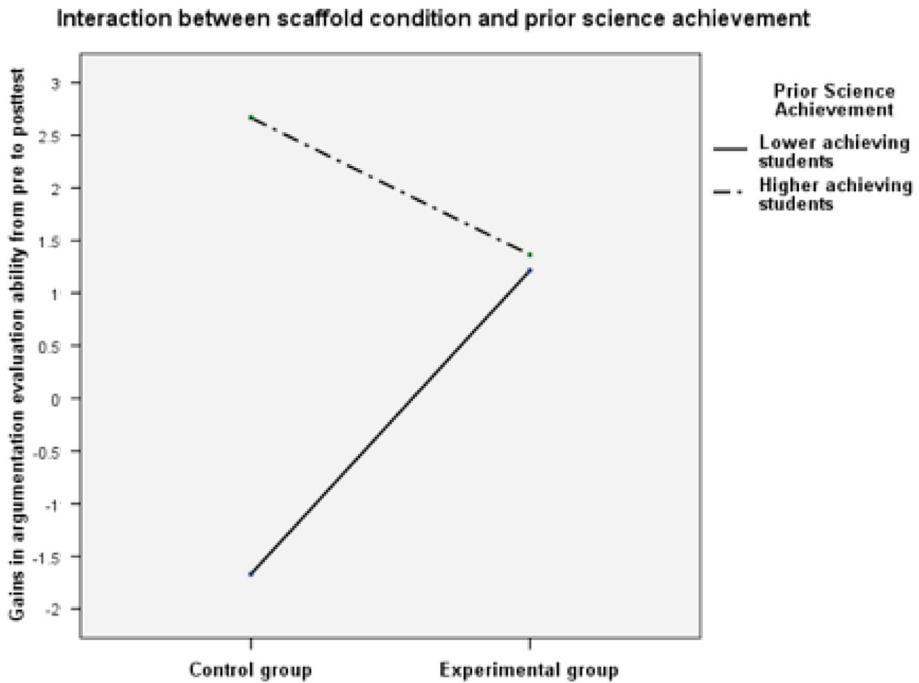
We also used ANOVA to examine the influence of computer-based scaffolds among students from different SES levels (See Table 1). There was no significant main effect of SES,  $F(1, 53) = 1.068$ ,  $p = 0.31$  or interaction between scaffold condition and SES,  $F(1, 53) = 2.22$ ,  $p = 0.14$  (see Table 3 for the source table). But trends favored experimental low-SES students, who achieved higher gain scores than control low-SES students,  $ES = 0.4$ .

### How does argumentation scaffolding influence group argument quality?

Up to four points were possible for group argument quality. Eleven groups received the maximum score. For control students, two out of six groups provided strong justification in their arguments and four groups provided strong interconnected justification. Out of 13

**Table 2** Two-way ANOVA (scaffold condition and prior science achievement) source table

Source	Sum of squares	df	Mean square	F	Sig.	Partial eta squared
Scaffold condition	7.11	1	7.11	0.71	0.41	0.013
Prior science achievement	57.38	1	57.38	5.69	0.021	0.097
Interaction	49.99	1	49.99	4.96	0.03	0.086
Error	534.12	53	10.078			
Total	719.00	57				



**Fig. 2** Interaction effect between scaffold condition and prior science achievement on argument evaluation ability

**Table 3** Two-way ANOVA (scaffold condition and SES) source table

Source	Sum of squares	df	Mean Square	F	Sig.	Partial eta squared
Scaffold condition	0.14	1	0.14	0.013	0.910	0.00
SES	11.46	1	11.46	1.068	0.31	0.02
Interaction	23.80	1	23.80	2.219	0.14	0.04
Error	568.43	53	10.73			
Total	719.00	57				

**Table 4** Group argument quality scores

Condition	Mean	SD	<i>N</i>
Control group	3.67	0.52	6
Experimental group	3.30	0.85	13
Total	3.42	0.77	19

The rubric was designed to assign scores from 1 (presentation of a solution with no justification) to 4 (presentation of a solution with a strong, interconnected justification)

experimental groups, three groups provided weak justification, three groups provided strong justification, and the rest provided strong interconnected justification.

A two-tailed t-test showed that there were no significant differences in argument quality between experimental and control students,  $t(17) = 0.94, p = 0.36$  (see Table 4),  $ES = 0.47$ .

What do middle school students use for support as they create evidence-based arguments, and why?

#### *Group E1: experimental condition*

Group E1 consisted of Melissa (scribe), Jimmy, Justin, and Aaron. They were assigned the stakeholder position of environmentalists.

*Epistemological criteria* In the interview, Justin noted, “If there’s a crop circle, you could [say], ‘Aliens did not come here. You do not have any proof.’ But if you had a picture of an alien landing in the crop circle, [you] have proof and [you] would be right.” Yet, during the project, they did not appear to know how to choose appropriate evidence to support their claims.

#### *Challenges*

*Understanding how to use the Connection Log* At first, Group E1 members did not know how to answer the *Connection Log*’s prompts. Jimmy asked the teacher for help.

*Interpreting data* The students struggled interpreting water quality data—specifically outliers and data from different locations. On Day 6, Aaron asked how to interpret an outlier:

Aaron: What are we supposed to do?

Melissa: We have to... look at the data... you see how... in the water temperature, it’s like 12.8, 12.8, 12.7, 12.9... And then it goes 21.8? So, I don’t know if that was just a typo or if it, like, really was raised that high.

*Providing a solution to the problem* Group E1 members appeared to think that they only needed to search for information rather than offer solutions. Based on usage data, this group spent the most time on the “Determine needed info” and “Find and Organize Info” stages. Although this group did better than other groups at interpreting water quality changes along the river and created several claims such as “All the factors are effecting [sic] the oxygen in the river which means there are not many bugs,” these claims were not included in the presentation.

On Day 14, when the teacher asked them how they would solve the problem, only Melissa and Aaron provided rough solutions. Melissa thought the county should “put like a wall, or a ditch.” Aaron thought they can “take out a whole bunch of the rocks in there and fill it back with dirt and press it down” to reduce turbidity. However, Aaron’s solution would make turbidity worse. These solutions were not considered further and did not make it into the presentation.

#### *What they used for support and why*

*Groupmate support: To learn how to operate the Connection Log* Group members often shared knowledge of how to operate the *Connection Log*. For example, they helped each other log into and edit their entries in the *Connection Log*.

*Teacher support: To learn how to operate the Connection Log* When unsuccessful obtaining support on how to use the *Connection Log* from groupmates, they asked the teacher. For example, when students were searching for information, they struggled to add found information. So the students turned to Mr. Thomas, who explained how to do so.

*Teacher support: To learn how to interpret the data* Mr. Thomas gave examples of how to find patterns in the data. This helped Melissa find that “there wasn’t a lot of bug life... in areas. There was a bunch in others, so... it’s polluted more in some areas.” Justin found, “the phosphate goes up a lot on the way down and the DO [dissolved oxygen] goes down.”

*Connection Log: To divide tasks among group members* Before using the *Connection Log* to divide tasks, the scribe assigned tasks to each group member in person. For example, on Day 7, Melissa asked Justin to examine phosphate data and asked Jimmy to look at dissolved oxygen data. Justin and Jimmy forgot and switched tasks on Day 10. After Melissa began to use the *Connection Log* to assign tasks, the students were clear on assigned tasks.

*Connection Log: To articulate ideas that could be critiqued* To come to consensus, students reviewed each other’s entries and gave feedback. For example, while reviewing her groupmates’ problem definitions, Melissa saw that Aaron and Justin’s problem definitions were the same. Thus, she asked them to revise their definitions.

*Connection Log: To develop their understanding of the problem* In each stage, students need to answer several prompts. For example, at the start of the unit, they developed their understanding of their stakeholder by answering prompts in the *Connection Log*.

Melissa: [Reading prompt questions in the *Connection Log*] “Who the problem affects.” Oh, okay. Who the problem affects. So, like...

Justin: Farmers? Ranchers?

Aaron: Everyone who lives next to it.

*Connection Log: To save information* During the interview, the students mentioned that the *Connection Log* can help them save their work. When asked about their opinion of the *Connection Log* Melissa said: “It helped us so that we could write down all of our information and... keep it in one spot so we could go back and look at it.”

### *Group E2: Experimental condition*

Group E2 was composed of Kristen (scribe), Billy, and Brad. Their stakeholder position was Farmers and Ranchers of Monroe (the next state south).

*Epistemological criteria* Kristen and the pair of Brad and Billy differed in their approach to assessing the quality of information. In the unit, students collected data from different areas of the Dale River, and sometimes one measurement (e.g., turbidity) differed substantially from the same measurement taken by other students at the same location. When encountering such data, Kristen “just skip[ped] around it”. She noted in the interview, “If it wasn’t like seventh graders, I’d probably trust it.” When asked if he followed the same approach, Billy shook his head. In the end, the group ignored outlier data. To assess the trustworthiness of information, Kristen judged the motives of the website creators. Brad and Billy ignored sites about rivers in other regions.

## Challenges

*Getting going* At the start of the unit, members of Group E2 struggled to understand their goal. For example, Kristen noted a need to define the problem in the *Connection Log*, but did not understand why. In the end, Group E2 members' problem statements were very similar.

Soon after defining the problem, the group experienced technical problems with their *Connection Log* accounts. This caused some delays in proceeding with unit work.

*Staying on track* Billy and Brad often got sidetracked. As a result, Kristen often needed to urge Brad and Billy to do assigned tasks.

### *What they used for support and why*

*Teacher and groupmate support: To learn how to operate the Connection Log.* When Group E2 members struggled, they tended to turn to different sources of support depending on the nature of their difficulty. When they did not know how to respond to *Connection Log* prompts, they tended to seek help from the teacher and groupmates.

*Connection Log: To guide their inquiry* The unit was paused for one week during hunting season. Remaining students were given index cards to write down key information to build their argument. Group E2 members noted, "With the cards, it was ... kind of just guessing. But with the *Connection Log*, it would tell us what to do and how to do it."

Without the *Connection Log*, Jake noted that during the unit they probably would have just "wing[ed] it." But as it were, they needed to define the problem and determine information to find and find it. It was not immediately apparent why this was useful, but near the end of the unit it became clearer, as they could "go back and find your information to be able to do claims and to get your information to be able to have the speech ready."

*Connection Log: To articulate ideas that could be critiqued* Students sought critique of articulated ideas from groupmates and the teacher. For example, after entering claims, Kristen told the teacher, "I don't know if we're done or what to do next." After reading the claims, the teacher noted, "Okay, the first one doesn't really have a full claim." He asked Kristen to explain her position, and then told her the elements of a claim. Then she worked further.

In another example, Brad typed in Found Information, "The dam is slowing the river and is causing less D. O. [dissolved oxygen] so if we get rid of the dam the river will get faster and have more D. O." Billy read this, and noted, "I don't think we're gonna cause them to... blow up the dam." Brad agreed. The idea of removing the dam did not make it into the presentation.

*Connection Log: To revisit ideas* The teacher often prompted students to go back through their argument in the *Connection Log*. When they did, they discussed ideas and often found room for improvement in their argument. For example, on Day 13, Kristen told Researcher 2, "I don't know if this is right" while pointing at the claim she made in the *Connection Log*. Researcher 2 looked at the evidence that Kristen had associated with the claim and said "Your information title needs to be descriptive enough that you can say the pollution is high with waste and car parts." Kristen asked for clarification: "So it would be what we found?" When Researcher 2 said yes, Kristen went back to the found information stage to elaborate.

### *Group E3: Experimental condition*

Taylor (scribe), Eric, and Angie (Group E3) represented environmentalists. In general, Taylor asked questions while Eric answered them.

*Epistemological criteria* In the interview, group members described the importance of good evidence to a solid argument. Eric noted one needs evidence to prove something.

### *Challenges*

*Understanding stakeholder position* The group struggled to understand what environmentalists want. They knew that environmentalists supported a healthy environment, but were unsure what was considered “healthy.” On Day 3, they discussed the problem definition:

Eric: How does it affect stakeholders?

Taylor: They affect me. But it’s affecting them.

Eric: You can’t swim in it.

The group understood poorly what environmentalists want from the river on Day 3. The group struggled for two days, and by the end of Day 5 had a better understanding of environmentalists’ viewpoints, but the topic was never addressed again until their presentation.

*Staying on track* The students often got off track, especially playing with their computers’ screen savers.

### *What they used for support and why*

*Teacher and groupmate support: To refocus* When the group became stuck, one of them would ask “Why is the river dirty?” to get back on track. For example, on Day 8 the group was discussing phosphates. Taylor asked what it was and neither Eric nor Angie could answer.

Taylor: Hey, what is it? Phosphate? What’s phosphate?

...

Angie: I don’t know.

Eric: We need information. Why is the river dirty? Hey guys, go back.

The question, “Why is the river dirty?” helped the group refocus on the task at hand.

*Connection Log: To guide inquiry* The *Connection Log* was integral to helping the group members accomplish tasks. Most often, they discussed tasks they were completing in the *Connection Log*. When asked if the *Connection Log* was helpful, all three group members responded yes, and commented they would use it again on a future project.

*Connection Log: provide a coherent system for organizing data* During the unit, Taylor used the *Connection Log* to gather data from groupmates. In the interview, group members noted that the *Connection Log* helped them keep data and evidence organized.

### *Group C1*

Group C1 was composed of Melissa, Megan, Allison, and Jon, and represented the stakeholder position of farmers in Madison. Megan was the scribe.

*Epistemological criteria* The group judged Internet sources based on whether the sources agreed with their beliefs. Megan said proving something means “that you believe in it” and that people will accept her ideas because she “looked on the Internet, and that took forever.”



## Challenges

*Understanding stakeholder* The group failed to understand their stakeholder position at the outset of the project. The students seemed to come from an environmentalist perspective. While turbidity and dissolved oxygen levels were not ideal, they did little to affect farmers. The group argued in favor of measures that were costly to farmers (e.g., discontinuing use of herbicides, not using land near the river) to mitigate problems that did not affect the farmers. Despite repeated efforts to point out this discrepancy, the group described such measures in their final presentation. In this passage, Mr. Thomas asked Alison about her part of the argument:

Mr. Thomas: And what is yours about?

Allison: Spraying weeds. And we need to stop, so the river....

Mr. Thomas: But you are the ones....

Allison: That we need to stop!

Mr. Thomas: Do you realize how many farmers are going to kill you when you say that?

...

Allison: But we could put sod down.

Mr. Thomas: Sod?!

Allison: Yeah, I looked it up. And then we don't have to spray, and then the river will be clean, and THEN we can irrigate!

Even after multiple discussions with Researcher 2, Researcher 3, and Mr. Thomas, the students failed to grasp that their stakeholders would prefer to not make any changes.

*Coordinating their argument* While there was some coordination, individual group members largely took one part of the subject matter and developed their own arguments without ensuring consistency with the rest of the group. Megan advocated for new irrigation methods to prevent excess phosphates, nitrates, and salts from entering the water, Allison proposed laying sod on fields before planting to create a weed barrier, and Melissa spoke about possible nitrate poisoning of livestock. As phosphate, nitrate, and salt levels were all too low to have any substantial impact on farmers and ranchers, all of these arguments and solutions conflicted with the best interests of the stakeholders. Jon's conclusion stated that the water quality was relatively good and did not affect farmers and ranchers enough to merit action. While this argument aligned with the interests of farmers and ranchers, it contradicted his groupmates' arguments.

## What they used for support and why

*Groupmate support: Find useful information from lives of group members* The group's strongest argument for discontinuing herbicide use was Allison's grandfather's success using herbicide less frequently than other local farmers. Though initially skeptical, Mr. Thomas was persuaded of the strategy's viability because Allison specified how it worked for an actual farmer.

*Teacher support: Procedural questions* A large proportion of questions they asked the teacher and research group members were procedural in nature, such as regarding the use of student-collected data and computers, and effective information-finding strategies.

*Teacher support: Assess quality of information and strategy* The group often lacked confidence about their abilities to complete the unit, and needed frequent reassurance and guidance to move forward productively. Megan asked about this more than other members.

*Teacher support: Focus effort* The teacher and research group members prompted the group's focus and directed their efforts. For example, Allison spent some time working on her argument about the negative effects of phosphates on cows. When Researcher 3 asked her what damage phosphates do and where she got her information, she realized that she had no evidence that phosphates are harmful to livestock. Then, she changed her focus to pesticides.

### *Cross-case analysis*

One would do well to notice that while some members of some groups demonstrated advanced epistemological understanding, or at least epistemological understanding that approached what one would need to effectively engage in argumentation, there was sometimes inconsistency within groups. Furthermore, at times certain elements of their epistemological approach that was less than desirable. For example, Kristen (of Group E2) recognized that it was best to ignore outliers in the data, but her reasoning for when to ignore outliers was less than desirable, as she noted that she skipped outliers because the readings were from seventh graders, noting that she would have investigated outliers found by adults more thoroughly. That view was not shared by her groupmates. Members of Groups E1 and E3 demonstrated somewhat advanced epistemological understanding in the interview, but this clashed with their presentations. Members of Group C1 demonstrated my-side bias in that they saw sources as valid if the latter agreed with their personal beliefs. Also telling is that Group C1's belief that their argument should be believed because they worked so hard on it.

Each group faced different challenges, with the exception that members of Group E2 and Group E3 struggled to stay on task. Members of different groups used the *Connection Log* to accomplish different goals, except that (a) both Group E1 and Group E2 used the *Connection Log* to articulate ideas that could then be critiqued, and (b) both Group E2 and Group E3 used the *Connection Log* to guide their inquiry. In the latter case, this similar use of the *Connection Log* may be traced to both groups struggling to stay on track. There also was variation in how different groups used teacher and groupmate support. For example, members of Group C1 largely relied on the teacher to assess the quality of information and group process. This makes sense in that their ability to assess the quality of information was less than ideal.

How can argumentation scaffolds be redesigned to provide stronger support?

### *Help students know that they need to provide a solution*

As previously noted, students can seek to use computer-based scaffolds in manners that align with their goals (Belland and Drake 2013). But students cannot use scaffolds in ways that the scaffolds do not support. Many students struggled to understand that they needed to provide a solution to the problem, rather than simply describe the problem. The teacher was sometimes able to help with this, but often the teacher would explain the need to provide a solution, and then many students would either still not understand that they needed to do so, or simply fail to provide a solution.

As we noted in a paper analyzing other data in the larger data set, Mr. Thomas performed well in providing one-to-one scaffolding support when compared to more experienced teachers in the literature (Belland et al. 2015). But constant reminders to provide solutions did not appear to be very productive. Furthermore, providing such reminders takes away from the crucial role of one-to-one scaffolding as a tool to evaluate and extend student reasoning (Belland 2014; van de Pol et al. 2010). Similar to the idea of using video cases to present PBL problems (Chan et al. 2010), in future iterations of the unit, the county commissioner will speak at the beginning of the unit explaining (a) what the students are doing in the unit and why, and (b) the need to create an argument in support of a problem solution at the end of the unit. We also updated *Connection Log* prompts to emphasize the need to create arguments supporting solutions, rather than explanations of the problem.

### *Help students become familiarized with what their audience expects*

To help students understand what the audience of the argument is looking for, we are adding a stage at the beginning of the *Connection Log* in which students need to identify (a) the audience, (b) what is important to the audience, (c) why the audience cares about the problem, and (d) what evidence the audience needs to support a problem solution.

### *Add running example*

Because many students struggled to use the *Connection Log* at the beginning of the unit, we added a running example related to student groups arguing about climate change, in which each student group represents a stakeholder group. The hypothetical students want to present their argument to the President of the United States. Students can see an overview of the argument and also what hypothetical students would write while using each page of the *Connection Log*. We chose climate change because it is a scientific issue (a) with which many students are familiar, and (b) about which fallacious arguments are often advanced. In the example, students can see how students arguing about climate change can (a) define the problem, (b) describe the audience, identify what is known and what needs to be known and how to find it, (c) evaluate credibility of web sites, (d) make claims, and (e) link evidence to claims.

## **Discussion**

### **Influence on argument evaluation ability among student subgroups**

As NGSS are implemented, K-12 teachers will need to think about how to help all of their students excel at scientific argumentation and solving authentic scientific problems (Achieve 2013; Wilson 2013). To this end, educators need to find evidence-based methods to help culturally diverse students and students who struggle succeed (P. Cuevas et al. 2005; Lynch 2001). Thus, it is interesting that we found greater influences among lower-achieving and low-SES students than among higher-achieving and average to high-SES students, respectively. At the same time, it is important to address the fact that we found no significant impact on argument evaluation gain scores among higher-achieving students.

### *Influence according to prior science achievement*

Argument evaluation test results indicated that experimental students with low prior science achievement benefitted significantly and substantially more than their control counterparts. In addition, there was no significant difference between higher-achieving students and their control counterparts. This is in line with previous research (Belland 2010; Belland et al. 2011) but the magnitude of the effect is twice as large. It is arguably more important that computer-based scaffolding impact lower-achieving than higher-achieving students. This aligns with a finding that teacher scaffolding was not enough to lead to substantial gains among lower-achieving students in content knowledge on the relationship between land use and water quality indicators (Azevedo et al. 2004). It also is in line with findings that the magnitude of the effect of reflective assessment on physics project scores was greater among lower-achieving than among higher-achieving middle school students (White and Frederiksen 1998). It goes against findings that middle school students need to be familiar with content before learning about how to effectively argue about it (Aufschnaiter et al. 2008).

When thinking about why the *Connection Log's* effectiveness varied according to achievement level, one can consider the expertise reversal effect, according to which the same instructional support that is effective among lower-achieving students may not be effective among higher-achieving students because its guidance is redundant to strategies that the higher-achieving students already have mastered (Roelle and Berthold 2013; Salden et al. 2010). Thus, when students engage in learning tasks or assessments, they need to consider both their mastered strategies and the newly acquired strategies in performing the task. This in turn leads to lower learning efficiency, in that high-ability learners need to attend to instructional messages that provide guidance on redundant strategies (Schnotz 2010). There is some evidence that one may avoid expertise reversal and achieve more consistent results among all students through performance-adapted customization (e.g., fading and adding) of scaffolding (Salden et al. 2010). However, a recent meta-analysis did not find that customization of scaffolding led to differences in average effect sizes (Belland et al. 2014).

One way expertise reversal may be explained is through an examination of what students needed to do: articulate answers to questions individually, and then come to consensus with groupmates. Group C1 was not alone in the control condition in not communicating well amongst themselves about their explanations and work. Control students mostly articulated explanations by themselves, and ran such explanations by the teacher. In Rivard (2004), lower-achieving eighth-grade students retained most in biology knowledge when they needed to discuss ecology problems and write explanations, rather than simply write explanations or discuss problems orally. In contrast, higher-achieving students performed best when they simply needed to write explanations (Rivard 2004). By also encouraging higher-achieving students to discuss explanations, the *Connection Log's* prompts may have led to less learning efficiency among higher-achieving students, as it may have interfered with what such students would have done with already mastered strategies (Schnotz 2010). This is a possible explanation of why there was no significant difference in gain scores between higher-achieving experimental and higher-achieving control students. Further research is needed.

The positive impact on argument evaluation ability among lower-achieving students may also relate to the experience of critiquing groupmates' responses as prompted by the *Connection Log*. The *Connection Log* encourages students to come to consensus with groupmates after articulating responses to prompts individually. Members of two small

groups selected for case studies from the experimental condition used the *Connection Log* to articulate ideas such that they could be critiqued. Members of Group C1 simply told their ideas to the teacher to receive feedback on the quality of the ideas. This lack of experience evaluating ideas may be why lower-achieving control students gained less on argument evaluation than lower-achieving experimental students, and is akin to research that shows that lack of experience creating arguments likely indicates low argument creation ability (Abi-El-Mona and Abd-El-Khalick 2011; Jonassen and Kim 2010). But this does not indicate why there was no difference among higher-achieving students. A possible explanation may be found in the literature on epistemology and argumentation. Typical 8th grade students in Israel were found to have an overall awareness of the difference between explanation and evidence (Glassner et al. 2005). Higher-achieving students may have been more familiar with argumentation norms than lower-achieving students, and had less room for improvement. Research indicates that middle school students with good understanding of argumentation norms will perform better at argument evaluation tasks than such students with poor understanding of argumentation norms (Weinstock et al. 2004).

Another explanation for the lack of a significant difference on argument evaluation gains between higher-achieving experimental and control students involves the possibility that *Connection Log* prompts interfered with higher-achieving students' pre-existing strategies for critiquing each other's articulated thoughts. This in turn could have led to poorer group functioning. The level of group functioning has been found to be the greatest predictor of achievement among high achieving 8th grade students (Webb et al. 2002). Simply put, the group functioning of lower-achieving students may have been enhanced by the encouragement of critiquing their groupmate's thoughts, while the same was not true among higher-achieving students. If true, this is further evidence of an expertise reversal effect. Further research is needed.

The lower-achieving experimental students had the highest argument evaluation posttest score among all student subgroups. While this is consistent with some prior research, it is important to note that not all computer-based scaffolding research has similar findings. For example, scaffolding to promote collaborative knowledge building was found to benefit both high-achieving and low-achieving elementary school students, but the magnitude of the effect was greater among high-achieving students (So et al. 2010).

#### *Influence among low-SES students*

We found preliminary evidence of a difference in influence of computer-based scaffolds according to socio-economic status. However, much caution is warranted in interpreting this finding, as this was not a statistically significant difference. This finding makes sense when interpreted in light of the literature (Cuevas et al. 2005; Lynch et al. 2005). Still, the exact mechanism by which differences in influence among low and average to high SES students arises is not clear from the literature. Describing and understanding differences in influence among diverse students is clearly important, as success must be promoted among all students (Boehner 2001; Lynch 2001). Further research is needed.

#### *Another potential reason behind differences in influence among student subgroups*

Not all students will interact with scaffolds in the same way, but rather may use scaffolds in ways aligned with their goals (Akhras and Self 2002; Belland and Drake 2013). These goals may influence what students get out of using the scaffold. For example, if students

use scaffolds to help them evaluate resources, then that may help students develop critical evaluation skills.

It is critical to note that students did not use the *Connection Log* to overcome all challenges. For example, Group E1 struggled to interpret the data. As it is a generic scaffold, the *Connection Log* could not support students in that task. Thus, Group E1 needed to rely on the contingent scaffolding of Mr. Thomas for support on this challenge. While Mr. Thomas provided teacher scaffolding well, there was only one of him and many students. This is in line with research on affordances that emphasizes that tools cannot have an unlimited number of affordances (Osieurak et al. 2010; Young et al. 2002).

#### *Lack of significant main effects on argument evaluation ability and group argument quality*

Our failure to find significant main effects could be due to several factors. First, the *Connection Log* may truly be effective among lower-achieving students, and ineffective among average-to-higher achieving students. If so, the lack of significant main effect would make sense. A larger sample size in future studies may elucidate this issue. But there are other possible explanations. First, experimental students needed to be trained in how to use the *Connection Log*. During the time it took experimental students to learn to use the *Connection Log*, control students could already begin to address the unit problem. Also, technical issues caused delays in experimental students' work. Thus, control students had essentially more time on task. Measuring exactly how much is not feasible because different groups and different students within groups faced different technical issues with the *Connection Log*. But on average it was likely 0.75 days (out of 10 days to analyze data and create arguments). Future research would benefit from training control students on a placebo software program. Next, in a separate study on a different part of the same dataset, we found that the teacher provided twice as much one-to-one scaffolding to control students as he did to experimental students. One-to-one scaffolding is one of the most powerful instructional interventions (VanLehn 2011), so control students may have received essentially the same amount of scaffolding support as experimental students.

#### Epistemological criteria and argumentation

Skills needed to succeed in argumentation include constructing and evaluating arguments (Perelman and Olbrechts-Tyteca 1958; van Eemeren et al. 2002). Epistemological beliefs critically influence this (Kuhn and Udell 2007; Weinstock et al. 2004). When examining the small groups selected for case studies, one notices that the experimental groups in general had more sophisticated epistemological criteria after using the *Connection Log* than did the control group at similar points in the unit.

#### *Epistemological criteria and argument evaluation*

To evaluate arguments effectively, one needs advanced epistemological understanding and practice (Driver et al. 2000; Kuhn and Udell 2007). While the case studies were not intended to describe the entirety of each group's respective class period, they can provide some insight. Members of Group C1 demonstrated beginning epistemological understanding in that they noted that people should believe their argument because it aligned with their personal beliefs and they worked hard. This is similar to what Hogan and Maglienti (2001) found among eighth grade students, who judged claims about a fictional

plant as well-founded when the latter meshed with prior beliefs. Applied to argument evaluation tasks, the same reasoning would lead to my-side bias (Glassner et al. 2005; Kuhn and Udell 2007), which causes one to not be open to exploring alternative hypotheses (Kyza 2009).

The epistemological beliefs of the experimental small groups described in the cases studies were closer to what would be considered sophisticated epistemological beliefs. For example, members of Group E2 looked for consistency in data, which is a criterion professional scientists use (Hogan and Maglienti 2001). Members of Groups E1 and E3 stressed the need for evidence, which is promising. Still, Kristen (Group E2)'s rationale for ignoring outliers because they were recorded by seventh graders could potentially lead to attacking arguments due to perceived untrustworthiness of the arguer (Weinstock et al. 2004). Epistemological beliefs are important because they influence ill-structured problem solving and argumentation abilities (Oh and Jonassen 2007). Also, while members of Group E1 espoused epistemic beliefs that were approaching sophistication, those beliefs were not always reflected in their actions. Inconsistency in application of epistemological beliefs was also found among high school students investigating microevolution among finches on the Galápagos Islands (Sandoval 2003).

### *Epistemological criteria and argument quality*

The case study component was not intended to provide a comprehensive description of all students, but it can provide some insight. Some evidence for beginning epistemological beliefs can be seen in most arguments. With more sophisticated epistemological beliefs, students may have produced more sophisticated arguments. For example, knowing that knowledge is not certain would likely lead one to acknowledge a need for careful construction of arguments (Chinn and Malhotra 2002; Weinstock et al. 2004). Nonetheless, the views of the control group from the case studies were concerning: that people should believe their argument because it was based on their personal beliefs and it took a long time to develop. A belief that arguments are good if they are based on personal beliefs is common among unaided middle school students (Hogan and Maglienti 2001; Yoon 2011).

Suboptimal epistemological beliefs in both conditions may not tell the whole story. The lack of significant differences in argument quality between conditions may have resulted from differences in teacher scaffolding provided. In another study on this unit but focused on teacher performance, we found that the teacher provided twice as much one-to-one scaffolding to control students as he did to experimental students (Belland et al. 2015). Teacher scaffolding is one of the biggest influences on learning and performance (Ruiz-Primo and Furtak 2006; VanLehn 2011), and can lead to an effect size of about 0.8—a large effect (VanLehn 2011).

### *Student desire to use the software in the future*

Our findings indicated that at least some of the students would like to use the software in the future. This is important in that motivation to learn is an essential prerequisite to effective learning (Pajares 1996; Wigfield and Eccles 2000), but motivation to use scaffolding is often lacking due to failure to account for motivation during scaffold design (Belland et al. 2013) and the fact that scaffolding often makes learning tasks harder (Reiser 2004; Simons and Ertmer 2006). Still, if scaffolding works, by definition, students should not need to use it again in the future as a result of having gained sufficient skill in the

scaffolded task. Thus, one might question whether students wanting to use scaffolding again is desirable. Further research is needed.

## Limitations and suggestions for future research

### Delimitations

The fact that the sample was not randomly selected from a greater population causes challenges for generalization to a greater population. But, the idea that one can arrive at a context-free generalization in educational research is questionable (Johnson and Onwuegbuzie 2004). It may be more fruitful to consider particularization of the case—the extent to which sufficient detail was provided about content so that readers can determine if the lessons learned in the study apply to a new context of interest (Eisenhart 2009; Stake 1978). From this perspective, it is important to note that half of the students lived on farms and this study was conducted in a rural setting in the Intermountain west of the USA—a desert region. This may have led the present participants to think differently about water quality than students who live in urban and/or non-desert environments. Next, the students had spent most of their schooling to date in teacher-centered classrooms, and thus PBL was entirely new to them. Students who have had more experience in student-centered learning environments would likely respond differently.

### Limitations due to implementation challenges

The unit was paused for 1 week due to hunting season, as over half of the students were absent. It took time once the unit started back up for students to rehabilitate to the unit. The students had never learned in a problem-centered format before, thus the process and goal of PBL were new to them. We will account for hunting season when scheduling future units.

Statistical power, especially for the comparison of argument quality, was very low. This could not be changed except to conduct a study at a different school, as participants included virtually the entire 7th grade student body of the participating school. With more participants, more robust statistical models such as generalized estimating equations could have been used.

Audio quality from Days 3 and 5 was poor due to microphone difficulties. On Day 3 students were defining the problem, and Day 5 students began to interpret the water quality data. In future research we will take care to verify microphone functionality well before unit start.

Due to technical difficulties, some recordings were missing from the middle of the unit—3 days out of 11 for Group E2, and for 1 day out of 11 for Group E1. This may have caused us to miss some instances of the groups interpreting the water quality data. We asked students about what they did on those days in the interview, and we also had access to what experimental students wrote in the *Connection Log* on those days, as well as log data.

The *Connection Log* had some technical issues that caused some students to not be able to use it at times. This led to some confusion and lost entries. When Mr. Thomas troubleshooted technical issues, he could not help students with content or process-related issues.



Using free and reduced lunch data to classify students according to SES relies on parents self-reporting income to schools and thus can suffer from presentation bias (Cruse and Powers 2006). Future research should explore alternative measures of familial SES.

## Conclusion

Improving middle school students' argumentation skill is central to preparing them for success in the 21st century (Abi-El-Mona and Abd-El-Khalick 2011). In this paper, we found a significant and substantial impact of computer-based scaffolding on the argument evaluation ability of lower-achieving students. We also found that students used the available support—computer-based scaffolding, teacher scaffolding, and groupmate support—in different ways to counter differing challenges. We made changes to the scaffolds on the basis of research results.

**Acknowledgments** This research was supported by Early CAREER Grant 0953046 from the National Science Foundation, awarded to the first author. Any opinions, findings, and/or conclusions are those of the authors and do not necessarily represent official positions of NSF.

## References

- Abi-El-Mona, I., & Abd-El-Khalick, F. (2011). Perceptions of the nature and “goodness” of argument among college students, science teachers, and scientists. *International Journal of Science Education*, 33(4), 573–605. doi:10.1080/09500691003677889.
- Achieve. (2013). Next generation science standards. Retrieved August 8, 2013, from <http://www.nextgenscience.org/next-generation-science-standards>.
- Akhras, F. N., & Self, J. A. (2002). Beyond intelligent tutoring systems: Situations, interactions, processes and affordances. *Instructional Science*, 30(1), 1–30. doi:10.1023/A:1013544300305.
- Azevedo, R., Winters, F. I., & Moos, D. C. (2004). Can students collaboratively use hypermedia to learn science? The dynamics of self- and other-regulatory processes in an ecology classroom. *Journal of Educational Computing Research*, 31(3), 215–245. doi:10.2190/HFT6-8EB1-TN99-MJVQ.
- Bell, P. (1997). Using argument representations to make thinking visible for individuals and groups. In R. Hall, N. Miyake, & N. Enyedy (Eds.), *Proceedings of CSCL'97: The second international conference on computer support for collaborative learning* (pp. 10–19). Toronto: University of Toronto Press.
- Belland, B. R. (2010). Portraits of middle school students constructing evidence-based arguments during problem-based learning: The impact of computer-based scaffolds. *Educational Technology Research and Development*, 58(3), 285–309. doi:10.1007/s11423-009-9139-4.
- Belland, B. R. (2014). Scaffolding: Definition, current debates, and future directions. In J. M. Spector, M. D. Merrill, J. Elen, & M. J. Bishop (Eds.), *Handbook of research on educational communications and technology* (4th ed., pp. 505–518). New York: Springer.
- Belland, B. R., Burdo, R., & Gu, J. (2015). From lecturing to scaffolding: One middle school science teacher's transformation. *Journal of Science Teacher Education*. doi:10.1007/s10972-015-9419-2.
- Belland, B. R., & Drake, J. (2013). Toward a framework on how affordances and motives can drive different uses of computer-based scaffolds: Theory, evidence, and design implications. *Educational Technology Research and Development*, 61, 903–925. doi:10.1007/s11423-013-9313-6.
- Belland, B. R., Glazewski, K. D., & Richardson, J. C. (2008). A scaffolding framework to support the construction of evidence-based arguments among middle school students. *Educational Technology Research and Development*, 56(4), 401–422. doi:10.1007/s11423-007-9074-1.
- Belland, B. R., Glazewski, K. D., & Richardson, J. C. (2011). Problem-based learning and argumentation: Testing a scaffolding framework to support middle school students' creation of evidence-based arguments. *Instructional Science*, 39(5), 667–694. doi:10.1007/s11251-010-9148-z.
- Belland, B. R., Kim, C., & Hannafin, M. (2013). A framework for designing scaffolds that improve motivation and cognition. *Educational Psychologist*, 48(4), 243–270. doi:10.1080/00461520.2013.838920.

- Belland, B. R., Walker, A., Kim, N., & Lefler, M. (2014). A preliminary meta-analysis on the influence of scaffolding characteristics and study and assessment quality on cognitive outcomes in STEM education. In *Presented at the 2014 Annual Meeting of the Cognitive Science Society, Québec City*.
- Blumer, H. (1969). *Symbolic interactionism: Perspective and method*. Englewood Cliffs: Prentice Hall.
- Boehner, J. A. H.R.1—No Child Left Behind Act of 2001, Pub. L. No. 107-110 (2001). Retrieved from <http://www.gpo.gov/fdsys/pkg/PLAW-107publ110/html/PLAW-107publ110.htm>.
- Buckland, L. A., & Chinn, C. A. (2010). Model-evidence link diagrams: a scaffold for model-based reasoning. In *Proceedings of the 9th International Conference of the Learning Sciences—Volume 2* (pp. 449–450). Chicago, IL, USA: International Society of the Learning Sciences. Retrieved from <http://dl.acm.org/citation.cfm?id=1854509.1854741>.
- Census. (2014). *State Median Income*. Retrieved from <https://www.census.gov/hhes/www/income/data/statemedian/>.
- Chan, L. K., Patil, N. G., Chen, J. Y., Lam, J. C. M., Lau, C. S., & Ip, M. S. M. (2010). Advantages of video trigger in problem-based learning. *Medical Teacher*, 32(9), 760–765. doi:10.3109/01421591003686260.
- Chinn, C. A., & Malhotra, B. A. (2002). Epistemologically authentic inquiry in schools: A theoretical framework for evaluating inquiry tasks. *Science Education*, 86(2), 175–218. doi:10.1002/sc.10001.
- Cho, K., & Jonassen, D. H. (2002). The effects of argumentation scaffolds on argumentation and problem-solving. *Educational Technology Research and Development*, 50(3), 5–22. doi:10.1007/BF02505022.
- Clark, D. B., & Sampson, V. D. (2007). Personally-seeded discussions to scaffold online argumentation. *International Journal of Science Education*, 29(3), 253–277. doi:10.1080/09500690600560944.
- Creswell, J. W., & Miller, D. L. (2000). Determining validity in qualitative inquiry. *Theory Into Practice*, 39(3), 124–130. doi:10.1207/s15430421tip3903\_2.
- Cruse, C., & Powers, D. (2006). *Estimating school district poverty with free and reduced-price lunch data*. Retrieved from <http://www.census.gov/did/www/saipe/publications/files/CrusePowers2006asa.pdf>.
- Cuevas, H. M., Fiore, S. M., & Oser, R. L. (2002). Scaffolding cognitive and metacognitive processes in low verbal ability learners: Use of diagrams in computer-based training environments. *Instructional Science*, 30(6), 433–464. doi:10.1023/A:1020516301541.
- Cuevas, P., Lee, O., Hart, J., & Deaktor, R. (2005). Improving science inquiry with elementary students of diverse backgrounds. *Journal of Research in Science Teaching*, 42(3), 337–357. doi:10.1002/tea.20053.
- Driver, R., Newton, P., & Osborne, J. (2000). Establishing the norms of scientific argumentation in classrooms. *Science Education*, 84(3), 287–312. doi:10.1002/(SICI)1098-237X(200005)84:3.
- Eisenhart, M. (2009). Generalization from qualitative inquiry. In K. Ercikan & W.-M. Roth (Eds.), *Generalizing from educational research: Beyond qualitative and quantitative polarization* (pp. 51–66). New York: Routledge.
- Ford, M. J. (2012). A dialogic account of sense-making in scientific argumentation and reasoning. *Cognition and Instruction*, 30(3), 207–245. doi:10.1080/07370008.2012.689383.
- Gelman, A. (2004). Exploratory data analysis for complex models. *Journal of Computational and Graphical Statistics*, 13(4), 755–779. doi:10.1198/106186004X11435.
- Giere, R. N. (1990). *Explaining science: A cognitive approach*. Chicago: University of Chicago Press.
- Glassner, A., Weinstock, M., & Neuman, Y. (2005). Pupils' evaluation and generation of evidence and explanation in argumentation. *British Journal of Educational Psychology*, 75, 105–118. doi:10.1348/000709904X22278.
- Hmelo-Silver, C. E. (2004). Problem-based learning: What and how do students learn? *Educational Psychology Review*, 16(3), 235–266. doi:10.1023/B:EDPR.0000034022.16470.f3.
- Hmelo-Silver, C. E., & Barrows, H. S. (2006). Goals and strategies of a problem-based learning facilitator. *Interdisciplinary Journal of Problem-Based Learning*, 1(1), 21–39. doi:10.7771/1541-5015.1004.
- Hogan, K., & Maglienti, M. (2001). Comparing the epistemological underpinnings of students' and scientists' reasoning about conclusions. *Journal of Research in Science Teaching*, 38(6), 663–687. doi:10.1002/tea.1025.
- Johnson, R. B., & Onwuegbuzie, A. J. (2004). Mixed methods research: A research paradigm whose time has come. *Educational Researcher*, 33(7), 14–26. doi:10.3102/0013189X033007014.
- Jonassen, D. H., & Kim, B. (2010). Arguing to learn and learning to argue: Design justifications and guidelines. *Educational Technology Research and Development*, 58(4), 439–457. doi:10.1007/s11423-009-9143-8.
- Klosterman, M. L., Sadler, T. D., & Brown, J. (2011). Science teachers' use of mass media to address socio-scientific and sustainability issues. *Research in Science Education*, 42(1), 51–74. doi:10.1007/s11165-011-9256-z.

- Kolstø, S. D. (2001). Scientific literacy for citizenship: Tools for dealing with the science dimension of controversial socioscientific issues. *Science Education*, 85(3), 291–310. doi:10.1002/sce.1011.
- Kuhn, D. (1991). *The skills of argument*. Cambridge: Cambridge University Press.
- Kuhn, D. (2010). Teaching and learning science as argument. *Science Education*, 94(5), 810–824. doi:10.1002/sce.20395.
- Kuhn, D., & Udell, W. (2007). Coordinating own and other perspectives in argument. *Thinking & Reasoning*, 13(2), 90–104. doi:10.1080/13546780600625447.
- Kyza, E. A. (2009). Middle-school students' reasoning about alternative hypotheses in a scaffolded, software-based inquiry investigation. *Cognition and Instruction*, 27(4), 277–311. doi:10.1080/07370000903221718.
- Loyens, S. M. M., Magda, J., & Rikers, R. M. J. P. (2008). Self-directed learning in problem-based learning and its relationships with self-regulated learning. *Educational Psychology Review*, 20(4), 411–427. doi:10.1007/s10648-008-9082-7.
- Luria, A. R. (1976). *Cognitive development: Its cultural and social foundations* (M. Lopez-Morillas & L. Solotaroff, Trans., M. Cole, Ed.). Cambridge, MA, USA: Harvard University Press.
- Lynch, S. (2001). "Science for all" is not equal to "one size fits all": Linguistic and cultural diversity and science education reform. *Journal of Research in Science Teaching*, 38(5), 622–627. doi:10.1002/tea.1021.
- Lynch, S., Kuipers, J., Pyke, C., & Szesze, M. (2005). Examining the effects of a highly rated science curriculum unit on diverse students: Results from a planning grant. *Journal of Research in Science Teaching*, 42(8), 912–946. doi:10.1002/tea.20080.
- Miles, M. B., & Huberman, A. M. (1984). Drawing valid meaning from qualitative data: Toward a shared craft. *Educational Researcher*, 13(5), 20–30. doi:10.3102/0013189X013005020.
- Nicolaidou, I., Kyza, E. A., Terzian, F., Hadjichambis, A., & Kafouris, D. (2011). A framework for scaffolding students' assessment of the credibility of evidence. *Journal of Research in Science Teaching*, 48(7), 711–744. doi:10.1002/tea.20420.
- Nunnally, J. C. (1978). *Psychometric theory*. New York, NY: McGraw-Hill.
- Oh, S., & Jonassen, D. H. (2007). Scaffolding online argumentation during problem solving. *Journal of Computer Assisted Learning*, 23(2), 95–110. doi:10.1111/j.1365-2729.2006.00206.x.
- Osborne, J. (2010). Arguing to learn in science: The role of collaborative, critical discourse. *Science*, 328(5977), 463–466. doi:10.1126/science.1183944.
- Osiurak, F., Jarry, C., & Le Gall, D. (2010). Grasping the affordances, understanding the reasoning: Toward a dialectical theory of human tool use. *Psychological Review*, 117(2), 517–540. doi:10.1037/a0019004.
- Pajares, F. (1996). Self-efficacy beliefs in academic settings. *Review of Educational Research*, 66(4), 543–578. doi:10.3102/00346543066004543.
- Patton, M. Q. (2002). *Qualitative research & evaluation methods* (3rd ed.). Thousand Oaks: Sage Publications.
- Perelman, C., & Olbrechts-Tyteca, L. (1958). *The new rhetoric: Treatise on argumentation*. Paris: Presses Universitaires de France.
- Quintana, C., Krajcik, J., & Soloway, E. (2003). Issues and approaches for developing learner-centered technology. *Advances in Computers*, 57, 271–321. doi:10.1016/S0065-2458(03)57006-1.
- Reiser, B. J. (2004). Scaffolding complex learning: The mechanisms of structuring and problematizing student work. *Journal of the Learning Sciences*, 13(3), 273–304. doi:10.1207/s15327809jls1303\_2.
- Rivard, L. P. (2004). Are language-based activities in science effective for all students, including low achievers? *Science Education*, 88(3), 420–442. doi:10.1002/sce.10114.
- Roelle, J., & Berthold, K. (2013). The expertise reversal effect in prompting focused processing of instructional explanations. *Instructional Science*, 41(4), 635–656. doi:10.1007/s11251-012-9247-0.
- Rose, S. L., & Barton, A. C. (2012). Should great lakes city build a new power plant? How youth navigate socioscientific issues. *Journal of Research in Science Teaching*, 49(5), 541–567. doi:10.1002/tea.21017.
- Ruiz-Primo, M. A., & Furtak, E. M. (2006). Informal formative assessment and scientific inquiry: Exploring teachers' practices and student learning. *Educational Assessment*, 11(3–4), 205–235. doi:10.1080/10627197.2006.9652991.
- Sadler, T. D., & Donnelly, L. A. (2006). Socioscientific argumentation: The effects of content knowledge and morality. *International Journal of Science Education*, 28(12), 1463–1488. doi:10.1080/09500690600708717.
- Salden, R. J. C. M., Aleven, V., Schwonke, R., & Renkl, A. (2010). The expertise reversal effect and worked examples in tutored problem solving. *Instructional Science*, 38(3), 289–307. doi:10.1007/s11251-009-9107-8.

- Sandoval, W. A. (2003). Conceptual and epistemic aspects of students' scientific explanations. *Journal of the Learning Sciences*, 12(1), 5–51. doi:10.1207/S15327809JLS1201\_2.
- Saye, J. W., & Brush, T. (2002). Scaffolding critical reasoning about history and social issues in multimedia-supported learning environments. *Educational Technology Research and Development*, 50(3), 77–96. doi:10.1007/BF02505026.
- Scheuer, O., Loll, F., Pinkwart, N., & McLaren, B. (2010). Computer-supported argumentation: A review of the state of the art. *International Journal of Computer-Supported Collaborative Learning*, 5(1), 43–102. doi:10.1007/s11412-009-9080-x.
- Schnotz, W. (2010). Reanalyzing the expertise reversal effect. *Instructional Science*, 38(3), 315–323. doi:10.1007/s11251-009-9104-y.
- Schraw, G. (2001). Current themes and future directions in epistemological research: a commentary. *Educational Psychology Review*, 13(4), 451–464. doi:10.1023/A:1011922015665.
- Sheskin, D. J. (2011). *Handbook of parametric and nonparametric statistical procedures* (5th ed.). Boca Raton: CRC Press.
- Simons, K. D., & Ertmer, P. A. (2006). Scaffolding disciplined inquiry in problem-based learning environments. *International Journal of Learning*, 12(6), 297–305.
- So, H.-J., Seah, L. H., & Toh-Heng, H. L. (2010). Designing collaborative knowledge building environments accessible to all learners: Impacts and design challenges. *Computers & Education*, 54(2), 479–490. doi:10.1016/j.compedu.2009.08.031.
- Stake, R. E. (1978). The case study method in social inquiry. *Educational Researcher*, 7(2), 5–8. doi:10.3102/0013189X007002005.
- Stanovich, K. E., & West, R. F. (2008). On the failure of cognitive ability to predict myside and one-sided thinking biases. *Thinking & Reasoning*, 14(2), 129–167. doi:10.1080/13546780701679764.
- Tal, T., & Kedmi, Y. (2006). Teaching socioscientific issues: Classroom culture and students' performances. *Cultural Studies of Science Education*, 1(4), 615–644. doi:10.1007/s11422-006-9026-9.
- U.S. Department of Health & Human Services. (2012). *2012 HHS poverty guidelines*. Retrieved from <http://aspe.hhs.gov/poverty/12poverty.shtml>.
- Van de Pol, J., Volman, M., & Beishuizen, J. (2010). Scaffolding in teacher–student interaction: A decade of research. *Educational Psychology Review*, 22(3), 271–296. doi:10.1007/s10648-010-9127-6.
- Van Eemeren, F. H., Grootendorst, R., & Snoeck Henkemans, A. F. (2002). *Argumentation: Analysis, evaluation, presentation*. Mahwah: Lawrence Erlbaum Associates.
- VanLehn, K. (2011). The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems. *Educational Psychologist*, 46(4), 197–221. doi:10.1080/00461520.2011.611369.
- Vellom, R. P., & Anderson, C. W. (1999). Reasoning about data in middle school science. *Journal of Research in Science Teaching*, 36(2), 179–199. doi:10.1002/(SICI)1098-2736(199902)36:2.
- von Aufschnaiter, C., Erduran, S., Osborne, J., & Simon, S. (2008). Arguing to learn and learning to argue: Case studies of how students' argumentation relates to their scientific knowledge. *Journal of Research in Science Teaching*, 45(1), 101–131. doi:10.1002/tea.20213.
- Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Cambridge: Harvard University Press.
- Walker, K. A., & Zeidler, D. L. (2007). Promoting discourse about socioscientific issues through scaffolded inquiry. *International Journal of Science Education*, 29(11), 1387–1410. doi:10.1080/09500690601068095.
- Webb, N. M., Nemer, K. M., & Zuniga, S. (2002). Short circuits or superconductors? Effects of group composition on high-achieving students' science assessment performance. *American Educational Research Journal*, 39(4), 943–989. doi:10.3102/01623737017002239.
- Weinstock, M., Neuman, Y., & Tabak, I. (2004). Missing the point or missing the norms? Epistemological norms as predictors of students' ability to identify fallacious arguments. *Contemporary Educational Psychology*, 29(1), 77–94. doi:10.1016/S0361-476X(03)00024-9.
- White, B. Y., & Frederiksen, J. R. (1998). Inquiry, modeling, and metacognition: Making science accessible to all students. *Cognition and Instruction*, 16(1), 3–118. doi:10.1207/s1532690xci1601\_2.
- Wigfield, A., & Eccles, J. S. (2000). Expectancy-value theory of achievement motivation. *Contemporary Educational Psychology*, 25(1), 68–81. doi:10.1006/ceps.1999.1015.
- Wilson, S. M. (2013). Professional development for science teachers. *Science*, 340(6130), 310–313. doi:10.1126/science.1230725.
- Wood, D., Bruner, J. S., & Ross, G. (1976). The role of tutoring in problem solving. *Journal of Child Psychology and Psychiatry*, 17(2), 89–100. doi:10.1111/j.1469-7610.1976.tb00381.x.
- Yoon, S. A. (2011). Using social network graphs as visualization tools to influence peer selection decision-making strategies to access information about complex socioscientific issues. *Journal of the Learning Sciences*, 20(4), 549–588. doi:10.1080/10508406.2011.563655.

- Young, M. F., DePalma, A., & Garrett, S. (2002). Situations, interaction, process and affordances: An ecological psychology perspective. *Instructional Science*, 30(1), 47–63. doi:[10.1023/A:1013537432164](https://doi.org/10.1023/A:1013537432164).
- Zhang, B., Liu, X., & Krajcik, J. S. (2006). Expert models and modeling processes associated with a computer-modeling tool. *Science Education*, 90(4), 579–604. doi:[10.1002/sce.20129](https://doi.org/10.1002/sce.20129).

**Brian R. Belland** is an Associate Professor of Instructional Technology and Learning Sciences at Utah State University. His research focuses on the use of technology to scaffold middle school students' argumentation during problem-based science units, and is supported by two National Science Foundation grants—an early CAREER grant and a REESE Grant. He has published in prominent journals including *Educational Psychologist*, *Computers & Education*, *Educational Psychology Review*, *Educational Technology Research & Development*, and *Instructional Science*.

**Jiangyue Gu** is a PhD candidate in the Department of Instructional Technology and Learning Sciences at Utah State University. Her research interests focus on the use of technology to scaffold middle school's arguments, and its influence on and interaction with middle school students' epistemic beliefs.

**Sara Armbrust** is a junior majoring in elementary education and early childhood education at Utah State University.

**Brant Cook** is a senior majoring in agricultural education at Utah State University.