

Problem-based learning and argumentation: testing a scaffolding framework to support middle school students' creation of evidence-based arguments

Brian R. Belland · Krista D. Glazewski · Jennifer C. Richardson

Received: 17 October 2009 / Accepted: 28 August 2010 / Published online: 10 September 2010
© Springer Science+Business Media B.V. 2010

Abstract Students engaged in problem-based learning (PBL) units solve ill-structured problems in small groups, and then present arguments in support of their solution. However, middle school students often struggle developing evidence-based arguments (Krajcik et al., *J Learn Sci* 7:313–350, 1998). Using a mixed method design, the researchers examined the use of computer-based argumentation scaffolds, called the *Connection Log*, to help middle school students build evidence-based arguments. Specifically we investigated (a) the impact of computer-based argumentation scaffolds on middle school students' construction of evidence-based arguments during a PBL unit, and (b) scaffold use among members of two small groups purposefully chosen for case studies. Data sources included a test of argument evaluation ability, persuasive presentation rating scores, informal observations, videotaped class sessions, and retrospective interviews. Findings included a significant simple main effect on argument evaluation ability among lower-achieving students, and use of the scaffolds by the small groups to communicate and keep organized.

Keywords Evidence-based arguments · Scaffolding · Middle school · Science education · English as a New Language

Introduction

Science is a way of thinking and acting that incorporates both approaches to investigating phenomena and interpreting and communicating the results of investigations (Bransford

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

K. D. Glazewski
Department of Curriculum and Instruction, New Mexico State University,
MSC 3CUR, P.O. Box 30001, Las Cruces, NM 88003-8001, USA

J. C. Richardson
Department of Curriculum and Instruction, Purdue University, West Lafayette, IN 47907, USA

et al. 2000; Hawkins and Pea 1987). Central to thinking like a scientist is the ability to create evidence-based arguments, or claims (e.g., deer populations are increasing exponentially because their natural predators are displaced) connected to supporting evidence (e.g., trends in wolf and deer populations) via premises (e.g., when predator populations are healthy, prey populations are healthy and sustainable; Bricker and Bell 2008; Driver et al. 1998; Perelman and Olbrechts-Tyteca 1958; van Eemeren et al. 2002). This is because the communication and support of ideas is crucial to the science inquiry process and the process by which scientific claims are evaluated (Bricker and Bell 2008).

One way to help students develop scientific argumentation ability is to have students engage in argumentation in science classrooms (Bricker and Bell 2008; Driver et al. 1998; Kuhn 2005). However, little science instruction at the K-12 level involves argumentation (Driver et al. 1998; Kuhn 2005). Centering science curricula on competing views of the nature of reality may allow the development of scientific argumentation ability as well as other knowledge about the scientific process (Apple 1975; Driver et al. 1998). At the heart of competing views on the nature of reality are ill-structured problems, or problems with (a) unclear or incomplete descriptions, (b) unclear problem elements (e.g., stakeholders, actions, and consequences), (c) many ways to approach the problem, (d) many possible solutions of which many can be “correct,” and (e) multiple solution evaluation criteria (Driver et al. 1998; Jonassen 2003; Osborne et al. 2004). There is no one right solution to an ill-structured problem, so scientists need to provide evidence supporting their proposed solution.

To change alternative conceptions and promote better scientific thinking, including argumentation, many authors have proposed that students engage in ill-structured problem solving in science classrooms (Alberts 2003; Chang and Barufaldi 1999; Duggan and Gott 2002; Guzetti et al. 1993; Kuhn 2005; Palincsar et al. 1993). One instructional approach that involves solving ill-structured problems, problem-based learning (PBL), is being used increasingly in K-12 contexts (Gallagher 1997; Torp and Sage 1998).

Creating effective arguments requires abstract thinking because learners must consider both rules of logic and the perspectives of audience members who may expect different evidential support for claims. In addition it requires that learners recognize that scientific theories are not universal truths but rather can be supported or not through argumentation. Many middle school students are not able to do this independently because they do not fully possess the ability to think abstractly (Berland and Reiser 2008; Inhelder and Piaget 1955; Krajcik et al. 1998; Kuhn et al. 1997; Sandoval and Reiser 2004) and they adhere to an objectivist/absolutist epistemology (Hofer and Pintrich 1997; Kuhn 2005; Sandoval and Morrison 2003). Furthermore, they lack the ability to engage in systems thinking, or think of scientific problems in terms of effects on systems rather than individual things (Inhelder and Piaget 1955; Kuhn et al. 2000). Unaided middle school students' inability to construct evidence-based arguments when participating in a PBL unit could be attributed to three major challenges: (1) adequately representing the central problem (Ge and Land 2004; Liu and Bera 2005), (2) determining and obtaining the most relevant evidence (Berland and Reiser 2008; Pedersen and Liu 2002–2003), and (3) synthesizing the information gathered to construct a sound argument (Cho and Jonassen 2002). Students participating in PBL units often represent problems based on surface-level details in the initial problem description (Ge and Land 2004; Jonassen 2003). Since PBL problems are ill-structured, surface-level details do not provide all information necessary to understand all problem elements and how they interact. These challenges are problematic because students who do not fully understand a problem cannot identify appropriate evidence to seek (Krajcik et al. 1998; Simons and Klein 2007). The information they do collect may not support their

claim (Krajcik et al.). For example, middle school students participating in a PBL unit with the driving question, “What is happening to the deer population in Cache County, Utah, and what can be done to control the problem?,” would likely think the evidence they need to collect simply includes deer populations in different years and hunting regulations. Unaided, they would not likely consider that the problem involves an interaction between an imbalance in the predator–prey relationship and the expansion of cities.

Research results do not support the effectiveness of direct instruction in helping students improve their argumentation skills (Knudson 1991; Marttunen and Laurinen 2001). Furthermore, it may be crucial for students to learn argumentation skills in the context of authentic problem solving (Clark and Sampson 2007; Kuhn 2005). To support this process, several researchers have used scaffolding to help K-12 students build evidence-based arguments (e.g., Bell 1997; Kyza and Edelson 2005; Liu and Bera 2005; Sandoval and Reiser 2004). Wood et al. (1976) defined scaffolding as tutoring, in which teachers or other more capable individuals provide support to students participating in an activity that they could not complete unaided. The goal of scaffolding is two-fold: (a) to provide temporary support to students as they perform tasks that they have difficulty performing unaided, and (b), to help students gain competency in the scaffolded tasks such that they *can* perform the tasks unaided (Puntambekar and Hübscher 2005). Hard scaffolds (computer or paper-based cognitive tools) can serve the same roles as soft scaffolds (scaffolding provided by more capable others), and are developed based on students’ anticipated needs during a PBL unit (Saye and Brush 2002). The emergence of hard scaffolds has helped teachers provide conceptual, metacognitive, procedural, and strategic support to students as they solve ill-structured problems (Hannafin et al. 1999; Saye and Brush).

Scaffolding is intended to bridge the gap between what students cannot do independently and what they can do independently (Wood et al. 1976). However, few studies on science scaffolding investigate the differential impact of scaffolds on students of differing ability levels. This may be because scaffolding to support students’ efforts during inquiry units has tended to be developed for and investigated with high- or average-achieving students (White and Frederiksen 1998). However, it seems logical that the gap between what students cannot do independently and what they can do independently would differ between lower-achieving and higher-achieving students, and thus scaffolding would have differential impacts between students of differing ability levels. Some authors have examined the differential impact of other scaffolds on students of different ability levels (e.g., Cuevas et al. 2002). For example, Cuevas et al. found that the use of diagrams to scaffold reading comprehension had a significant effect on students’ ability to apply knowledge learned in reading passages to new scenarios only among students with low verbal ability.

In a recent paper, we described the process by which middle school students create evidence-based arguments during PBL units and proposed guidelines for the creation of hard scaffolds to support middle school students’ creation of evidence-based arguments (Belland et al. 2008). To create evidence-based arguments, students (a) represent the problem, (b) analyze the audience, (c) determine needed information and gather it, (d) develop claim, and (e) gather additional evidence and link evidence to claim (Belland et al. 2008). Using the process and the guidelines (a) embed scaffolds within a system, (b) have students articulate their thoughts, and (c) focus on the development of conceptual, strategic, and procedural hard scaffolds, we designed and developed non-context bound hard scaffolds, called the *Connection Log*, to help middle school students build evidence-based arguments (Belland et al. 2008). The purpose of this study is to document the impact and use of the *Connection Log*.

Research questions

1. What is the impact of hard scaffolds on argument evaluation ability?
2. What is the impact of hard scaffolds on argument quality?
3. How and why do middle school science students use hard scaffolds to construct an argument while participating in a PBL unit?

Method

Research design

We used a mixed method design in order to address different question types. Quantitative and qualitative methods were used concurrently at all stages of the study: research design/data collection, data analysis, and data interpretation (Onwuegbuzie et al. 2004).

Quantitative method

The research questions, “What are hard scaffolds’ impact on individual argument evaluation ability?” and “What are hard scaffolds’ impact on group argument quality?” were primarily addressed by quantitative methods. We used a two-factor nested experiment (Giesbrecht and Gumpertz 2004). As is common in K-12 settings, treatment and control groups were based upon intact classrooms. As such, students were nested within four different class sections. The independent variable was scaffold condition (explained in “[Independent variables](#)” section).

Qualitative method

To address the question “How and why do middle school students use hard scaffolds to construct an argument while participating in a PBL unit,” we investigated in depth how the members of a small group—one each from the lower- and higher-achieving experimental conditions—worked together and used the scaffolds. We designed and conducted this investigation through the lens of symbolic interactionism. Kinney et al. (2003) argued that one cannot fully understand what goes on in schools without considering the interactive processes by which students engage in tasks. One way researchers can investigate interactive processes in schools is by using the lens of symbolic interactionism (Kinney et al.). Two premises are fundamental to understanding symbolic interactionism. First, social reality (social dynamics of organizations such as classrooms or schools) is not externally imposed on individuals, but constructed through their interactions with other individuals (Sandstrom and Fine 2003; Stryker 2001). Second, the way people *interact* with *symbols*, defined as other people or things, is dependent on the *meaning* that they assign to the symbols (Blumer 1969). The meaning that they assign to symbols is influenced by both past and present interactions with the symbol and other symbols, and the extent to which those interactions responded to the needs resulting from their challenges.

Setting and participants

Participants

The setting was four seventh-grade science class sections in a low-SES (about half of the student body received free or reduced lunch) school located in a small Midwestern community. The school had laptop carts with enough laptops for an entire class.

The teacher identified two classes as higher- and two as lower-achieving based on their performance (grades on tests and impressions) during the school year up to the unit time, which was in February. The lower-achieving class means on a pretest of argument evaluation ability (see the “[Argument evaluation ability](#)” section within data collection for a description of the test) were significantly lower than those of the higher-achieving classes, $F(1, 80) = 8.16, p < 0.01, ES = 0.61$. There were no significant differences in pretest scores between the two lower-achieving class sections, $F(1, 80) = 1.14, p = 0.29, ES = 0.28$. There were no significant differences in pretest scores between the two higher-achieving classes, $F(1, 80) = 0.004, p = 0.95, ES = 0.07$. Thus, there were two matched higher-achieving classes and two matched lower-achieving classes. We randomly assigned one higher- and one lower-achieving class to each condition: scaffold and no scaffold. Students in the experimental condition received the additional support of the *Connection Log*, computer-based scaffolds designed to support students’ creation of evidence-based arguments (described in the procedures and “[Independent variables](#)” sections). The teacher had a master’s degree and 4 years’ experience facilitating PBL. By the time of the unit, most students had experienced PBL units both the academic year of the study (2006–2007) and the previous year. Eighty-six students from four class sections taught by the same teacher participated in the unit.

Unit

Data were gathered during a 2-week PBL unit on the Human Genome Project (HGP). The driving question was “How can a \$3 million grant be used to expand on or protect the public from the findings of the human genome project?” Student groups chose stakeholder positions (e.g., adopted children). The unit problem was ill-structured in that there were multiple valid solution paths and solutions for any given stakeholder group. For example, adopted children could propose to use the grant money to improve methods for adopted children to find their biological parents or to fund genetic testing for adopted children since they don’t know what “runs in their family.” Such solutions are not inherently right or wrong but can only be evaluated based on evidential support.

Qualitative subsample

One group was chosen for in-depth examination from each period that used the scaffolds. The group from the lower-achieving section was composed of Robert, Alejandra, and Erin (note: all names have been changed). The teacher noted they were all low-achieving students based on their grades to that point in the school year. The group seemed particularly information-rich because they were likely to face communication problems: Robert and Erin were both native English speakers, while Alejandra was an English as a New Language (ENL) student with low oral English proficiency who had arrived in the US 6 months before (her native language was Spanish). The group from the higher-achieving section consisted initially of two members, Daniel and Megan. The teacher noted that each

was a high-achieving student based on their grades to that point in the school year. We chose the group because Daniel and Megan provided a contrast to the Lower-Achieving Group in that they communicated proficiently. Also, Daniel was outgoing while Megan was reserved. Claudia, an ENL student (her native language was Spanish) who was high-achieving, joined the group 3 days before the persuasive presentation because her original partner was not able to finish the unit. We predicted that each group would face different challenges, and thus, our theoretical frame indicated they would use computer-based scaffolds in a different manner.

Independent variables

Connection Log

There were two levels of this independent variable: *Connection Log* and no *Connection Log*. Two class sections were randomly assigned to use the *Connection Log* during the PBL unit. The other two class sections completed the same unit without the support of the *Connection Log*.

The *Connection Log* is a database-driven website designed to scaffold middle school students' creation of evidence-based arguments during PBL units. At the basis of the *Connection Log* is a conceptual framework (Belland et al. 2008) that defines (a) evidence-based arguments, (b) difficulties that middle school students have in creating evidence-based arguments, and (c) the process by which students create evidence-based arguments in PBL units (e.g., develop claim), and provides guidelines for the development of hard scaffolds to support the process (e.g., embed scaffolds within a system and have students articulate their thoughts). Central to the framework is that scaffolding can be developed to counter student difficulties with generic processes rather than content-specific questions. The first author designed the *Connection Log* during an instructional strategies seminar to support middle school students in the completion of the steps in the process of creating evidence-based arguments. First, an instructional analysis was conducted that indicated the types of supports needed in the scaffolds. Then a storyboard, or paper-based representation of the scaffolds, was created. The seminar professor (the second author) reviewed the instructional analysis once and the storyboard twice. Suggestions on scaffold simplification were incorporated. Ten 7th grade science students (who were not participants in this study) explained in their own words the actions prompted by the storyboard in order to assess readability. A seventh grade science teacher suggested language simplifications. When words could not be simplified, students could see definitions by positioning their mouse over unfamiliar words. Figures 1 and 2 are screenshots of the *Connection Log*.

Students register for the *Connection Log* in groups. The scaffolds are organized in six stages corresponding to the steps in the argument creation process (Belland et al. 2008):

1. *Define problem*, in which students state the problem in their own words and then come to consensus on a definition broken into stakeholder, what is happening, and how it affects stakeholders categories.
2. *Determine needed information*, in which students decide on information about the problem that they need to find, and strategies for finding it.
3. *Find needed information*, in which students find and record needed information.
4. *Organize information*, in which students organize found information according to the categories of stakeholder, what is happening, and how it affects stakeholders and decide if information is relevant.


Connection Log		What is a Claim?	Make a Claim	Compare with Group Mates		
Stages Define Problem > Determine Needed Info > Find Needed Info > Organize Info > Develop Claim > Link Evidence to Claim >	Develop Claim: Make a Claim <i>By Myself</i> Here is the relevant evidence that you have so far:					
	Stakeholder Info	<input type="text"/>				
	What is Going On	<input type="text"/>				
	How it Affects the Stakeholders	<input type="text"/>				
	Other Info	<input type="text"/>				
You are logged in as: You are currently on stage: Develop Claim 		Do you have enough evidence now to make a claim or two? If not, please go back to the determine info we need to know stage by clicking on the link to the left. If you think you can make a claim either about what is going on or what can be done to improve the situation, type the claim(s) in the box below. If you type more than one claim, please press enter after each one. <input type="text"/>				

Fig. 1 Screenshot of Stage 5, Step 2 of the *Connection Log*. In the previous step, students learned what a claim is in the context of an evidence-based argument

5. *Develop claim*, in which students assert a possible problem solution.
6. *Link evidence to claim*, in which students link specific, relevant data to their assertions and build an argument.

Each stage consists of 2–4 steps that are performed either individually or as a group. Each step requires articulation of responses to questions: Students type their responses and click “next” to record the responses to the database. In Step 1, students individually type responses to a question (e.g., determine information to find based on the group problem definition). In Step 2, students read what their groupmates wrote and the group comes to consensus on a response to the question (e.g., what the group should find). As part of this process, students (a) read guidelines about appropriate claims, evidence, or arguments, depending on the stage, (b) break their groupmates’ and their own input into categories (pertaining to stakeholder, what is happening, and how it affects the stakeholders), and (c) debate in their group to develop a consensus response that incorporates elements on which they all agree. In Steps 3 or 4 (if present), groups assign tasks to individual members such as information to find and/or determine strategies to accomplish tasks. When students attempt to skip a stage or a step, they receive an error message. Additionally, on steps where groupmate responses are not submitted, students receive a message to encourage their groupmates to submit their answers. The teacher additionally encouraged students to use the scaffolds when they asked questions

Connection Log

Select Evidence to Support Claims | Test Evidence | Break Down Claims | Put it All Together

Stages

- Define Problem >
- Determine Needed Info >
- Find Needed Info >
- Organize Info >
- Develop Claim >
- Link Evidence to Claim >

Link Evidence to Claims: Break Down Claims

With My Group Mates

You may have found that you could not find evidence that really supports a claim. You may not be able to find any evidence that directly supports your claim for three reasons:

- Your claim is too broad, and needs to be broken down
- You don't have enough evidence
- Your claim is just too far off-base

The chances are your claim is just too broad and needs to be broken down into sub-claims. Breaking a claim into subclaims is a lot like breaking a long sentence into several shorter sentences. So if your claim were "rainforests are being deforested because people buy a lot of wooden furniture," you might break the claim into at least two subclaims:

- The manufacture of wooden furniture is one of the biggest uses of rainforest wood
- If people didn't buy as much wooden furniture, demand for wood from the rainforest would decrease.

Below are your group's claims. **The scribe** should copy the claim or claims that you could not support, and paste it in the bottom box. Try to break it down into subclaims and type them in the bottom box.

You are logged in as:

You are currently on stage:
Link Evidence to Claim

Here are your group's claims:

As a group, talk about breaking the claims into sub-claims. **The scribe** should type your subclaims here:

Fig. 2 Screenshot of Stage 6, Step 3 of the *Connection Log*

that were answered in the scaffolds. Generic questions work because they use information previously entered by students (e.g., problem definition) in subsequent questions (e.g., about what further information is needed).

As an example of the use of the *Connection Log*, Student A representing private citizens in a PBL unit may write a claim that the HGP should be stopped, and attempt to support it with evidence that genetic information can cause people to be denied insurance companies. But his groupmate may claim that the HGP benefits private citizens because it can help people determine if they would pass on genes from a genetic disorder to their children. The groupmates would need to come to consensus on their claim and on how to construct the argument in the *Connection Log* after articulating their ideas.

Data collection

Argument evaluation ability

One dependent variable was argument evaluation ability, defined as ability to evaluate evidential support for a claim. An argument evaluation test was adapted with permission

Jenny and Angie argue about whether cell phone use by drivers causes accidents. Jenny says, "Cell phone use by drivers causes accidents." Angie asks, "Why?"
 To what extent do the following statements help Jenny prove that cell phone use by drivers causes accidents:

	Doesn't help	Helps a little	Helps a lot
"A large percentage of car accidents happen when at least one driver is talking on a cell phone"	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
"Drivers can't concentrate on the road when they are talking on a cell phone"	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Fig. 3 An example item from argument evaluation test

from the test used by Glassner et al. (2005). The original test was composed of six claims and two statements each that were meant to support an argument goal (prove or explain the claim). In both the original test and the modified version, students needed to indicate (a) the extent to which the statements supported the argument goal, and (b) which statement best supported the argument goal. Modifications that were made included: (1) as we were only interested in students' evaluation of proof, we dropped the explanation goal, (2) due to time constraints, we dropped two claims, and (3) to make the test more relevant to American middle school students we changed three claims. See Fig. 3 for an example item. Cronbach's α coefficients of reliability of the pretest and posttest were 0.70 and 0.77, respectively. A middle school science teacher and a middle school language arts teacher reviewed the test for content validity, and they noted that it tested the content it was supposed to test, and in a manner understandable to middle school students. Additionally, two seventh grade students, one advanced fifth-grade student, and one eighth-grade student reviewed the test to provide evidence of readability and face validity. They said that they did not have difficulty reading the test, and that they thought it measured students' ability to distinguish between good and bad arguments.

Group argument quality

The dependent variable was group argument quality. Video of the persuasive presentation was transcribed, and two raters blind to treatment condition rated the entire transcript, and then met to come to consensus. The rubric (see Appendix Table 5) was developed based on our theoretical framework, which holds that an evidence-based argument consists of a clear claim connected with relevant, supporting evidence via premises. Using the rubric, raters assigned numerical scores for claim, evidence, and connection of claim to evidence quality. Initial interrater agreement was 0.63 as measured by Cohen's Kappa. Then the raters met to come to consensus.

How and why students used the scaffolds

Consistent with the theoretical frame of symbolic interactionism, we collected data to determine the meaning that the scaffolds held to students, and how that related to how students used the scaffolds. We (a) videotaped each group during the unit, and transcribed verbatim all dialogue, (b) retrieved what each group member typed in the *Connection Log*, (c) conducted prompted, retrospective interviews of approximately 30 min each with each group, and (d) engaged in informal observations of student problem solving approaches during the entire unit. In each interview a unique, approximately 20-min video containing

scenes from the videotaped class sessions prompted participants' recollection of how they used the *Connection Log* and why. In the informal observations, we looked for ineffective student approaches (e.g., multiple group members searching for the same things at the same time), off-task behaviors, and progressions in student ideas. This range of objective and subjective data types contributes to the trustworthiness of conclusions (Lather 2003).

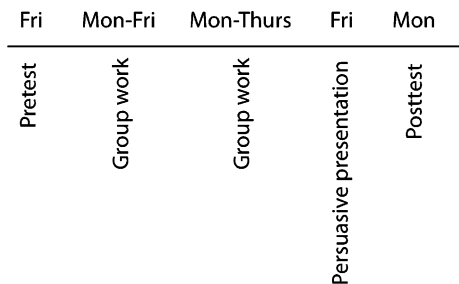
Procedures

Events proceeded according to the timeline in Fig. 4. The pretest was given on the Friday before the unit started. Students worked in groups of 2–3 for 50 min per day for 9 days. Students in the experimental condition could use the *Connection Log* (described in materials) as an additional means of support while students in the control condition could not. The teacher facilitated the unit for 9 days, and in the final day each group made a persuasive presentation. On Day 1, the teacher explained the unit goals and what students needed to do to accomplish the goals (e.g., work collaboratively with group mates, use time wisely, and apply learned information to own experience and values). The teacher explained that during the persuasive presentation students had to (1) introduce themselves (from the perspective of their stakeholder group), (2) state their purpose, (3) state three points about the HGP as they pertain to their stakeholder group, (4) state their use of the grant money, and (5) summarize their presentation.

On subsequent days, students worked in small groups: after assuming the perspective of their assigned stakeholder groups, they pursued learning issues (e.g., stakeholder positions and goals related to the HGP) to begin to understand how members of their stakeholder groups perceived the HGP and what they could do with the fictional 3 million dollar grant to further their stakeholder group's goals in relation to the HGP. For example, the doctor group could decide that (1) more specific types of research should be done on genes that cause particular chronic diseases, and (2) the \$3 million grant could be used to benefit everyone by engaging in that type of research. When they neared the end of the unit, students began creating promotional brochures and arguments to use in the persuasive presentation. Most groups developed posters to support their arguments, and some developed additional materials such as charts and fictional business cards to hand to the judge.

On the last day of the unit, each group had 4 min to present its arguments in a persuasive presentation and the grant winner was decided based on the judge's rating of the strength of the arguments. Each group's persuasive presentation was filmed. The day after the unit's completion, the posttest measure was administered. The time span between pre and posttest was 17 days. Then the persuasive presentations were rated by two raters blind to treatment condition according to the rubric in Appendix Table 5.

Fig. 4 Timeline of study



Data analysis

Argument evaluation scores

The data set did not meet the assumptions for ANCOVA because Levene's test indicated inequality of variances. Therefore, we compared argument evaluation scores using nested ANOVA. If nested effects were significant, simple main effects were calculated (Keppel 1982). The sample size was not sufficient to perform more integrated latent variable analyses that provide error-free estimates of model parameters and effect sizes. To compensate for the inability to conduct integrated latent-variable analyses, effect sizes were corrected for attenuation due to reliability. As noted by Hunter and Schmidt (2004), observed effect sizes are lower than actual effect sizes due to less than perfect reliability. Effect sizes corrected for attenuation are equal to the observed effect size divided by the square root of the test reliability.

Group argument quality ratings

We compared ratings using nested MANOVA. If effects were significant, we ran follow-up ANOVAs. If nested effects were significant, we calculated simple main effects (Keppel 1982).

How and why students used the scaffolds

We followed the iterative process of qualitative research described by Miles and Huberman (1984) of data collection, data reduction, data display, and conclusion-drawing/verification. In the phase of data reduction, all data was coded. We conducted coding, as with all analysis, through the lens of symbolic interactionism. For example, we focused much coding on identifying challenges faced by students and the extent to which the *Connection Log* addressed the challenge and/or was recognized as helping by the students. The coding scheme was created through a literature review and the stages and steps of the *Connection Log* in a three-step process. First, we created an initial coding scheme from theory-based argument creation processes and middle school student problem-solving processes. Then, we reviewed the transcripts to modify the coding scheme to account for patterns manifest in the data. Subsequently, we applied the evolving coding scheme to all transcripts. Upon coding completion, we developed themes by looking at text and video formats of all passages to which a code was applied.

Data display was conducted to lead to initial conclusions but also to revise the coding scheme and thus the initial conclusions. In data display, researchers display data graphically (e.g., graphs, causal networks) to establish potential causal paths and make sense of the data (Miles and Huberman 1984). Then we generated and verified conclusions. In the verification process we paid particular attention to parsimony of the conclusions in light of our theoretical frame, or the extent to which the conclusions concisely and accurately explained the meaning the scaffolds held to the participants and the way participants used the scaffolds.

Preliminary conclusions arose during the data reduction and data display stages, but they remained tentative as we progressed through these stages. The preliminary conclusions evolved as we triangulated by searching in the data displays and coded data for confirming and contrary data and in a range of different data types (video, interview, database, informal observations, persuasive presentation ratings) (Glaser and Strauss 1967;

Lather 2003). We also checked validity of conclusions by examining the meaning of outliers (Miles and Huberman 1984).

Results

What is the impact of hard scaffolds on argument evaluation ability?

The main effect of the *Connection Log* on argument evaluation ability was significant at $\alpha = 0.1$, $F(1, 82) = 2.99$, $p = 0.09$, $ES = 0.35$. After correction for attenuation, the effect size was 0.41, which means that the average student who used the *Connection Log* scored 0.41 standard deviations better on the argument evaluation posttest than the average student who did not use the *Connection Log*.

There was a significant nested effect of the *Connection Log* on argument evaluation ability, $F(2,82) = 6.48$, $p < 0.01$. As can be seen in Fig. 5, the magnitude of the effect of the *Connection Log* differed between subgroups. Partition of the nested effect showed that *Connection Log* had a significant simple main effect on the posttest performance of students from the lower-achieving classes, $F(1,82) = 6.07$, $p = 0.01$, $ES = 0.61$. After correction for attenuation, the effect size among lower-achieving students was 0.69. In other words, the average lower-achieving student who used the *Connection Log* scored 0.69 SD higher than the average lower-achieving control student. The difference between the higher-achieving experimental and control students was not statistically significant, but the effect size favored the higher-achieving experimental students, $ES = 0.15$. After correction for attenuation, the effect size was 0.17. So the average higher-achieving student who used the *Connection Log* scored 0.17 standard deviations above the average higher-achieving student who did not use the *Connection Log*. Means and standard deviations for each class are listed in Table 1.

What is the impact of hard scaffolds on argument quality?

We found no significant main effect, $A = 0.88$, $F(3, 25) = 1.16$, $p = 0.34$, and no significant nested effect $A = 0.66$, $F(6, 50) = 1.95$, $p = 0.09$. As can be seen in Fig. 6, the experimental Higher-Achieving Group scored better on claim than the control Higher-Achieving Group, $ES = 0.56$. After correction for attenuation, the effect size was 0.7. So higher-achieving students who used the *Connection Log* performed 0.7 standard deviations

Fig. 5 Argument evaluation posttest means by achievement level and scaffolding condition

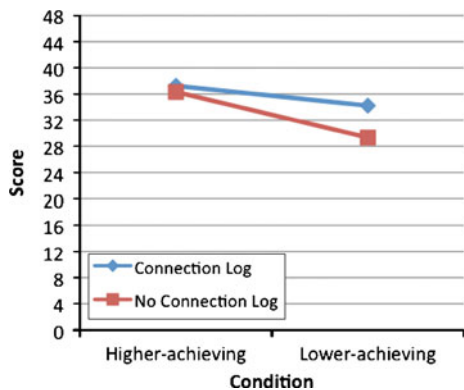
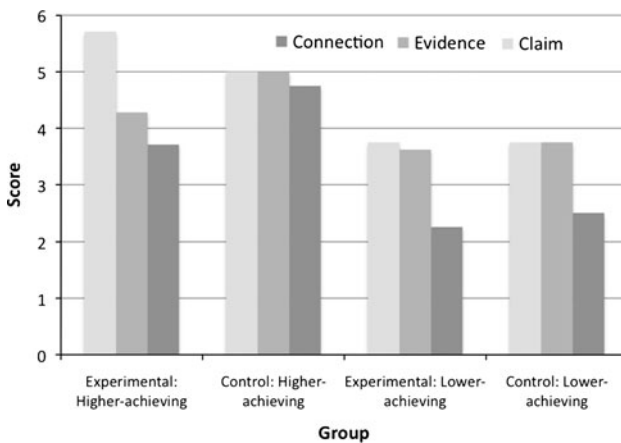


Table 1 Argument evaluation scores by period

Condition (higher or lower-achieving)	<i>n</i>	M	SD	SE
Experimental	41	35.24	6.1	0.95
Higher-achieving	15	37.26	4.35	1.12
Lower-achieving	26	34.23	6.74	1.32
Control	45	32.69	8.67	1.29
Higher-achieving	22	36.32	6.42	1.37
Lower-achieving	23	29.33	9.23	1.93

**Fig. 6** Argument quality scores by achievement level and condition

better than those who did not use the scaffolds. However, they did not fare as well on evidence and connection, where the uncorrected effect sizes were -0.6 and -0.72 , respectively. After correction for attenuation, the effect sizes were -0.75 and -0.91 respectively. Among the lower-achieving students, experimental students performed exactly the same on claim as control students, but slightly below control students on evidence and connection, where the uncorrected effect sizes were -0.07 and -0.18 respectively. After correction for attenuation, the effect sizes were -0.09 and -0.23 , respectively. Means and standard deviations are listed in Table 2.

How and why do middle school science students use hard scaffolds to construct an argument while participating in a PBL unit?

To address this question for each group, we first describe each member's experiences and their perceptions of challenges during the unit. This description is meant to illustrate the social realities of working within the group as experienced by each member. We then describe how and why the members used the *Connection Log*, how their use of the *Connection Log* may be attributed to the challenges they faced due to the social realities, how their thoughts progressed during the unit, and how their performance compares to

Table 2 Claim, evidence, and connection scores by period

Condition (higher-or lower-achieving)	<i>n</i>	Claim		Evidence		Connection	
		<i>M</i>	SD	<i>M</i>	SD	<i>M</i>	SD
<i>Experimental</i>							
Higher-achieving	15	5.71	0.75	4.28	1.38	3.71	1.38
Lower-achieving	26	3.75	1.98	3.62	1.5	2.25	0.49
<i>Control</i>							
Higher-achieving	22	5.0	1.51	5.0	1.07	4.75	1.49
Lower-achieving	23	3.75	1.28	3.75	1.98	2.5	1.77

Note: Maximum score = 6

each other and the control groups. Both the lower-achieving and the Higher-Achieving Groups chose the stakeholder position of bone marrow transplant doctors. After presenting the case studies of each selected group, we present a summary of the approaches used by groups from the experimental and control conditions, as indicated by informal observations.

Lower-Achieving Group

Robert perceived that the unit was one of the most difficult projects that year: “This was definitely one of the hardest I thought because it’s sort of like untouched territory. Like ‘cause I don’t know much about it.” Erin perceived that communication in the group was difficult. During several incidents during the unit, Robert appeared to either ignore or misunderstand a question from Erin. In addition, Robert often went to another part of the classroom, noting in the interview, “Sometimes I like try to be alone so I can think and meditate.” Sometimes Robert would leave to meditate after assigning tasks to Erin or Alejandra. Video evidence indicated that this often led to confusion, and Erin and Alejandra often talked to friends afterwards about unrelated topics, perhaps due to uncertainty about tasks. Alejandra’s limited proficiency in English also posed a problem. She noted: “I was listening... Because I was trying to help, but they always talk, and I was trying to listen, but Erin talks too fast.”

How they used the Connection Log Erin, Robert, and Alejandra used the *Connection Log* to (a) put down their ideas, (b) aid communication, and (c) think of ideas, as discussed in greater detail in the following paragraphs.

To put down their ideas. In the interview, Erin noted, and her groupmates agreed, that the group input their ideas into the *Connection Log*: “We put down our ideas, and what we thought, and the things we needed to find.” Erin said she would use the *Connection Log* for another similar project “because it helps you organize.”

Robert noted that recording their ideas in the *Connection Log* facilitated the debate of their ideas: “[The *Connection Log*] definitely [helped] with groups, ‘cause like if they told me something... and they put something different in [the *Connection Log*], I could ask ‘em about [what they meant].” Erin also appeared to benefit from using the scaffolds to present her ideas so that they could be debated, as in the following passage from Day 5:

Erin: [Showing what she typed in Stage 1 Step 2 of Connection Log] Robert can you help? I’ve tried a hundred different things but they don’t sound right.

Robert: Anyways. Cancer patients. What's happening? People are getting cancer every day, and then they're getting bone marrow. Well, I don't know, that sounds pretty good.

Because Erin articulated the problem definition ("People are getting cancer everyday [*sic*] and getting [*sic*] the Transplant, and it does not always work"), the group could weigh its merits.

This use of the *Connection Log* is interesting from a symbolic interactionist perspective because Erin noted that during past units ideas always sounded better in her head than when she communicated them orally. Through past experience she likely constructed a perception that when she communicated ideas orally they would never be as clear as they sound in her head. But, with the *Connection Log*, she could represent and revise her thoughts until they more closely represented what she intended. She and her groupmates appeared to construct a meaning of the *Connection Log* as a place where they could articulate ideas that could be correctly interpreted and debated.

To aid communication. As noted previously, oral communication between group members was sometimes difficult for several reasons. Alejandra noted that the *Connection Log* helped them communicate: "Well that [the *Connection Log*] helped me a lot because first Robert was like good and then he was not, and he didn't talk a lot, so that helped me know what they were thinking." Reading what Robert and Erin wrote helped Alejandra stay involved in the group and aware of what was happening because, as she noted, her reading and writing ability in English was more developed than her oral comprehension and expression.

From a symbolic interactionist perspective, the group's use of the *Connection Log* to aid communication is interesting because Erin and Alejandra noted that communication was one of the biggest challenges of the unit. The group members appeared to come to see and use the *Connection Log* as a tool that could help their communication.

To think of ideas. Robert said that he would use the *Connection Log* for a similar unit in the future because "it really helped me think of ideas... what we were supposed to look for." In this passage on Day 5, Robert talked with Erin about a prompt that asked students to write what else they needed to find:

Erin: Robert! We're supposed to put more information, like why are they getting cancer. Like, like, why their bone marrow is going bad.

Robert: Wait, I don't know. I know. Oh, we need to know why they are doing [*inaudible*] We don't know... This is like overduty. Well, what's giving them cancer, what protein? ...If we can figure out what proteins are giving them cancer, we can figure out like before they get the cancer, they're gonna have cancer.

Before this passage, Robert had noted only needing to define bone marrow transplants. However, after reading the scaffolds, Robert struggled and then mentioned other things they needed to research: "What's giving them cancer, what protein?" He then responded to the question about what else they needed to find by typing, "why they are getting cancer" and "Is there another treatment other than the transplant when their Marrow has completely gone bad" in the information to find field.

From the symbolic interactionist perspective, we note that when the group communicated orally they had difficulties sharing their ideas effectively. Questions were raised in students' minds when they discussed what they read in the *Connection Log*, and they

sought to perform research based on those questions. Thus they appeared to construct a meaning of the *Connection Log* as a tool that helped them think of ideas.

Progression of their ideas Initially, Robert wrote that his group needed to address “what causes bone marrow to go bad?” He identified the stakeholders as “people with cancer,” and noted, “their bone marrow goes bad” and this affects them because they “could possible [sic] die.” Erin wrote that the stakeholders were “Cancer Victims” and the problem was “People are getting cancer everyday [sic] and getting [sic] the Transplant, and it does not always work.” She identified what is happening as “Their bown [sic] marrow isn’t working properly” and that this affects the stakeholders because “They might die.” Alejandra wrote that the problem was “Is there a cure for bone marrow?” She identified the stakeholders as “CancerVictims [sic],” and said, “THEY [sic] have to get aTransplant [sic] which might work” and that this affects them because “They could live or die.”

When they came to consensus, they wrote that the stakeholders were “people with cancer” whose “bone marrow goes bad and they have to get transplants” and that this affects them because they “could die.” As additional information they needed to find, they wrote “why they are getting cancer?” and “Is there another treatment other than the transplant when their Marrow has completely gone bad?”. Robert chose to address the question “What does the Bone Marrow do?”. However, when he found information and entered it into the *Connection Log*, technical difficulties prevented it from being written into the database. He thus wrote what he found in a Word document, which we were unable to collect before he logged out. Eventually they discovered that bone marrow transplant complications result from differences in DNA between donor bone marrow and the donor recipient.

In the persuasive presentation, Robert proposed using the grant to develop a method to clone a patient’s healthy bone marrow cells and using the cloned cells to replace the diseased cells. To illustrate the scope of the problems in bone marrow transplantation, they cited how many bone marrow transplants take place each year in the US and how many die from leukemia-related diseases.

Higher-Achieving Group

Daniel noted that early in the unit, “I was like, ‘I don’t know how we’re gonna get through this.’ ‘Cause like starting from scratch, it was pretty hard to get like this here and that there and then put it together.” He explained later that getting started was the hard part, and that once they started it was easier “just ‘cause we knew what we were doing, and we like know what we need, and we decided how we were gonna use it.” Megan perceived that the unit was more difficult than a previous unit because, “you kinda had to figure out what you were, what you’re doing, how you do it.” Claudia noted that the project was hard because “[in a previous unit]... it was easier for me to find information.”

How they used the Connection Log Daniel, Claudia, and Megan used the *Connection Log* to (a) get and stay organized, (b) serve as a reference for later, and (c) ensure inclusion of all required presentation parts.

To get and stay organized. As Daniel explained, “It helped us organize... ‘cause it had it like in segments.... And it helped us put [the presentation] together so it didn’t look confused and sloppy.” Just before this passage, he pasted the introduction into the found

information part of the *Connection Log*, “Okay, we need to undo some of this stuff [referring to information entered into the *Connection Log*]. [Pause] Okay, shouldn’t we shorten this up?... How’s that?” Reflecting on this episode, Daniel mentioned that the categories of the *Connection Log* helped him notice that he was repeating himself when he pasted the introduction into the text box.

Their use of the *Connection Log* to facilitate organization is interesting from a symbolic interactionist perspective in that Daniel and Megan noted that the two biggest initial challenges were organization and the ill-structured nature of the unit. Daniel noted that they could overcome the challenges by responding to the *Connection Log*’s “segments” of what they needed to find and putting the segments back together for the presentation. During past units, in which they had not used the *Connection Log*, they noted having trouble staying organized.

To serve as a reference for later. Daniel noted that the group valued being able to access what they wrote later in that they did not have to duplicate their efforts: “Once you type it in here, and then you forget, or you need to know what you need [to find], you can just go back here and say, ‘oh I need this to put into this project to make it make sense.’” Megan noted, “If we did need something we could just... we could look on there and see what we put.”

Using a symbolic interactionist lens, this is interesting because Daniel noted that it was challenging initially to envision how to fit all found information together. The group members constructed a meaning of the *Connection Log* as a tool to hold information for later reference.

To ensure inclusion of all required parts of the presentation. The Higher-Achieving Group also appeared to use the *Connection Log* to help ensure that their speech included all required parts. When reading group notes in the *Connection Log* on the last day before the persuasive presentation, Daniel noticed that the group did not include a detailed explanation of bone grafts. He thus assigned to Megan the task of writing an explanation of bone grafts.

From a symbolic interactionist perspective this makes sense because Daniel and Megan were unsure what they needed to put into the final presentation. Including all required parts of the presentation essentially meant that they included a claim, evidence, and connection of claim to evidence. The only other required parts of the presentation were an explanation of DNA and the HGP and an explanation of who their stakeholders were. Daniel’s group’s experience appeared to lead them to construct a meaning of and use the *Connection Log* as a tool that could help them address the central problem of the unit by creating an evidence-based argument.

Progression of their ideas When asked to define the problem at the beginning of the unit, Megan wrote that her group’s stakeholder was “bone marrow transplant doctors” and that they needed to determine “if we deserve 3 million dollars” so that they could “progress [she could mean improve treatment for] patients in need.” She noted that they needed to find out “how many transplant [sic] are done per year.” Daniel also wrote that the stakeholder was “bone marrow transplant doctors.” However, he provided more detail on the task: he wrote that the group needed to determine the group “if we deserve 3 million dollars for further study of bone marrow transplant [sic]” so that they could “progress [sic] patients in need.” After discussing their answers and coming to consensus, Megan and Daniel agreed that they needed to determine “if we deserve 3 million dollars for further study of bone marrow transplant” so that they could “progress [sic] patients in need.” To

address the problem, they wrote that they needed “to now [*sic*] more about what bone marrow transplant doctors do and more about the genome project.”

Their initial problem definition was a bit unclear. One may guess that “progress patients in need” means something akin to “improve the lot of patients in need,” but that is speculation. The definition seemed to imply that they saw their goal as answering a closed-ended question—did they deserve the 3 million dollar grant? Answering such a question would not lead to the development of an evidence-based argument.

Subsequently, they decided that they needed to find “how many transplant [*sic*] are done per year,” “what we are,” “what we do,” and “about how long will it take to find this cure.” This information was more relevant to addressing the ill-structured problem of the unit than what they had originally discussed finding. In their final categorization of found information, they wrote that the stakeholder was “Bone marrow transplant doctors... we give bone marrow to people with cancer and other diseases [including] leukimia [*sic*], breast cancer.” They noted needing “to prove that that [*sic*] 3 million should be ours for different discoveries [*sic*] and cures” and that the HGP could help them “see all the people who have the disease and how bad they have it.” It appeared that between the first problem definition and their evolved problem definition, they moved from thinking that they needed to answer a closed-ended question to thinking that they needed to solve a more open-ended problem.

The Higher-Achieving Group’s argument was not stored in the database due to technical problems of unknown origins that prevented students’ progression beyond the stage where they made their initial claims. Due to the camera angle, video evidence did not indicate what they typed into the backup scaffolds (word documents that were unfortunately not saved).

During the persuasive presentation, Daniel, Claudia, and Megan pretended to be bone marrow transplant surgeons. They described bone marrow and bone marrow transplants in more depth and gave more statistics related to leukemia-related disease incidence than the Lower-Achieving Group. They claimed they would use the grant to make transplants “more efficient” and increase cures. However, they ran out of time and did not detail how they would use the grant money. This could be an indication of poor planning as all groups had the same amount of time, and the Higher-Achieving Group failed to practice more than once.

Comparison of Higher- and Lower-Achieving Groups’ persuasive presentation and posttest performances with respective control classes

Lower-Achieving Group compared to lower-achieving and higher-achieving control classes

The Lower-Achieving Group performed better on claim, evidence, and connection ratings (See Table 3) than the lower-achieving control class. In addition, the Lower-Achieving Group performed 0.66 times and 0.83 times the SDs better than the higher-achieving control group on Claim and Connection, respectively. They scored 0.93 SDs lower on the Evidence scale. All group members scored better than the control class on the posttest (See Table 4). In addition, their average score, 39, was 0.41 times the SD above the posttest mean score of students from the higher-achieving control class.

Higher-Achieving Group compared to the higher-achieving control class

The Higher-Achieving Group performed better on claim, but worse on evidence and connection ratings (See Table 3) than the higher-achieving control class. A possible reason

Table 3 Persuasive presentation rating scores by small group

Small group	Claim			Evidence			Connection		
	Score	Control mean	SD from control mean ^a	Score	Control mean	SD from control mean ^a	Score	Control mean	SD from control mean ^a
Lower-Achieving Group	6	3.75	+1.76	4	3.75	+0.13	6	2.5	+1.98
Higher-Achieving Group	6	5	+0.66	4	5	-0.93	2	4.75	-1.84

^a Control mean for Lower-Achieving Group is the mean of the lower-achieving control class; control mean for Higher-Achieving Group is the mean of the higher-achieving control class

Table 4 Posttest scores by small group and members

Group	Member	Posttest score	SD from control mean ^a
Lower-Achieving Group		39	+1.05
	Robert	40	+1.16
	Erin	42	+1.37
	Alejandra	35	+0.61
Higher-Achieving Group		35	-0.2
	Daniel	38	+0.26
	Megan	33	-0.52
	Claudia	34	-0.36

^a Control mean for Lower-Achieving Group is the mean of the lower-achieving control class; control mean for Higher-Achieving Group is the mean of the higher-achieving control class

is that the group ran out of time during the persuasive presentation. But other possible reasons include that one of their members joined the group late. Daniel scored higher and Megan and Claudia lower than the control class on the posttest (See Table 4).

Comparison of Lower- and Higher-Achieving Groups' posttest performances

The Lower-Achieving Group scored 0.62 times the SD higher than the Higher-Achieving Group on the posttest. The average pretest score for the Lower-Achieving Group was 33, which was 0.12 SD above the pretest score of the Higher-Achieving Group, who scored 32.33. While the Lower-Achieving Group, on average, performed better on the pretest, the effect size was very small, while the effect size for the posttest was medium according to Cohen (1969). In addition, a marginal outlier—a score of 2.04 SD below the class mean, a very large effect size—lowered the pretest average of the Higher-Achieving Group.

Results of observations

Informal observations indicated that students from the experimental condition used the *Connection Log*. Experimental students tended to delegate responsibility for searching for different types of information. For example, in the cloning group in one of the experimental periods, one member looked for information on how cloning can be accomplished, while another looked into ethical issues raised about cloning. Also common in groups in

the experimental condition was the practice of writing down found information in the *Connection Log*, and referring to it later. A group from an experimental period whose stakeholder position was adopted children, for example, looked in the *Connection Log* to refer to an absent student's research about issues faced by adopted children. This was possible because what the absent student had found was in the database, and not in her head or notebook.

An approach that appeared to be typical among many groups in the control condition is represented by that of a group whose stakeholder position was adopted children. One member of the group asked me (the first author) on Day 4 to help her find a web page that stated how many adopted children there are in Indiana. I found two links, but due to wireless network difficulties, she could not get the pages to load immediately. I told them to access the pages later during the period. They finally got the page to come up, but then various members of the group asked me on Days 6 and 7 to help them find how many adopted children there are in Indiana. As another example of this approach, members of the endangered species group in the other control period continually looked for the same information about what Pandas eat, or for the same types of pictures of Pandas. Much of this redundancy could have resulted from students not writing down what they found. Indeed, when I asked group members from the control condition why they were looking for the same information, they often noted that it was for two reasons. They either did not write down what they had found previously, or another member had found the information on a previous day, but was not present on the day in question. Another thing that members of many groups from the control condition did is look for the same information at the same time using the same search terms at their individual computers. They would thus find many of the same web pages, and look through them at the same time.

Discussion

Summary and discussion of important results

An interesting finding was that of a significant and substantial simple main effect on the argument evaluation ability of lower-achieving students. In addition, the effect among lower-achieving students was approximately twice the magnitude as the effect among all students collectively. This indicates that a system of networked hard scaffolds may have the potential to help lower-achieving middle school students improve their ability to critically evaluate evidential support for a claim. It is important to consider alternative explanations for the difference in performance between groups. As is common in K-12 research, individual students were not randomly assigned to condition, but rather nested in classrooms. There were no significant differences in argument evaluation ability on a pretest (a) between the two lower-achieving section and (b) between the two higher-achieving sections. This helps to ensure, but does not guarantee, that any differences on post-unit measures are due to the treatment. Developed through the three previous years facilitating the unit, teacher support for the argumentation process was uniform between the experimental and control sections, so any difference in argumentation ability was likely due to the *Connection Log*. Furthermore, improvement in argument evaluation ability is an important first step towards improvement in ability to create evidence-based arguments, as students who cannot distinguish between a sound and an unsound evidence-based argument would not likely be able to create a sound evidence-based argument (Perelman and Olbrechts-Tyteca 1958). Many teachers avoid setting higher-order thinking instructional

goals for lower-achieving students because they think that these students are incapable of thinking at a high level (Raudenbush et al. 1993; Zohar and Dori 2003). This study shows that lower-achieving students can improve substantially in at least one type of higher-order thinking with the proper support of a relatively brief duration—about 2 weeks.

Also, the case studies provided evidence that the *Connection Log* may help middle school students engage in the argument creation process more effectively. Specifically, it may have the potential to help students define the problem better, look for more relevant information, and construct a more coherent argument, which are three of the five steps in the evidence-based argument creation process (Belland et al. 2008). Again, it is important to consider alternative explanations for the results. Bias may have led students to be overly positive in their portrayal of how they used the *Connection Log*. However, to counter that possibility we always looked for confirmation of themes from multiple data sources and we always looked for counterpatterns. In this study, the *Connection Log* appeared to help students in the groups selected for case studies (a) realize that their task involved answering an open-ended question, (b) determine and find more relevant information than at unit beginning, and (c) generate an effective argument as determined by their performance on the persuasive presentation.

Implications for scaffold design guidelines

Because three scaffold design guidelines were used to inform the design of the *Connection Log*—(a) embed scaffolds within a system, (b) have students articulate their thoughts, and (c) focus on the development of conceptual, strategic, and procedural hard scaffolds—it is important to consider how features that emerged from the use of the guidelines either supported or hindered student learning. Please note that this is only one study and thus can only provide preliminary support for the guidelines. Further research is needed.

Have students articulate their thoughts

Articulation of research results and opinions appeared to play an important role in the experiences of students in the Lower-Achieving and Higher-Achieving Groups. Members of both groups appeared to benefit from being able to read and debate what each other wrote in the *Connection Log* so that they could (a) compare ideas, (b) communicate, and (c) organize.

Compare ideas Erin noted that articulation allowed her thoughts to become more coherent both to her groupmates and to herself. This in turn allowed the group to weigh the merit of Erin's ideas. This parallels findings that articulation helps firmly root ideas in students' minds (Bell 1997; Chi et al. 1989; Nussbaum 2002). Additionally, sometimes there were inconsistencies between what students said and what they wrote, and this allowed students such as Robert to ask for clarification.

Communicate Alejandra (Lower-Achieving Group) appeared to be able to stay involved with her group due to her groupmates' articulation of their thoughts. Alejandra noted having trouble understanding what her group members said, but she could understand what they wrote. Language learners often feel comfortable reading and writing before they feel comfortable speaking and listening in the new language (Hadley 1993). As PBL involves students collaborating to create a viable solution to a problem, success in PBL depends on

effective communication (Lindblom-Ylänne et al. 2003). ENL students make up a large and rapidly growing proportion (19%) of school-aged children in the US (National Center for Educational Statistics 2006). We did not videotape groups in the control condition; thus we could not analyze the experimental groups at a very detailed level. However, observations indicated that in one of the control classes a translator aided five ENL students with low oral English proficiency by translating everything that the teacher or their groupmates said. This was problematic because the ENL students often did not know what they were supposed to do, the translator asked the teacher or us, she or we responded, and then he translated what we said for the ENL students. We did not observe the students talking to their groupmates.

Student absences can pose problems during group projects in K-12 schools. The *Connection Log* appeared to help assuage these problems because students needed to articulate their thoughts in writing. When students are absent, the *Connection Log* database displays what they had written. For example, when Robert left the group to “meditate,” the aural line of communication with Robert was temporarily severed. Erin and Alejandra noted being able to keep going on the project by reading what Robert had written in the *Connection Log*. In our observations of other groups who used the *Connection Log* we noticed some whose group members were absent doing some of the same things. In contrast, informal observation indicated that members of at least three groups in the control condition tried to imagine what an absent group member had been assigned to find, what he/she had already found, and what they needed to find in his/her absence.

Organize Also interesting is that both groups noted using the *Connection Log* to get and stay organized. Central to the ability of students to organize their thoughts using the *Connection Log* is articulation, because this allowed students to access the ideas and organize them. Though we never observed members of either group during previous PBL units, we can say from experience that typical middle school students left to their own devices, regardless of whether they are higher- or lower-achieving, write down what they find on pieces of notebook paper. Rarely in those situations is all information found by different group members put together in an organized format. Organization is a challenge for students engaged in inquiry units, who often spend more time trying to keep organized than pursuing learning issues (Blumenfeld et al. 1991; Hmelo et al. 2000).

Focus on the development of conceptual, strategic, and procedural hard scaffolds

No *Connection Log* scaffolds were overtly metacognitive. Metacognition refers to an individual’s ability to monitor and evaluate the extent to which he/she understands something (Flavell 1979). Metacognitive scaffolds can be defined as scaffolds that explicitly tell students to question their own understanding (Hannafin et al. 1999). Our literature review indicated that students tend not to use them (Brush and Saye 2001; Oliver and Hannafin 2000), and we concluded that (a) it is unwise to create scaffolds that would not be used, and (b) teachers can provide better metacognitive support. We built into the process model of the scaffolds the opportunity for students to question their groupmates’ understanding through the process of argumentation. We noticed during observations and from the video that many students questioned their groupmates’ ideas. Often, one groupmate would write something in the *Connection Log* and another would read it. In such a case the latter would often say something like “Do cloned people really have health

problems?” Sometimes the questioned group mate would say something like “I dunno.” But sometimes, he/she would attempt to explain. Video data and informal observations indicated that students in the experimental condition often questioned their groupmates’ understanding. In other words, rather than students engaging in metacognition related to their own thoughts, their groupmates evaluated their thoughts, and, through a process of argumentation, students modified their ideas as they worked toward consensus. Kuhn (1999) suggested that through the process of having students engage in argumentation they can ultimately gain metacognitive skills. However, we did not collect data on students’ metacognitive skills. Methodological limitations also prevented an analysis of how the teacher scaffolded students on a metacognitive level.

Limitations and suggestions for future research

Limitations included technical problems, possible bias related to the first author’s presence, the scope of the study, teacher effect, and non-random assignment. The *Connection Log* was a new system and so we expected some technical problems. A limited number of technical errors in the *Connection Log* caused some data to not be entered into the database. The origins of the technical problems were not clear, but as the *Connection Log* is a web-based system, it required communication via the Internet between laptops and a database housed off site. The laptops connected to the Internet via a wireless connection, and the connection faded at least twice. On occasion the variables being sent to the database (what the students wrote) may have been misinterpreted by the database. This was problematic for two reasons. First, since student responses built off previous responses, students had to retype information that was not sent to the database. Second, the data was not available for analysis, or for determining the extent to which students used all parts of the scaffolds. Future research should use additional measures (e.g., observation checklist) to verify that experimental students use scaffolds; this in turn can help establish a cause and effect relationship between use of the scaffolds and differences in scores on dependent measures (O’Donnell 2008).

The first author was present during the entire unit in all control and experimental class periods. To counter any influence potential bias had on findings, we (a) encouraged students to be honest about their experiences during the unit and (b) confirmed all themes and looked for counterpatterns with multiple data sources.

Due to finite resources, no studies examine all factors that affect student learning. As we were interested in students’ ability to create evidence-based arguments, we focused largely on the totality of the argumentation process, and were thus unable to fully describe students’ problem definition. A richer description of students’ problem definitions during argumentation would be suggested in future research.

The teacher had developed over time soft scaffolding strategies to support her students’ creation of evidence-based arguments. Informal observations indicated that she used these strategies equally with students in both conditions. However, using the *Connection Log* among students taught by a different teacher may lead to different results.

Despite the lack of random assignment of individual students, having one higher-achieving and one lower-achieving class in experimental and control conditions allowed us to determine with strong, but not absolute, confidence whether the *Connection Log* caused a difference in either group argument quality or argument evaluation ability. But there is clearly a need for more research to determine if the results that we found in this study hold true with different students from the same or different schools. Our theoretical lens—symbolic interactionism—implies that different students will use the *Connection Log*

differently since they face diverse challenges during PBL. Furthermore, it cannot be assumed that just because the *Connection Log* produced a simple main effect of medium magnitude among lower-achieving students in this study that it always will. Lower-achieving students at other schools may struggle with argumentation for different reasons than the students in this study and may need different supports. The only way to find out is through further research.

The lack of random selection and assignment of participants may have led to threats to internal validity such as differential selection of participants (Ary et al. 2002). We controlled for preexisting differences in argument evaluation ability by administering a pretest to test for significant differences between class periods in argument evaluation ability. Also, the use of intact classrooms may have led to greater ecological external validity, or the extent to which the research setting is similar to the settings in which the treatment would be used in non-research settings (Ary et al.).

Future research should use a transfer test that requires students to generate an argument given a unique situation. It is also important to determine if students from similar and different student populations use the *Connection Log* in a similar manner during similar units. Future research should also use a wider rubric to assess students' persuasive presentations.

A major limitation of this study is the depth of reporting of the control students' processes. When compared to the data sources used to describe control students' processes, data sources describing experimental students' processes were more extensive and allowed for more rich descriptions of processes. Close examination of control students' problem solving processes would allow future researchers to compare and contrast strategies used by control and experimental students, and thereby isolate reasons for superior performance by experimental or control students. For example, students in the control condition may not articulate their ideas prior to discussing them with their groupmates, and this may lead to inferior performance, as articulation is crucial to science learning (Puntambekar and Kolodner 2005; Sandoval and Reiser 2004).

In addition, future research should examine when and how the *Connection Log* or other systems like it could be faded, as the literature base currently does not describe well the fading of hard scaffolds (Pea 2004; Puntambekar and Hübscher 2005). Our guidelines that suggested that all scaffolds be part of a system and that students articulate their thoughts caused us to design the *Connection Log* in such a way that fading it would have been difficult. Each student needed to type his/her answers to prompts for groupmates to be able to read each other's input. This renders infeasible one of the only existing models for the fading of hard scaffolds that involves students deciding when they do not need the former any more (Puntambekar and Hübscher).

Conclusion

Students will face unique challenges in the twenty-first century, and to help them prepare, schools need to incorporate authentic inquiry in school curricula (Carnegie Council on Adolescent Development 1989; Jackson and Davis 2000). Creating evidence-based arguments is central to scientific thinking and student success during PBL units (Bricker and Bell 2008; Clark and Sampson 2007; Jonassen 2003; Osborne et al. 2004). Our study provided further evidence that scaffolding can support students' development of argumentation abilities. Specifically, it produced a medium effect on lower-achieving students' argument evaluation ability, a crucial first step in argumentation and notable in a unit of only 2 weeks. But we also examined closely how individual student groups used the

scaffolds in the context of a PBL unit. More evidence from additional studies is needed, but scaffold designers may be advised to create scaffolds that require middle school students to articulate their thoughts in a networked system.

Acknowledgments The authors thank Drs. Peg Ertmer and Brian French for their many helpful comments on this research.

Appendix

See Table 5.

Table 5 Persuasive presentation rating rubric

Argument component	Score	Criteria
Claim	6	Group makes assertion that is related to the problem, the HGP/DNA, and the group's stakeholder position. The assertion is clear <i>and</i> complete
	4	Group makes assertion that is related to the problem, the HGP/DNA, and the group's stakeholder position. The assertion is either not clear <i>or</i> not described in enough detail to be complete
	2	Group makes assertion that is related to the problem, and the HGP/DNA, but not the group's stakeholder position. The assertion is <i>neither</i> clear <i>nor</i> described in enough detail to be complete
	0	Group does <i>not</i> make assertion or the assertion is not related to the HGP/DNA
Evidence	6	Group provides evidence with claim. The evidence is clear <i>and</i> described in enough detail
	4	Group provides evidence with claim. The evidence is either not clear <i>or</i> not described in enough detail
	2	Group provides evidence with claim. The evidence is <i>neither</i> clear <i>nor</i> described in enough detail
	0	Group does <i>not</i> provide evidence with claim <i>Note:</i> This would also apply if the evidence has nothing at all to do with the claim
Connection of claims to evidence	6	Group clearly shows relevance of evidence to its associated claim, <i>and</i> the pertinence of the combination of the evidence and claim to their stakeholder position
	4	Group shows relevance of evidence to its associated claim, but they do <i>not</i> present the link between the evidence and claim clearly <i>or</i> they do not establish the pertinence of the combination of the claim and the evidence to their stakeholder position
	2	Group shows relevance of evidence to its associated claim, but they <i>neither</i> present the link between the evidence and claim clearly <i>nor</i> establish the pertinence of the combination of the claim and the evidence to their stakeholder position
	0	Group does not show the relevance of evidence to its associated claim <i>Note:</i> If they don't have a claim and/or evidence, they cannot get higher than a zero here

References

- Alberts, B. (2003). *On creating a "scientific temper"*. *Science literacy for the twenty-first century* (pp. 57–61). Amherst, NY: Prometheus.
- Apple, M. W. (1975). The hidden curriculum and the nature of conflict. In W. Pinar (Ed.), *Curriculum theorizing: The reconceptualists* (pp. 95–119). Berkeley, CA: McCutchan.
- Ary, D., Jacobs, L. C., & Razavieh, A. (2002). *Introduction to research in education*. Belmont, CA: Wadsworth.
- Bell, P. (1997). Using argument representations to make thinking visible for individuals and groups. In R. Hall, N. Miyake, & N. Enyedy (Eds.), *Proceedings of CSCLE '97* (pp. 10–19). Toronto: University of Toronto Press.
- 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*, 401–422.
- Berland, L. K., & Reiser, B. J. (2008). Making sense of argumentation and explanation. *Science Education, 93*(1), 26–55.
- Blumenfeld, P. C., Soloway, E., Marx, R. W., Krajcik, J. S., Guzdial, M., & Palincsar, A. (1991). Motivating project-based learning: Sustaining the doing, supporting the learning. *Educational Psychologist, 26*(3&4), 369–398.
- Blumer, H. (1969). *Symbolic interactionism: Perspective and method*. Englewood Cliffs, NJ: Prentice Hall.
- Bransford, J. D., Brown, A. L., & Cocking, R. R. (2000). *How people learn: Brain, mind, experience, and school*. Washington: National Academies Press.
- Bricker, L. A., & Bell, P. (2008). Conceptualizations of argumentation from science studies and the Learning Sciences and their implications for the practices of science education. *Science Education, 92*(3), 473–498.
- Brush, T., & Saye, J. (2001). The use of embedded scaffolds with hypermedia-supported student-centered learning. *Journal of Educational Multimedia and Hypermedia, 10*, 333–356.
- Carnegie Council on Adolescent Development. (1989). *Turning points: Preparing American youth for the 21st century*. Washington, DC: Author.
- Chang, C., & Barufaldi, J. P. (1999). The use of a problem-solving-based instructional model in initiating change in students' achievement and alternative frameworks. *International Journal of Science Education, 21*(4), 373–388.
- Chi, M. T. H., Bassok, M., Lewis, M. W., Reimann, P., & Glaser, R. (1989). Self-explanations: How students study and use examples in learning to solve problems. *Cognitive Science, 13*, 145–182.
- 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.
- Clark, D. B., & Sampson, V. D. (2007). Personally-seeded discussions to scaffold online argumentation. *International Journal of Science Education, 29*(3), 253–277.
- Cohen, J. (1969). *Statistical power analysis for the behavioral sciences*. New York: Academic Press.
- 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 learning. *Instructional Science, 30*, 433–464.
- Driver, R., Newton, P., & Osborne, J. (1998). Establishing the norms of scientific argumentation in classrooms. *Science Education, 84*(3), 287–312.
- Duggan, S., & Gott, R. (2002). What sort of science education do we really need? *International Journal of Science Education, 24*(7), 661–679.
- Flavell, J. F. (1979). Metacognition and cognitive monitoring: A new era of cognitive-developmental inquiry. *American Psychologist, 34*(10), 906–911.
- Gallagher, S. A. (1997). Problem-based learning: Where did it come from, what does it do, and where is it going? *Journal for the Education of the Gifted, 20*, 332–362.
- Ge, X., & Land, S. M. (2004). A conceptual framework for scaffolding ill-structured problem-solving processes using question prompts and peer interactions. *Educational Technology Research and Development, 52*(2), 5–22.
- Giesbrecht, F. G., & Gumpertz, M. L. (2004). *Planning, construction, and statistical analysis of comparative experiments*. Hoboken, NJ: Wiley.
- Glaser, B. G., & Strauss, A. L. (1967). *The discovery of grounded theory: Strategies for qualitative research*. Chicago: Aldine.
- 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.

- Guzetti, B. J., Snyder, T. E., Glass, G. V., & Gamas, W. S. (1993). Meta-analysis of instructional interventions from reading education and science education to promote conceptual change in science. *Reading Research Quarterly*, 28, 116–161.
- Hadley, A. O. (1993). *Teaching language in context* (2nd ed.). Boston: Heinle and Heinle.
- Hannafin, M., Land, S., & Oliver, K. (1999). Open-ended learning environments: Foundations, methods, and models. In C. M. Reigeluth (Ed.), *Instructional design theories and models: Vol. II: A new paradigm of instructional theory* (pp. 115–140). Mahwah, NJ: Lawrence Erlbaum.
- Hawkins, J., & Pea, R. D. (1987). Tools for bridging the cultures of everyday and scientific thinking. *Journal of Research in Science Teaching*, 24(4), 291–307.
- Hmelo, C. E., Holton, D. L., & Kolodner, J. L. (2000). Designing to learn about complex systems. *The Journal of the Learning Sciences*, 9(3), 247–298.
- Hofer, B. K., & Pintrich, P. R. (1997). The development of epistemological theories: Beliefs about knowledge and knowing and their relation to learning. *Review of Educational Research*, 67(1), 88–140.
- Hunter, J. E., & Schmidt, F. L. (2004). *Methods of meta-analysis: Correcting error and bias in research findings* (2nd ed.). Thousand Oaks, CA: SAGE Publications.
- Inhelder, B., & Piaget, J. (1955). *De la logique de l'enfant à la logique de l'adolescent: Essai sur la construction des structures opératoires formelles* [From the logic of the child to the logic of the adolescent: Essay on the construction of formal operations]. Paris: Presses Universitaires de France.
- Jackson, A. W., & Davis, G. A. (2000). *Turning points 2000: Educating adolescents in the 21st century*. New York: Teacher's College Press.
- Jonassen, D. (2003). Using cognitive tools to represent problems. *Journal of Research on Technology in Education*, 35, 362–381.
- Keppel, G. (1982). *Design and analysis: A researcher's handbook* (2nd ed.). Englewood Cliffs, NJ: Prentice-Hall.
- Kinney, D. A., Rosier, K. B., & Harger, (2003). The educational institution. In L. T. Reynolds & N. J. Herman-Kinney (Eds.), *Handbook of symbolic interactionism* (pp. 575–599). Lanham, MD: AltaMira Press.
- Knudson, R. E. (1991). Effects of instructional strategies, grade, and sex on students' persuasive writing. *Journal of Experimental Education*, 59, 141–152.
- Krajcik, J., Blumenfeld, P. C., Marx, R. W., Bass, K. M., Fredricks, J., & Soloway, E. (1998). Inquiry in project-based science classrooms: Initial attempts by middle school students. *Journal of the Learning Sciences*, 7, 313–350.
- Kuhn, D. (1999). A developmental model of critical thinking. *Educational Researcher*, 28(2), 16–25.
- Kuhn, D. (2005). *Education for thinking*. Cambridge, MA: Harvard University Press.
- Kuhn, D., Black, J., Keselman, A., & Kaplan, D. (2000). The development of cognitive skills to support inquiry learning. *Cognition & Instruction*, 18(4), 495–523.
- Kuhn, D., Shaw, V., & Felton, M. (1997). Effects of dyadic interaction of argumentative reasoning. *Cognition & Instruction*, 15(3), 287–315.
- Kyza, E., & Edelson, D. C. (2005). Scaffolding middle school students' coordination of theory and practice. *Educational Research and Evaluation*, 11, 545–560.
- Lather, P. (2003). Issues of validity in openly ideological research. In Y. S. Lincoln & N. K. Denzin (Eds.), *Turning points in qualitative research: Tying knots in a handkerchief* (pp. 185–215). Walnut Creek, CA: AltaMira Press.
- Lindblom-Ylänne, S., Pihlajamäki, H., & Kotkas, T. (2003). What makes a student group successful? Student–student and student–teacher interaction in a problem-based learning environment. *Learning Environments Research*, 6, 59–76.
- Liu, M., & Bera, S. (2005). An analysis of cognitive tool use patterns in a hypermedia learning environment. *Educational Technology Research and Development*, 53(1), 5–21.
- Marttunen, M., & Laurinen, L. (2001). Learning of argumentation skills in networked and face-to-face environments. *Instructional Science*, 29, 127–153.
- Miles, M. B., & Huberman, A. M. (1984). Drawing valid meaning from qualitative data: Toward a shared craft. *Educational Researcher*, 13(5), 20–30.
- National Center for Educational Statistics. (2006). *The condition of education 2006: Indicator 7: Language Minority School-Age Children*. Retrieved April 7, 2007, from: http://nces.ed.gov/programs/coe/2006/pdf/07_2006.pdf.
- Nussbaum, E. M. (2002). Scaffolding argumentation in the social studies classroom. *The Social Studies*, 93(2), 79–83.
- O'Donnell, C. L. (2008). Defining, conceptualizing, and measuring fidelity of implementation and its relationship to outcomes in K 12 curriculum intervention research. *Review of Educational Research*, 78, 33–84.

- Oliver, K., & Hannafin, M. J. (2000). Student management of web-based hypermedia resources during open-ended problem solving. *Journal of Educational Research, 94*(2), 75–92.
- Onwuegbuzie, A. J., Jiao, Q. G., & Bostick, S. L. (2004). *Library anxiety: Theory, research, and applications*. Lanham, MD: Scarecrow Press.
- Osborne, J., Erduran, S., & Simon, S. (2004). Enhancing the quality of argumentation in school science. *Journal of Research in Science Teaching, 41*(10), 994–1020.
- Palincsar, A. S., Anderson, C., & David, Y. M. (1993). Pursuing scientific literacy in the middle grades through collaborative problem solving. *The Elementary School Journal, 93*(5), 643–658.
- Pea, R. D. (2004). The social and technological dimensions of scaffolding and related theoretical concepts for learning, education, and human activity. *The Journal of the Learning Sciences, 13*, 423–451.
- Pedersen, S., & Liu, M. (2002–2003). The transfer of problem-solving skills from a problem-based learning environment: The effect of modeling an expert's cognitive processes. *Journal of Research on Technology in Education, 35*, 303–320.
- Perelman, C., & Olbrechts-Tyteca, L. (1958). *La nouvelle rhétorique: Traité de l'argumentation* [The new rhetoric: Treatise on argumentation] (Vols. 1–2). Paris: Presses Universitaires de France.
- Puntambekar, S., & Hübscher, R. (2005). Tools for scaffolding students in a complex learning environment: What have we gained and what have we missed? *Educational Psychologist, 40*, 1–12.
- Puntambekar, S., & Kolodner, J. L. (2005). Toward implementing distributed scaffolding: Helping students learn science from design. *Journal of Research in Science Teaching, 42*(2), 185–217.
- Raudenbush, S. W., Rowan, B., & Cheong, Y. F. (1993). Higher-order instructional goals in secondary schools: Class, teacher, and school influences. *American Educational Research Journal, 30*(3), 523–553.
- Sandoval, W. A., & Morrison, K. (2003). High school students' ideas about theories and theory change after a biological inquiry unit. *Journal of Research on Science Teaching, 40*(4), 369–392.
- Sandoval, W. A., & Reiser, B. J. (2004). Explanation-driven inquiry: Integrating conceptual and epistemic scaffolds for scientific inquiry. *Science Education, 88*, 345–372.
- Sandstrom, K. L., & Fine, G. A. (2003). Triumphs, emerging voices, and the future. In L. T. Reynolds & N. J. Herman-Kinney (Eds.), *Handbook of symbolic interactionism* (pp. 1041–1057). Lanham, MD: AltaMira Press.
- 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.
- Simons, K. D., & Klein, J. D. (2007). The impact of scaffolding and student achievement levels in a problem-based learning environment. *Instructional Science, 35*, 41–72.
- Stryker, S. (2001). Traditional symbolic interactionism, role theory, and structural symbolic interactionism: The road to identity theory. In J. H. Turner (Ed.), *Handbook of sociological theory* (pp. 211–231). New York: Kluwer Academic/Plenum Publishers.
- Torp, L., & Sage, S. (1998). *Problems as possibilities: Problem-based learning for K-12 education*. Alexandria, VA: Association for Supervision and Curriculum Development.
- van Eemeren, F. H., Grootendorst, R., & Snoeck Henkemans, A. F. (2002). *Argumentation: Analysis, evaluation, presentation*. Mahwah, NJ: Lawrence Erlbaum Associates.
- White, B. Y., & Frederiksen, J. R. (1998). Inquiry, modeling, and metacognition: Making science accessible to all students. *Cognition and Instruction, 16*(1), 3–118.
- Wood, D., Bruner, J., & Ross, G. (1976). The role of tutoring in problem-solving. *Journal of Child Psychology and Psychiatry, 17*, 89–100.
- Zohar, A., & Dori, Y. J. (2003). Higher-order thinking skills and low-achieving students: Are they mutually exclusive? *The Journal of the Learning Sciences, 12*(2), 145–181.