Supporting Students' Construction of Scientific Explanations Using Scaffolded Curriculum Materials and Assessments

Katherine L. McNeill, David J. Lizotte, & Joseph Krajcik
University of Michigan

Ronald W. Marx
University of Arizona

contact info:
Center for Highly Interactive Computing in Education
610 E. University Ave., Ann Arbor, MI, 48109-1259
734-647-4226
kmcneill@umich.edu

Paper presented at the annual meeting of the American Educational Research Association, April, 2004, San Diego, CA. Electronic version available at: http://hi-ce.org/aera2004.html

The research reported here was supported in part by the National Science Foundation (REC 0101780 and 0227557). Any opinions expressed in this work are those of the authors and do not necessarily represent either those of the funding agency or the University of Michigan.

Abstract

We investigated the influence of scaffolding on students' scientific explanations over an eight-week middle school chemistry unit. Students received a focal lesson on an explanation framework and then completed investigation sheets containing explanation component scaffolds over the unit. Students received one of two treatments: Continuous, involving detailed scaffolds, or Faded, involving less supportive scaffolds over time. We analyzed their investigation sheets and pretests and posttests. During the unit, students in the Continuous treatment provided stronger explanations than those in the Faded treatment. Yet on the posttest for the items without scaffolds, the Faded group gave stronger explanations than the Continuous group for certain content areas.

Supporting Students' Construction of Scientific Explanations Using Scaffolded Curriculum Materials and Assessments

Recent science reform efforts and standards documents advocate that students develop scientific inquiry practices (American Association for the Advancement of Science, 1993; National Research Council, 1996). "Learning science involves young people entering into a different way of thinking about and explaining the natural world; becoming socialized to a greater or lesser extent into the practices of the scientific community with its particular purposes, ways of seeing, and ways of supporting its knowledge claims" (Driver, Asoko, Leach, Mortimer, & Scott, 1994, p 8). One prominent inquiry practice in both the standards documents and research literature is the construction, analysis, and communication of scientific explanations. Although researchers cite explanations as important for classroom science, they are frequently omitted from classroom practice (Kuhn, 1993; Newton, Driver & Osborne 1999) and few research studies have examined the effectiveness of instructional practices in helping students construct explanations (Reznitskaya & Anderson, 2002). Our work focuses on an eight-week standards-based chemistry curriculum designed to support seventh grade students in their construction of scientific explanations. We investigated the effects of instructional and assessment scaffolds aimed at helping students construct scientific explanations.

The Importance of Scientific Explanations

Explanation construction is essential for science classroom practice for a variety of reasons. Research into scientists' practices portrays a picture where scientists construct arguments or explanations including weighing evidence, interpreting text, and evaluating claims (Driver, Newton, & Osborne, 2000). Previous research in science education demonstrates that students who engage in explanation change or refine their image of science as well as enhance their understanding of the nature of science (Bell & Linn, 2000). Scientific explanations frame the goal of inquiry as understanding natural phenomenon, and articulating and convincing others of that understanding (Sandoval and Reiser, 1997). Lastly, constructing explanations can enhance students' understandings of the science content (Driver, Newton & Osborne, 2000). A deep understanding of science content is characterized by the ability to explain phenomena (Barron et. al. 1998).

The science standards documents also reflect the importance of incorporating explanation in students' learning of science (American Association for the Advancement of Science, 1993; National Research Council, 1996). In particular, the National Research Council stresses the importance of explanation by including them in four of their five essential features of classroom inquiry (National Research Council, 2000). The table below (adapted from the National Research Council, 2000, p. 25) shows how explanations thread through many different aspects of classroom inquiry.

Table 1 Essential Features of Classroom Inquiry **Bold** added for emphasis

Learners are engaged in scientifically oriented questions.
Learners give priority to evidence, which allows them to develop and evaluate **explanations** that address scientifically oriented questions.
Learners formulate **explanations** from evidence to address scientifically oriented questions.
Learners evaluate their **explanations** in light of alternative **explanations**, particularly those reflecting scientific understanding.
Learners communicate and justify their proposed **explanations**.

Although the concept of scientific explanations permeates the research literature, there is not one accepted definition or framework for what counts as a scientific explanation. Furthermore, some researchers use the term "explanation" while others use the term "argument." At times their definitions of these terms are very similar, while other researchers differentiate between the two. In our work, we chose to use the term "explanation" in order to be consistent with science standards (American Association for the Advancement of Science, 1993; National Research Council, 1996), which the teachers we work with need to address in their classroom practice.

Our Definition of Scientific Explanation

Like many science educators our framework for scientific explanation (Bell & Linn, 2000; Driver, et al., 2000; Jiménez-Aleixandre, Rodríguez, & Duschl, 2000; Sandoval, 2003; Zembal-Saul, et al., 2002), used an adapted version of Toulmin's (1958) model of argumentation. Our explanation framework includes three components: a claim, evidence (similar to Toulmin's data), and reasoning (a combination of Toulmin's warrants and backing). The claim is an assertion or conclusion that answers the original question. The evidence is scientific data that supports the student's claim. This data can come from an investigation that students complete or from another source, such as observations, reading material, archived data, or other sources of information. The data needs to be both appropriate and sufficient to support the claim. The reasoning is a justification that links the claim and evidence and shows why the data counts as evidence to support the claim by using the appropriate scientific principles. Consequently, we use a general explanation framework (claim, evidence, and reasoning) across different content areas in science.

Educational researchers disagree about whether an explanation framework has to be domain specific or if the same framework can be used across all domains of science. Whether or not teaching general strategic knowledge proves useful for reasoning in context, is a complicated and unresolved issue (Perkins & Salomon, 1989). Passmore and Stewart argue that scientific inquiry is domain specific, which needs to be taken into consideration when designing curriculum (2002). Domain specific knowledge determines the types of questions asked, the methods used, and what counts as evidence (Passmore & Stewart, 2002; Sandoval, 2003). For example even in two areas of ecology, a scientist studying population dynamics may use mathematical models as evidence while a scientist studying field ecology may use frequency or counts of species. Although we agree that domain specific knowledge influences the inquiry practice, we conjecture that a general and useful explanation framework can be adapted across all domains of science. Argument is a form of thinking that transcends the particular content to which it refers (Kuhn, 1993). An explanation model, such as Toulmin's, can be used to assess the structure of an explanation, but it cannot determine the correctness of the explanation (Driver, Newton & Osborne, 2000). Scientific argument or explanation is a way of thinking that is not domain specific, but the general framework does need to be adapted to the specific content and context.

Explanations in Classroom Practice

Explanations are rarely a part of classroom practice (Kuhn, 1993; Newton, Driver & Osborne 1999). Furthermore, prior research on explanation in science classrooms has demonstrated some student difficulties in constructing and evaluating explanations. For example, students have difficulty using appropriate evidence (Sandoval & Reiser, 1997) and including the backing for why they chose the evidence (Bell & Linn, 2000) in their written explanations. During classroom discourse, discussions tend to be dominated by claims with little backing to support their claims (Jiménez-Aleixandre, Rodríguez & Duschl, 2000). Consequently, we focus on how to support students' construction of scientific explanations in classrooms. Specifically, we examined the role of instructional and assessment scaffolds in supporting students' written explanations.

Scaffolding Student Learning

First, we briefly discuss the history of scaffolding and why the research suggests that fading should be beneficial for greater student learning. Wood, Bruner and Ross (1976) originally introduced the term "scaffolding" in 1976. They introduced the term in the context of adult-child interactions where the more knowledgeable adult tutors the child to complete a task they would be unable to do on their own. With the help of scaffolds, learners can complete more advanced activities and engage in more advanced thinking and problem solving (Bransford et al., 2000). We define scaffolds as the supporting structures provided by people or tools to promote learning. Scaffolds can promote different types of knowledge like metacognitive expertise, inquiry abilities, and content knowledge. Although Wood et al. did not originally connect scaffolding to Vygotsky's zone of proximal development, a number of educational researchers since then have explicitly made this connection (Hogan & Pressley, 1997; Palincsar & Brown, 1984). The zone of proximal development (ZPD) defines the area between a child's independent problem solving capabilities and the level of potential problem solving capabilities with the guidance of people or tools (Vygotsky, 1978). Stone argues that scaffolds allow students to achieve a higher level of understanding within their Zone of Proximal Development (Stone, 1993). Scaffolds are the supporting structures provided by those people or tools to promote learning. In order for a scaffold to promote student understanding, it needs to reside within a students' current ZPD. If a scaffold provides too much information, the student will not be challenged to learn more. The scaffold should provide just enough information that the learner may make progress on his/her own (Hogan & Pressley, 1997).

In their study of reciprocal teaching, Palincsar and Brown (1984) discuss Vygotsky's idea that at first the parent or expert guides much of a child's cognitive activities and over time the child takes on more and more of those responsibilities. Eventually, the child performs the activities by themselves, without the help of the scaffolds. In fact, Wood et al. (1976) described scaffolding as a flexible process contingent on what a child knows and the characteristics of the learning task. This suggests that scaffolds should be adjusted over time rather than remaining constant in order to allow students greater responsibility over their own learning. Palincsar and Brown's study (1984) supports this idea of adjusting scaffolds based on students' understanding.

In studying teacher-student interactions during reciprocal teaching, they found that initially the teacher provided modeling, feedback, and practice to students in order to meet the student's needs. Over time as the student became better able to complete a task, the teacher provided less support. By the end, the teacher's role was one more of supportive audience and the student had taken over the expert responsibilities. This shift to greater control over knowledge construction resembles the shift from child to adult status where adults retain a more regulatory role controlling the cognitive interaction in their ZPD (Scardmalia & Bereiter, 1991). This suggests that by fading the scaffolds provided by the teacher the student was able to obtain more ownership and responsibility over the task. Over time, as students gain knowledge and abilities, the scaffolds may fade providing the students with greater autonomy and ownership of their knowledge construction.

Traditionally, ZPD has been discussed in terms of one-on-one interactions. There has been little research on teacher-student scaffolding in whole class settings (Hogan & Pressley, 1997). Hogan and Pressley argue that one of the reasons there has been little research in this area is because in a large classroom a teacher cannot possibly interact with every child individually. Ideally, the teacher can react to the current situation and modify the scaffolds based on the students' needs. When a teacher addresses the whole class they are confronted with multiple zones of proximal development. There is concern that teacher-student scaffolding cannot be carried out effectively in large groups (Stone, 1998).

One possible solution to this problem is having students work in groups and then scaffolding those groups. But this can still be problematic because of the number of groups in a classroom. Another possibility is to provide students with tools, such as computers or written materials, which provide students with scaffolds. Here the interaction is between the student and the computer or written materials. Because external tools (like computers or written artifacts) cannot include the dynamics of adult-child or even peer interactions, they can be seen as limited in the use of the scaffolding metaphor (Stone, 1998). Palincsar argues that one way researchers "…have hobbled the use of scaffolding is by attributing scaffolding only to interactions that occur between individuals, and typically between individuals of significantly different expertise…it is helpful to recall that ZPDs include not only people but also artifacts, and that ZPDs are embedded in activities and contexts" (1998, p. 371). Consequently, we are interested in the role of written scaffolds in instruction and assessment.

Written Scaffolds

Although previous research suggests fading encourages greater student independence, the majority of these studies have looked at adult-child interactions where the scaffolds can be individualized for the particular student's needs. Written scaffolds obviously do not have that advantage though they have been shown to increase student learning. One example of the benefits of written scaffolds includes the Thinkertools curriculum created by White and Frederiksen (1998; 2000). They designed their curriculum to scaffold students' development of scientific inquiry processes, modeling, and metacognitive skills and knowledge. In order to develop metacognitive skills, they developed a set of reflection prompts. Such as "Being Systematic" and "Being Inventive." At the end of each phase of the inquiry cycle, students evaluate their work using the two most relevant Reflective Assessments. The students' research book provides them with a prompt that includes the title of the self-assessment, such as "Being Systematic" as well as a description "Students are careful, organized, and logical in planning and carrying out their work. When problems come up, they are thoughtful in examining their progress and deciding whether to alter their approach or strategy" (1998, p 26). These descriptive prompts provide students with guidelines for the task. To evaluate the effectiveness of the metacognition prompts, White and Fredericksen compared two versions of the curriculum, one with reflection prompts and one without reflection prompts. They found that students who received the reflective prompts resulted in greater understanding of the inquiry practices.

Davis (2003) also examined the role of scaffolds or directed prompts in supporting students' reflection. In this case, she integrated the scaffolds into the Knowledge Integration Environment (KIE) software. She investigated the role of two different types of reflection prompts, generic prompts and directed prompts. She found that generic prompts were more productive for student reflection than directed prompts. A variety of technology tools have been created to scaffold students' learning, such as Computer Supported Intentional Learning Environments or CSILE (Scardamalia & Bereiter, 1991), Artemis (Krajcik, Blumenfeld, Marx & Soloway, 2000), WorldWatcher (Edelson, Gordon & Pea, 1999), and The Galapagos Finches (Reiser et al, 2001). Our study builds off of this work as well as scaffolding research on written scaffolds, and adult-child interactions. The question remains whether you should fade written scaffolds when there is no individualization afforded such as in adult-child interactions. We address this question for the construction of scientific explanations.

Explanation Scaffolds

Providing students with prompts or contextualized scaffolding, can encourage a deep learning approach in students where they are more apt to articulate their reasoning about how and why something occurs. If students do not initially provide their reasoning, prompting can result in students articulating their thoughts (Chinn & Brown, 2000). In scaffolding students' explanation construction, we attempted to make our explanation framework clear to students in order to facilitate their understanding of what an explanation is and how to create one. Making scientific thinking strategies explicit to students can facilitate their use and understanding of these strategies (Herrenkohl, Palincsar, DeWater, & Kawasaki, 1999). More specifically, revealing the tacit framework of scientific explanation through scaffolds can facilitate students' explanation construction (Reiser et al, 2001). We hoped that by providing students with our explanation framework, we would encourage deeper thinking and promote students translation of their thinking into written text.

Previous research on using scaffolds in science to promote students written explanations has focused on *content specific* scaffolds for different explanation components (e.g. Bell & Linn, 2000; Sandoval, 2003; Zembal-Saul, et al, 2002). Content specific scaffold provide students with hints about what content knowledge to use or incorporate into their explanation. For example, in Sandoval's work he provides content scaffolds that support students in using the correct data in their natural selection explanations, such as "The factor in the environment exerting a pressure is…" (2003). These studies found that the scaffolds helped students construct explanations in that particular context and content area.

Research on explanation from other disciplines has emphasized *generic explanation* scaffolds (e.g. reading, Reznitskaya & Anderson, 2002; debate Kuhn & Udell, 2001, Kuhn & Udell, 2003). Generic explanation scaffolds help students understand a general framework for their explanation regardless of the content area. For example, Kuhn & Udell in working with middle school students on debating capital punishment provide students with general scaffolds for the different components of their arguments, such as "generating reasons", "supporting reasons with evidence", and "examining and evaluating opposing-side's reasons." They found that students' provided with scaffolds showed advancement not only in capital punishment debates, but also in assessments involving other social issues (2001; 2003). Furthermore, White and Fredrickson (1998; 2000) found that their reflective scaffolds promoted student learning of inquiry processes. An examination of these scaffolds reveal that they are in fact generic prompts because the same prompts could be used regardless of content. The scaffolds promote general metacognition, which is not content specific. In Wood, Bruner, and Ross' original discussion of scaffolds, they also discuss the importance of repetition. One of the factors determining their choice of tasks was "…to make its underlying structure repetitive so that experience at one point in task mastery could potentially be applied to later activity, and the child could benefit from after-the-fact knowledge" (91). This supports the idea of using a generic prompt, which can be repeated regardless of the content of the task.

Combining generality and context specificity in instruction can result in greater student understanding and ability to use cognitive skills (Perkins & Salomon, 1989), such as constructing scientific explanations. Since both content specific and generic explanation scaffolds provide benefits, we decided to create explanation scaffolds that included both aspects. Our scaffolds included generic components for claim, evidence and reasoning that we repeated regardless of task. For example, the evidence component of the explanation scaffold took the following format, "_______ Pieces of Evidence (Provide ______ pieces of data that support your claim that _______.)" The three blanks in this example changed depending on context. For example, the following two scaffolds were provided for two different learning tasks, "<u>Two</u> Pieces of Evidence (Provide <u>two</u> pieces of data that support your claim that <u>the nail and the wrench are the same or different substances</u>.)" and "<u>Three</u> Pieces of Evidence (Provide <u>three</u> pieces of data that support your claim that <u>new substances were or were not formed</u>.)" In each case, the number of pieces of evidence changes from two to three and the claim they are trying to support changes. But the portion about providing data to support their claim also repeats. We hoped that by using this repeated format that students would understand how this same format could be used across multiple contexts. While the type of evidence changed, students always needed to provide evidence that supported their claim.

Based on this research, we felt that the repetition of the scaffolds and the generic nature of the prompts could facilitate students' learning of a general explanation framework to apply to all content areas. Yet we still wondered whether fading the written prompts would be effective. Scaffolds should be sensitive to students' current understanding and provide just enough information that students can proceed on their own (Hogan & Pressley, 1997). As students begin to learn the explanation framework, the scaffolds should be adjusted or faded to students' current understanding. This forces students to think about what they have learned from the previous scaffolds and apply their knowledge to the current learning task. But the danger of fading a written scaffold is that since it is not individualized it may fade too quickly and reside outside of a child's ZPD. Previous research providing fifth grade students with *content-specific* written scaffolds found that fading scaffolds resulted in less student learning (Lee, 2003). However, we conjecture that with older students, repetitive, more *generic explanation* scaffolds could fade to produce greater student learning.

Introduction of Explanation

Although we wanted to investigate the role of written explanation scaffolds, we decided that the written scaffolds were not a sufficient introduction to help middle school students create scientific explanation. In order for scaffolding to be successful, a child must have some prior understanding of what is to be accomplished (Wood, Bruner, & Ross, 1976; Stone, 1998). This is important because it means in an instructional setting an inquiry process, like constructing explanations, should not be introduced through scaffolding. Instead, the teacher needs to first help students understand the inquiry process before they can effectively use the scaffolds embedded in the curriculum. For example, Chen and Klahr (1999) found that providing students with the rationale behind controlling variables in science experiments, as well as examples of unconfounded comparisons before completing investigations, resulted in greater learning relative to students who did not receive the explicit instruction.

In our unit, we accomplish this by devoting an entire lesson to scientific explanation where the teacher introduces the explanation framework and models how to construct explanations. Although the lesson focuses on scientific explanation, it also includes content. Again, we combine both generic explanation supports and content-specific supports. The teacher introduces scientific explanation in the context of whether fat and soap are the same or different substance. We hope that by introducing explanations to students in this manner, when they later write their own explanations they are able to utilize the support in the explanation scaffolds.

The Relationship Between Explanation and Science Content

In our analysis of students' explanations, we examine both their ability to construct explanations and their understanding of the science concepts. If students perform poorly on one explanation, we cannot tell if their performance was because of their lack of content knowledge or their lack of understanding of the particular learning task. In order to construct an accurate scientific explanation, students need to understand both the content and how to construct a scientific explanation. Metz argues that, "…the adequacy of individuals' reasoning is strongly impacted by the adequacy of their knowledge of the domain within which the reasoning is tested. Thus, inside the research laboratory and beyond, cognitive performance is always a complex interaction of scientific reasoning capacities and domain-specific knowledge" (2000, p. 373). If students have difficulty with any of those components, they will be unable to write an accurate scientific explanation. For example, if a student does not understand the content even though they understand how to write an explanation, they will be unable to construct an accurate explanation. Consequently, we need to look beyond one explanation to hypothesize why students may be having difficulty with explanations.

Instructional Context

Using a learning-goals-driven design model (Reiser, Krajcik, Moje, & Marx, 2003), we developed a middle school chemistry unit (McNeill et al, 2003). Learning-goals-driven design uses key learning goals identified from the national standards (American Association for the Advancement of Science, 1993; National Research Council, 1996) to guide all phases of curriculum and assessment design. We used this design model to develop an eight-week project based unit addressing the driving question (Krajcik, Czerniak, & Berger, 1999), "How can I make new stuff from old stuff?" Students investigated how you can make soap from lard. During the instructional sequence, students completed other investigations, each time cycling back to soap and fat. Each cycle helped them delve deeper into the science content to initially understand substances, then properties, chemical reactions, conservation of mass, and the particulate nature of matter.

Besides the science concepts, our other key learning goals focused on inquiry abilities such as the construction of scientific explanations. As mentioned earlier, in order to introduce students to scientific explanations, we developed a focal lesson. First students wrote explanations using their own data and their prior understanding of scientific explanations. Then the teacher led a discussion about scientific explanation in order to make the framework (claim, evidence, and reasoning) explicit to students. The teacher also modeled the construction of scientific explanations through the use of hypothetical examples of weak and strong explanations. Using this framework and models, students revised their original explanations. After the focal lesson, the students wrote six more explanations over the unit.

Students recorded the results of their investigations and scientific explanations on student investigation sheets. These investigation sheets contained the explanation component scaffolds we described earlier, which combined both content specific and generic explanation scaffolds. We created two scaffold treatments, Continuous and Faded. The Continuous group received the same type of scaffold on all six investigation sheets. This scaffold provided detailed information about each explanation component. For example, for evidence the sheet said, "Three Pieces of Evidence (Provide three pieces of data that support your claim that new substances were or were not formed.)" followed by three prompts on the sheet labeled Evidence #1, Evidence #2, and Evidence #3 for the students to record their response. The Faded group received investigation sheets, which had three types of scaffolds that provided less detail over the six sheets. The first type of scaffold was the same as the Continuous group. An example of the intermediate scaffold for evidence was "Evidence (Provide data that support your claim.)." The last type of scaffold simply stated, "Remember to include claim, evidence, and reasoning," with no specific prompts about the different components.

Method

Participants

Participants included 6 teachers and 331 seventh grade students from schools in the Mid-west. Three of the teachers and 260 of the students in 9 classes were from public middle schools in a large urban area. The majority of these students were African American and from lower to lower-middle income families. The other three teachers and 71 students in 5 classes were from an independent middle school in a large college town. The majority of these students were Caucasian and from middle to upper-middle income families.

Assessment Data

Two types of assessment data were collected: student investigation sheets and pretest and posttest data. For the student investigation sheets, all three components of explanations (claim, evidence, and reasoning) were scored separately. All questions were scored by one rater. We then randomly sampled 20% of the student sheets and a second independent rater scored them. The average inter-rater reliability was above 85% for each component (claim, evidence, and reasoning) for each of the seven explanations.

All students completed the same pretest and posttest, which consisted of 30 multiple-choice and eight open-ended items. Only students who completed all parts of the test were included in the analysis. Due to high absenteeism in the urban schools and the necessity of students being in class for all four days of testing, only 220 students took all parts of the pre- and posttest assessments. We randomly selected 20% of the pretests for students that we did not have posttests. In our missing data analysis, we only examined the multiple-choice portion of the test since the majority of students missing complete exams were excluded because they were missing the open-ended items. These students' multiple-choice scores (M = 13.2, SD = 3.8) did not significantly vary from those of the other students who had complete test data (M = 12.5, SD = 4.1) in their respective classes, t (172) = 0.901, *ns*.

For this study, we focused on the four open-ended items, which asked the students to write a "scientific explanation" (see Appendix). Each student received all four items, two items had the detailed Continuous-type scaffolds and the other two items had no scaffolds. We created two versions of the test to counterbalance which items had scaffolds and no scaffolds across students. For all four questions, we scored the different components of explanation (claim, evidence, and reasoning) separately. All items were scored by one rater. Twenty percent of the tests were randomly chosen and scored by a second independent rater. The average inter-rater reliability was above 85% for each component of the four test questions.

Study Design

We randomly assigned classes of students to the Continuous and Faded groups so that teachers with multiple classes taught both groups. For example, if a teacher taught two classes of seventh grade science, we assigned one class the Continuous treatment and the other class the Faded treatment. We charted students' explanations through successive stages of the unit. For students in the Continuous scaffold treatment, scaffolds on the investigation sheets were identical in Stages I, II, and III, whereas students in the Faded scaffold treatment received progressively less detailed scaffolds through Stages I, II, and III. Stages I, II, and III each involved two investigation sheets for explanation.

Of the 220 students who completed both the pre- and posttest, 129 completed the focus lesson and at least one of the investigation sheets for each of Stages I, II, and III. We charted explanations of those 129 students through the unit. However, we used the larger sample of students for all quantitative analyses of pre and posttest data.

Results and Discussion

Our analyses address three questions: 1) How do the different scaffold treatments (Continuous or Faded) within the unit influence students' explanations on the student investigation sheets? 2) Do the scaffold treatments during the unit have different effects on students' explanations on posttest items with and without scaffolds? and 3) Does the influence of scaffolds vary depending on the content and component of the explanations?

Scaffold Treatments and Explanations Within the Unit

Students in both the Faded and Continuous groups had significant pre-posttest gains during the unit on all three components of explanation (Table 2). We are interested in whether there was a treatment effect for students' explanations throughout the unit.

Figures 1, 2, and 3 chart the mean scores for claims, evidence, and reasoning, respectively, through stages of the unit for the students in the Continuous (n = 52) and Faded (n = 77) treatments who completed the requisite investigation sheets. We performed a series of t-tests to determine whether students' scores differed at each stage of the unit according to the scaffold treatment they received. For this analysis at the Pre- and Posttest Stages, we collapsed students' scores across test items with and without scaffolds. Significant differences between the treatment groups manifested only for students' claims as shown in Figure 1. For the Pretest Stage, students in the Faded group (M = 0.70, SE = 0.04) had higher claim scores than students in the Continuous group (M = 0.53, SE = 0.05), t (127) = 2.60, $p < .05$. For Stage III, students in the Faded group ($M = 1.01$, $SE = 0.04$) again had higher claim scores than those in the Continuous group ($M = 0.85$, $SE = 0.07$), $t\ (87) = 2.08$, $p < .05$[1]. There were no other significant differences between the treatment groups.

Although the Continuous and Faded groups did not have significantly different evidence and reasoning scores at any individual stage of the unit, Figures 1-3 reveal important trends in students' scores across the different stages. First, students' scores for claim, evidence, and reasoning increased considerably from the pretest to the focal lesson on scientific explanations. For the focal lesson, students did not receive scaffolds on their investigation sheets; rather, the teacher provided students with support. Consequently, this increase is due to the teacher support and not the written scaffolds, which suggests that the role of the teacher is considerably important. We conjecture that the differences between the treatments are not significantly different during the instructional stages for evidence and reasoning because teachers discussed students' explanations in class. Since teachers taught both Faded and Continuous classes, the influence of the teacher practices may have caused the treatment groups explanations to be similar during instruction. The Faded and Continuous groups received the same scaffolds in Stage I, then in Stage II and Stage III the Faded group received less detailed scaffolds. Another trend in both Figure 2 and Figure 3 is that when scaffolds diverge for the two treatments, students in the Continuous group had higher evidence and reasoning scores on the investigation sheets than those in the Faded group. While this difference is not significant, it is interesting that the pattern switches on the posttest where the Faded group had higher evidence and reasoning scores than the Continuous group. We further unpack this trend on the posttest below.

Influence of Scaffold Treatments During the Unit on Posttest Explanations

We examined whether the Faded and Continuous scaffolds treatments influenced students' explanations on the test items using the entire sample of 220 students ($n = 97$ for continuous treatment; $n = 123$ for faded treatment). We tested whether a scaffold treatment effect was present for test items with scaffolds, without scaffolds, or both types, by performing separate ANOVAs on students' posttest claim, evidence, reasoning, and composite scores for items with and without scaffolds. For each ANOVA, Scaffold Treatment (continuous, faded) was the fixed factor and the appropriate pretest score was the covariate. The effect of Scaffold Treatment was marginally significant in one analysis: reasoning scores on posttest items without scaffolds were higher for students in the Faded treatment than the Continuous treatment $F\ (1, 217) = 3.28$, $p = .07$. Figure 4 shows the mean reasoning scores for items with and without scaffolds. This suggests that fading written scaffolds during instruction might result in greater student gains for the reasoning component of explanation for items without scaffolds.

To further tease apart this effect on reasoning, we examined the substance/property and chemical reaction explanations separately to evaluate the role of science content. We found that this effect of scaffold treatments on reasoning scores for test items without scaffolds applied to explanations about substance/property phenomena but not chemical reaction phenomena. We performed separate ANOVAs on students' posttest reasoning scores for substance/property and chemical reaction items, with scaffolds and without scaffolds; Scaffold Treatment (continuous, faded) was the fixed factor and the appropriate pretest score was a covariate for each analysis. For substance/property items, the effect of Scaffold Treatment on students' reasoning scores was significant for items without scaffolds, $F\ (1, 217) = 3.95$, $p < .05$. Figure 5 shows that the mean reasoning score for substance/property items without scaffolds was higher for students in the

[1] Note that results of this test were corrected for unequal variance between the groups.

Faded treatment than the Continuous treatment. For comparison, Figure 5 also includes the mean reasoning scores for substance/property items with scaffolds. Again, the Faded treatment is higher, but this difference is not significant. Consequently, fading written scaffolds during instruction had a positive effect for items without scaffolds on the posttest. For items that included scaffolds, the type of scaffolding treatment was not significant.

For chemical reaction items, there were no significant effects of Scaffold Treatment on students' reasoning scores. Specifically, reasoning scores for students in the Faded and Continuous treatments did not differ either for chemical reaction items without scaffolds (M = 0.39, SE = 0.04 for Faded; M = 0.33, SE = 0.04 for Continuous) or for those with scaffolds (M = 0.40, SE = 0.04 for Faded; M = 0.35, SE = 0.04 for Continuous).

To summarize, students who received the Faded treatment had significantly higher reasoning scores on posttest items without scaffolds than those who received the Continuous treatment, but only for substance/property items. These results suggest that the scaffold treatment had different effects on the different components of explanation (claim, evidence, and reasoning) and the different content areas (substance/property and chemical reactions). To explore possible causes of these differential effects, we examined other sources of data from the enactment.

Relationship Between Science Content and Scientific Explanations

To further investigate this relationship between the science content and students' ability to construct explanations, we looked at students' performance on the multiple-choice items for both substance/property and chemical reaction. We determined the correlations between students' posttest multiple-choice and explanation scores for each content area. Not surprisingly, there is a relationship between these two scores. Students' scores on the substance/property multiple-choice items were significantly correlated with their substance/property explanations, rs (220) = 0.37 for claim, 0.35 for evidence, and 0.52 for reasoning, ps < .001. Students' scores on the chemical reaction multiple-choice items were significantly correlated with their chemical reaction explanations, rs (220) = 0.45 for claim, 0.40 for evidence, and 0.41 for reasoning, ps < .001. Students who had higher multiple-choice scores in a content area also had higher explanation scores in that area. We examined students' posttest scores comparing substance/property items to the chemical reaction items. We found that students scored higher on the set of substance/property items (M = 3.90, SE = 0.07) than the equally-weighted set of chemical reaction items (M = 3.65, SE = 0.08), t (219) = 3.96, p < .001. Students have a stronger understanding of the substance/property content.

One possible reason students who received faded scaffolds had higher substance/property reasoning but not different chemical reaction reasoning than those who received continuous scaffolds is because the students had a stronger understanding of the substance/property content. Metz (2000) argues that children's reasoning is highly influenced by their knowledge of the domain. If students do not understand the content, they are not able to construct valid scientific explanations. Consequently, we might not have seen a unit scaffold treatment effect for students' chemical reaction explanations because their understanding of the content was not strong enough to demonstrate their ability to construct scientific explanations.

Student Difficulty with the Reasoning Component of Explanations

Another trend in our results was that the scaffolds appeared to have the strongest influence on the reasoning component of the explanations. We hypothesize that this greater influence is because students had the most difficulty with the reasoning component of explanation. Figures 1-3 display students claim, evidence, and reasoning scores for each of the treatments across the unit. These figures show that claim and evidence are consistently higher than reasoning, regardless of treatment, at the beginning, middle, and end of the unit. For example, on the posttest the average score across all the students for claim was 3.20, for evidence 2.73, and for reasoning 1.76 (Maximum score for all three components is 5.0).

We also found that during students' discussions, that they had a difficult time with the reasoning component of scientific explanations. During the focus lesson, students are introduced to the concept of scientific explanation. After the class discusses explanation and the three components (claim, evidence, and reasoning), students work on writing their own scientific explanations. During the writing of this initial explanation, the investigation sheets do not include any scaffolds. Students must rely on the discussion and any notes the teacher has provided them to support their construction. We videotaped a group of students from one of the urban public middle schools as they worked on their explanations and captured the following conversation.

S1: What does the reasoning mean?
S2: To explain your statement.
S3: Telling why.
S1: Why is the evidence.
S3: Um. Ok.
S2: Ok.
S1: So how does the reasoning help you. I mean also why?
[Pause. Students look at each other.]
S2: I have no idea.
S3: Me either.
S2: Well, its all the same thing
S1: What is the reasoning?
S2: (points to the wall where it says Explanation: Claim, Evidence, and Reasoning) Uh. The explain is the claim –
S1: What is the reasoning?
S2: The reasoning is the evidence.
S1: What is the reasoning? What is the reasoning?
S4: You ask them.
S3: Why are you asking that?
S1: She said we gotta write claim, evidence, and reasoning. Reasoning is the same thing as evidence.
S3: No. Reasoning is why you think it and then you tell the evidence.
S1: Hmm?
S3: Reasoning is why you think and then you tell the evidence like when we put -
S2: - like the melting point -
S3: - the fat in the oil -
S2: - the melting point and the solubility
S3: See?
S1: Thank you.

This conversation suggests that these students did not have a clear understanding of reasoning during the focal lesson. They equated evidence and reasoning as the same. The students did not know that the reasoning should include the scientific principle, while their evidence should include the data. We hypothesize that students' difficulty with reasoning made the scaffolds particularly important for this component of scientific explanation throughout the unit. This may be why we see an effect for the Faded treatment for students' reasoning scores, but not their evidence or claim scores. By fading the scaffolds, this may have forced students to revisit this question of "What is the reasoning?" instead of relying on the support of the written scaffolds. This may have better equipped them to provide the reasoning on the posttest when no scaffolds were provided.

Conclusion

Originally, Wood et al. (1976) described scaffolding as flexible process that relies both on what a child knows and the learning task. Scaffolding should provide just enough information that learners can make progress independently (Hogan & Pressley, 1997), which suggests that scaffolds should be adjusted or faded as students learn. Although the fading of scaffolds has been found to be beneficial for students during teacher-student interactions (Palincsar & Brown, 1984), there has been little research examining whether the fading of written scaffolds promotes student learning. Our findings suggest that fading written scaffolds that include both content specific and generic explanation supports may better equip students to write explanations when they are not provided with support. Yet the effect of scaffolds on student learning may depend on both the difficulty of the content and the particular inquiry practice.

We found a relationship between students' content knowledge and their ability to construct explanations. For the content areas where students' understanding was weaker, they also constructed weaker scientific explanations. Students' performance depends on both their reasoning capabilities and their understanding of the science content (Metz, 2000). Consequently, when students construct poor explanations it is difficult to tease out if students' difficulty stems from their lack of understanding of the content or their lack of understanding of scientific explanations. The relationship between content understanding and explanation construction suggests that one reason the scaffolds did not appear to influence students' explanations on the chemical reaction questions on the posttest may be because the students did not understand the chemical reaction content well enough to demonstrate their understanding of scientific explanations.

Fading the scaffolds also had a significant effect on students' reasoning scores, but not their claims or evidence. We believe that the scaffolds had the strongest effect on students' reasoning scores because this was the most difficult component for students. Previous research has found that students have difficulty providing the backing for their claims and evidence both in their written explanations (Bell & Linn, 2000) and during classroom discussion (Jiménez-Aleixandre, et al, 2000). We found that students reasoning scores were lower than both their claim, and evidence throughout the unit. Since students had difficulty with reasoning, perhaps by fading the scaffolds we forced students to think about and apply what they had learned from the previous scaffolds to the current learning task. When the scaffolds were removed on the posttest, this may be why students in the Faded treatment scored higher.

Educators need to pay careful attention to the relationship between the type, detail and duration of scaffolds. This relationship may depend on the difficulty of the science content, students' prior experiences, as well as other characteristics of the learning environment. For example, Lee's (2003) study found that younger students did not benefit from the fading of content specific scaffolds. We found that fading scaffolds that combined both content specific and generic components with 7th grade students did result in their greater learning of scientific explanations. This suggests that there is not a general rule about fading written scaffolds that applies to all learners and instructional contexts.

Besides the fading of scaffolds, the format and language of the scaffolds may be important for student learning. For example, the lower reasoning scores on the posttest (compared to claim and evidence) may be the result of the materials not providing enough support for this component. Stone (1993) discusses how successful scaffolding involves the construction of shared definitions in a particular situation or context. Although Stone discusses the importance of a shared discourse in adult-child interactions, we believe that this shared definition is also important in written scaffolds. One of the reasons students had difficulty constructing the reasoning portion of their explanations may be because they did not share the same understanding of reasoning as we intended in the curriculum. Using a different scaffolding format or language in the scaffold, may have been more effective. The design of scaffolded instructional materials requires the careful consideration of multiple factors.

References

American Association for the Advancement of Science. (1993). *Benchmarks for science literacy.* New York: Oxford University Press.

Barron, B., Schwartz, D., Vye, N., Moore, A., Petrosino, A., Zech, L., Bransford, J., & The Cognition and Technology Group at Vanderbilt. (1998). Doing with understanding: lessons from research on problem- and project-based learning. *The Journal of the Learning Sciences. 7 (3&4),* 271-311.

Bell, P., & Linn, M. (2000). Scientific arguments as learning artifacts: Designing for learning from the web with KIE. *International Journal of Science Education. 22 (8),* 797-817.

Bransford, J., Brown, A., & Cocking, R. (Eds.). (2000). *How people learn: Brain, mind, experience and school.* Washington D.C.: National Academy Press.

Chen, Z., & Klahr, D. (1999). All other things being equal: Acquisition and transfer of the control of variables strategy. *Child Development*. *70*(5). 1098-1120.

Chin, C. & Brown, D. E. (2000). Learning in science: A comparison of deep and surface approaches. *Journal of Research in Science Teaching, 37*(2), 109-138.

Davis, E. (2003). Prompting middle school science students for productive reflection: generic and directed prompts. *The Journal of the Learning Sciences. 12*(1), 91-142.

Driver, R., Asoko, H., Leach, J., Mortimer, E., & Scott, P. (1994). Constructing scientific knowledge in the classroom. *Educational Researcher, 23* (7), 5-12.

Driver, R., Newton, P. & Osborne, J. (2000). Establishing the norms of scientific argumentation in classrooms. *Science Education. 84 (3),* 287-312.

Edelson, DC, Gordin, D.N., & Pea, R.D. (1999). Addressing the challenges of inquiry-based learning through technology and curriculum design*.* *Journal of the Learning Sciences,* 8(3/4) 391-450.

Herrenkohl, L. R., Palinscar, A. S., DeWater, L. S., & Kawasaki, K. (1999). Developing scientific communities in classrooms: A Sociocognitive Approach. *The Journal of the Learning Sciences. 8*(3&4), 451-493.

Hogan, K., & Pressley, M. (1997). *Scaffolding student learning: Instructional approaches and issues*. Cambridge, MA: Brookline Books.

Jiménez-Aleixandre, M. P., Rodríguez, A. B., & Duschl, R. A. (2000). “Doing the lesson” or “doing science”: argument in high school genetics. *Science Education, 84*, 757-792.

Kuhn, D. (1993) Science as argument: Implications for teaching and learning scientific thinking. *Science Education, 77,* 319-338.

Kuhn, D., & Udell, W. (2001). The path to wisdom. *Educational Psychologist. 36 (4),* 261-264.

Kuhn, D & Udell, W. (2003). The development of argument skills. *Child Development. 74 (5),* 1245-1260.

Krajcik, J., Blumenfeld, P., Marx, R., & Soloway, E. (2000). Instructional, curricular, and technological supports for inquiry in science classrooms. In J. Minstrell & E. v. Zee (Eds.), *Inquiring into Inquiry Learning and Teaching in Science* (pp. 283-315). Washington D.C.: AAAS.

Krajcik, J., Czerniak, C. M., & Berger, C. (1999). *Teaching children science: A project-based approach*. Boston, MA: McGraw-Hill.

Lee. H. S. (2003). *Scaffolding elementary students' authentic inquiry through a written science curriculum.* Unpublished doctoral dissertation, University of Michigan, Michigan.

Metz, K. E. (2000). Young children's inquiry in biology: Building the knowledge bases to empower independent inquiry. In J. Minstrell & E. H. van Zee (eds.), *Inquiry into inquiry learning and teaching in science* (pp. 371-404). Washington, DC: American Association for the Advancement of Science.

McNeill, K. L., Lizotte, D. J, Harris, C. J., Scott, L. A., Krajcik, J., & Marx, R. W. (2003, March). *Using backward design to create standards-based middle-school inquiry-oriented chemistry curriculum and assessment materials*. Paper presented at the annual meeting of the National Association for Research in Science Teaching, Philadelphia, PA.

National Research Council. (1996). *National science education standards.* Washington, DC: National Academy Press.

National Research Council. (2000). *Inquiry and the national science education standards: A guide for teaching and learning.* Washington, D.C.: National Academy Press.

Newton, P., Driver, R., & Osborne, J. (1999). The place of argumentation in the pedagogy of school science. *International Journal of Science Education*, 21 (5), 553-576.

Palincsar, A. S. (1998). Keeping the metaphor of scaffolding fresh – A response to C. Addison Stone's "The metaphor of scaffolding: Its utility for the field of learning disabilities." *Journal of Learning Disabilities, 31*(4), 370-373.

Palincsar, A. S., & Brown, A. L. (1984). Reciprocal teaching of comprehension-fostering and comprehension-monitoring activities. *Cognition and Instruction*, 1, 117-175.

Passmore, C. & Stewart, J. (2002). A modeling approach to teaching evolutionary biology in high schools. *Journal of Research in Science Teaching, 39*(3), 185-204.

Perkins, D. N. & Salomon, G. (1989). Are cognitive skills context-bound? *Educational Researcher*, *18*(1), 16-25.

Reiser, B. J., Krajcik, J., Moje, E. B., & Marx, R. W. (2003, March). *Design strategies for developing science instructional materials*. Paper presented at the Annual Meeting of the National Association for Research in Science Teaching, Philadelphia, PA.

Reiser, B., Tabak, I., Sandoval, W., Smith, B., Steinmuller, F., & Leone, A. (2001). BGuILE: Strategic and conceptual scaffolds for scientific inquiry in biology classrooms. In S.M. Carver & D. Klahr (Eds.), *Cognition and instruction: Twenty-five years of progress* (pp. 263-305). Mahwah, NJ: Erlbaum.

Reznitskaya, A., & Anderson, R. C. (2002). The argument schema and learning to reason. In C. C. Block, & M. Pressley (Eds.), *Comprehension instruction: Research-based best practices* (pp. 319-334). New York: The Guilford Press.

Sandoval, W. A. & Reiser, B. (1997, March). *Evolving explanations in high school biology*. Paper presented at the Annual Meeting of the American Educational Research Association. Chicago, IL.

Sandoval, W. (2003). Conceptual and epistemic aspects of students' scientific explanations. *The Journal of the Learning Sciences, 12*(1), 5-51.

Scardamalia, M. and Bereiter, C. (1991). Higher levels of agency for children in knowledge building: A challenge for the design of new knowledge media. *The Journal of the Learning Sciences (1)* 1.37-68.

Stone, C. A. (1993). What is missing in the metaphor of scaffolding? In E. A. Forman & N. Minick & C. A. Stone (Eds.), *Contexts for learning: Sociocultural dynamics in children's development* (pp. 169-183). New York: Oxford University Press.

Stone, C. A. (1998). The metaphor of scaffolding: its utility for the field of learning disabilities. *Journal of Learning Disabilities, 31*(4), 344-364.

Toulmin, S. (1958). *The uses of argument.* Cambridge, UK: Cambridge University Press.
Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. London, England: Cambridge University Press.
White, B., & Frederiksen, J. (1998). Inquiry, modeling, and metacognition: Making science accessible to all students. *Cognition and Instruction, 16*(1), 3-118.
White, B., & Frederiksen, J. (2000). Metacognitive facilitation: An approach to making scientific inquiry accessible to all. In J. Minstrell & E. v. Zee (Eds.), *Inquiring into Inquiry Learning and Teaching in Science* (pp. 283-315). Washington D.C.: AAAS.
Wood, D., Bruner, J. S. & Ross, G. (1976). The role of tutoring in problem solving. *Journal of Child Psychology and Psychiatry. 17,* 89-100.
Zembal-Saul, C., Munford, D., Crawford, B., Friedrichsen, P. & Land, S. (2002). Scaffolding preservice science teachers' evidence-based arguments during an investigation of natural selection. *Research in Science Education*, *32 (4),* 437-465.

Appendix: Scientific Explanation Test Questions

Question #1: Examine the following data table:

	Mass	**Soluble in Water**	**Melting Point**	**Color**
Solid 1	65 g	Yes	136 °C	yellow
Solid 2	38 g	Yes	175 °C	white
Solid 3	21 g	No	89 °C	white
Solid 4	65 g	Yes	175 °C	white

Write a **scientific explanation** that answers the question: Are any of the solids the same substance?

Question #2: Maya has two liquids, hexane and ethanol. She determines a number of measurements for the two liquids and then mixes them together. After mixing the liquids, they form two separate layers, layer A and layer B. Maya uses an eyedropper to take a sample from each layer, and she determines a number of measurements for each.

	Volume	**Mass**	**Density**	**Solubility in Water**	**Melting Point**
hexane	25 cm^3	16.5 g	0.66 g/cm^3	No	-95 °C
ethanol	40 cm^3	31.6 g	0.79 g/cm^3	Yes	-114 °C
layer A	8 cm^3	6.3 g	0.79 g/cm^3	Yes	-114 °C
layer B	8 cm^3	5.3 g	0.66 g/cm^3	No	-95 °C

Write a **scientific explanation** that answers the question: Did a chemical reaction occur when Maya mixed hexane and ethanol?

Appendix: Scientific Explanation Test Questions

Question #3: Examine the following data table:

	Volume	**Density**	**Melting Point**	**Color**
Liquid 1	58 ml	0.93 g/cm^3	-98 °C	no color
Liquid 2	13 ml	0.79 g/cm^3	26 °C	no color
Liquid 3	38 ml	13.6 g/cm^3	-39 °C	silver
Liquid 4	21 ml	0.93 g/cm^3	-98 °C	no color

Write a **scientific explanation** that answers the question: Are any of the liquids the same substance?

Question #4: Carlos has two liquids, butanic acid and butanol. He determines a number of measurements for the two liquids and then mixes them together. After heating and stirring the liquids, they form two separate layers, layer A and layer B. Carlos uses an eyedropper to take a sample from each layer, and he determines a number of measurements for each.

	Volume	**Mass**	**Density**	**Solubility in water**	**Melting Point**
Butanic acid	10.18 cm^3	9.78 g	0.96 g/cm^3	Yes	-7.9 °C
Butanol	10.15 cm^3	8.22 g	0.81 g/cm^3	Yes	-89.5 °C
Layer A	2 cm^3	1.74 g	0.87 g/cm^3	No	-91.5 °C
Layer B	2 cm^3	2.0 g	1.0 g/cm^3	Yes	0 °C

Write a **scientific explanation** that answers the question: Did a chemical reaction occur when Carlos mixed butanic acid and butanol?

Table 2: Faded and continuous performance for claim, evidence, and reasoning

Treatment	Maximum Score	Pretest *M* (*SD*)	Posttest *M* (*SD*)	*t*-Value[a]	Effect Size[b]
Faded (n = 123)					
Claim	5.0	2.37 (1.46)	3.23 (1.61)	5.77***	0.59
Evidence	5.0	1.47 (1.21)	2.47 (1.59)	7.47***	0.83
Reasoning	5.0	0.39 (0.63)	1.92 (1.62)	11.12***	2.43
Continuous (n = 97)					
Claim	5.0	1.88 (1.35)	3.17 (1.46)	9.36***	0.96
Evidence	5.0	1.10 (1.06)	2.25 (1.49)	7.97***	1.08
Reasoning	5.0	0.42 (0.66)	1.55 (1.45)	7.97***	1.71

[a]One-tailed paired *t*-test:
[b]Effect Size: Calculated by dividing the difference between posttest and pretest mean scores by the pretest standard deviation.
*** $p < .001$

Figure 1. Claim Scores Over the Unit

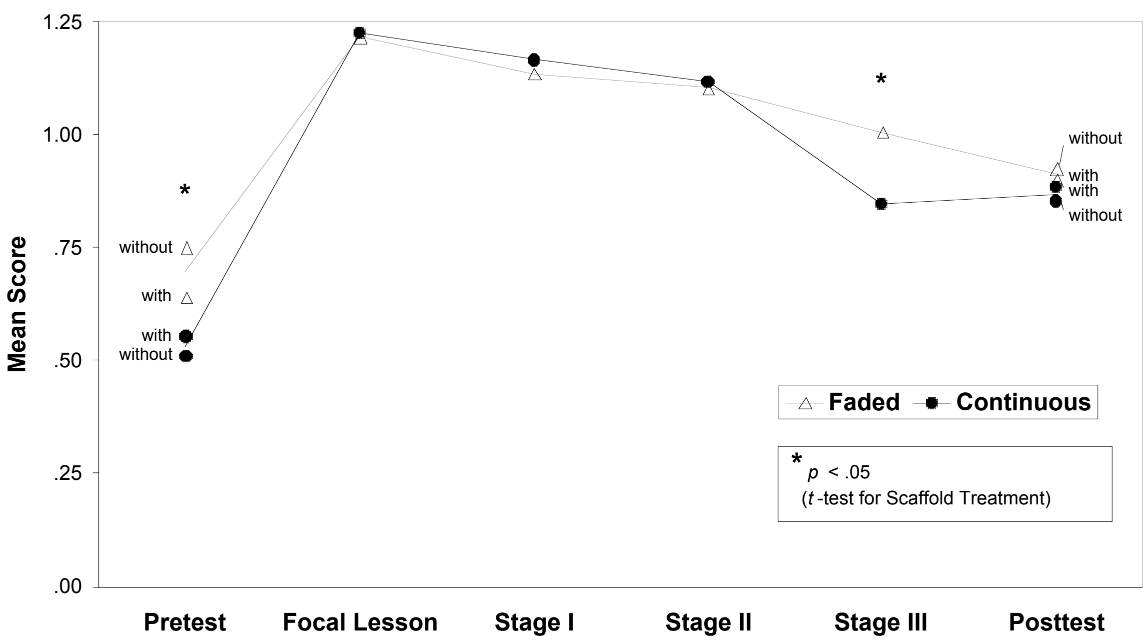

Figure 2. Evidence Scores Over the Unit

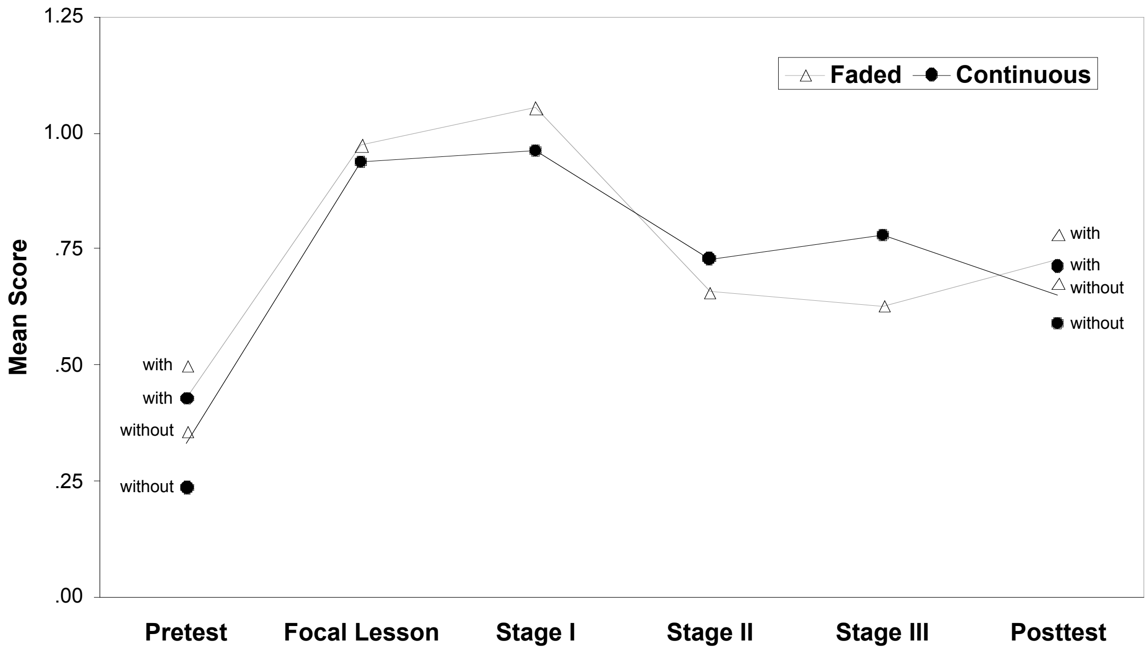

Figure 3. Reasoning Scores Over the Unit

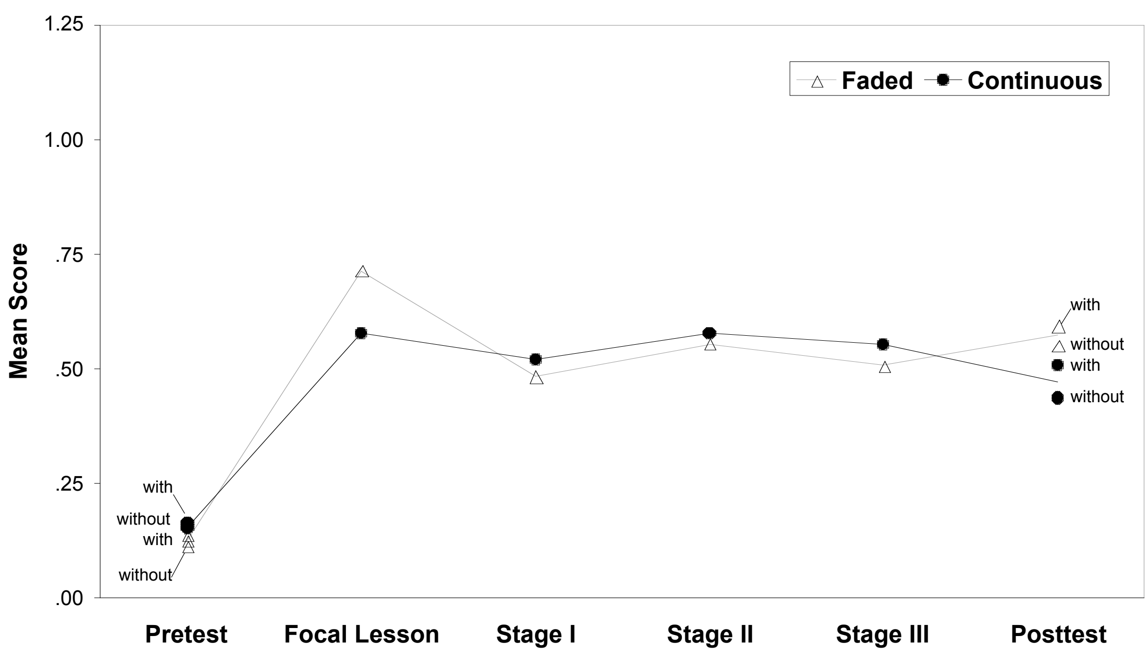

Figure 4. Influence of Scaffold Treatments on Posttest Reasoning for All Content

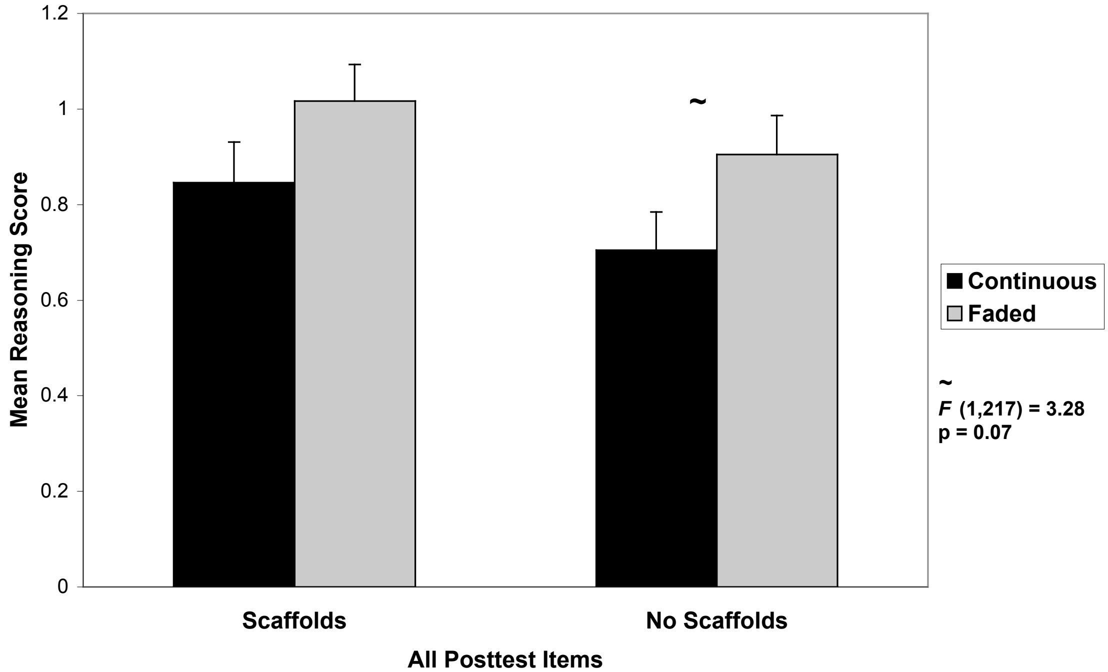

Figure 5. Influence of Scaffold Treatments on Posttest Reasoning for Substance/Property Items