Identifying Teacher Practices that Support Students' Explanation in Science

Katherine L. McNeill, David J. Lizotte & Joseph Krajcik

University of Michigan

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, 2005, Montreal, Canada.

This research was conducted as part of the Investigating and Questioning our World through Science and Technology (IQWST) project and the Center for Curriculum Materials in Science (CCMS), supported in part by the National Science Foundation grants ESI 0101780 and ESI 0227557 respectively. 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

Teacher practices are essential for supporting students in scientific inquiry practices, such as the construction of scientific explanations. In this study, we examine whether four different teacher instructional strategies, defining scientific explanation, making the rational of scientific explanation explicit, modeling scientific explanation, and connecting scientific explanations to everyday explanations, influence students' ability to construct scientific explanations during a middle school chemistry unit. Thirteen teachers enacted the chemistry unit with 1197 seventh grade students. Approximately two weeks into the unit, the curriculum materials introduce scientific explanations and have students construct and revise their initial scientific explanation. We videotaped each teacher's enactment of this focal and then coded the videotape for the four instructional strategies. We created a hierarchical linear regression model using these codes as well as students pre and posttest scores to examine the influence of these instructional strategies on student learning. We found that when a teacher discussed the rationale behind scientific explanation and modeled how to construct an explanation, that these instructional practices had a positive impact on student learning, while connecting scientific explanation to everyday explanation had a negative impact on student learning. The effect of defining scientific explanation depended on whether or not the teacher also discussed the rationale behind scientific explanation. If a teacher discussed the rationale behind explanation, defining explanation had a positive impact on student learning while if a teacher did not discuss the rationale behind scientific explanation, defining the components had a negative impact on student learning.

Identifying Teacher Practices that Support Students' Explanation in Science

Classrooms are complex systems where many factors influence student learning including tools, teachers, and peers (Lampert, 2002). Recent research (Reiser et al., 2001) and reform documents (American Association for the Advancement of Science, 1993; National Research Council, 1996) argue that the role of the teacher is essential in structuring and guiding students' understanding of scientific inquiry, a key learning goal in recent science education reform efforts. Teachers need to support students in making sense of these scientific practices (Driver, Asoko, Leach, Mortmer, & Scott, 1994). We are interested in how different teacher practices during the enactment of the same instructional unit influence students' understanding of one specific scientific inquiry practice, the construction of scientific explanations.

Role of Teachers in Inquiry

It is not enough to acknowledge that teachers play a critical role. We need to know what their role is in order in order to help support teachers in the difficult task of creating an inquiry-oriented classroom. Teachers have difficulty helping students with scientific inquiry practices such as asking thoughtful questions, designing experiments, and drawing conclusions from data (Marx, Blumenfeld, Krajcik & Soloway, 1997). Many science teachers may not have the appropriate expertise to create an inquiry-based learning environment (Krajcik, Mamlok, & Hug, 2001). Teachers need to learn new ways of teaching to promote scientific inquiry, which may differ from their own earlier socialization into school science as students (Lee, 2004; Metz, 2000) and their own orientations based on their experiences as students (Magnusson, Krajcik & Borko, 1999). Although teachers have difficulty supporting students in these practices, there is little research that provides guidance on what types of teacher practices may help students with scientific inquiry.

Research literature about inquiry classrooms often does not describe the classroom practices, rather classroom inquiry is summarized as "doing science", "hands-on science", or "real-world science" (Crawford, 2000). Furthermore, researchers often label a classroom as inquiry-oriented based on the nature of the curriculum materials used by the teacher and not by what the teacher and students are actually doing (Flick, 1995). Since teachers' beliefs about the nature of science, student learning, and the role of the teacher substantially affect their enactment of inquiry curriculum (Keys & Bryan, 2001), this raises the question of how using inquiry materials actually translates into inquiry-oriented classrooms. There is probably a range of inquiry occurring in these research studies labeled as exploring inquiry-oriented classrooms. Like other researchers (Flick, 2000; Keys & Bryan, 2001), we argue that there are few research studies that actually examine teachers' practices in inquiry classrooms.

Scientific Explanation

One prominent scientific inquiry practice in both the standards documents (AAAS, 1993; NRC, 1996) and recent research literature in science education is the construction of scientific explanations or arguments (e.g. Bell & Linn, 2000; Driver, Newton & Osborne, 2000; Jiménez-Aleixandre, Rodríguez, & Duschl, 2000; Kelley & Takao, 2002; Sandoval, 2003; Zohar & Nemet, 2002). Explanations refer to how or why a phenomenon occurs (Chin & Brown, 2000). An argument is an assertion with a justification (Kuhn, 1991) or a standpoint that is justified or defended for a particular audience (van Eemeren, et al., 1996). In our work, we use the word "explanation" to align with the national and state science standards that our teachers need to address, but our work builds off literature in both of these areas. Our goal is to help students construct scientific explanations about phenomena where they justify their claims using appropriate evidence and scientific principles.

A key goal for science education is to help students seek evidence and reasons for the ideas or knowledge claims that we draw in science (Driver et. al., 2000). Engaging students in scientific explanation and argumentation is a fundamental aspect of scientific inquiry (Duschl & Osborne, 2002). Helping students engage in this practice may help shift their view of science away from science as a static set of facts to science as a social process where knowledge is constructed. Bell and Linn (2000) found that there is a correlation between students' views about the nature of science and the arguments that they construct. They suggest that engaging students in this practice may help refine their image of science. Furthermore, engaging in scientific explanation may help students construct a deeper understanding of the content knowledge. For example, Zohar and Nemet (2002) found that students engaged in a unit on argumentation skills through dilemmas in human genetics, learned greater biological content knowledge than a comparison group who learned genetics in a more traditional manor.

Although engaging in scientific explanation is an important learning goal for students, students often have difficulty articulating and defending their knowledge claims (Sadler, 2004). Kuhn (1991) investigated both children and adults' ability to construct arguments and found that this practice often did not come naturally to them. They often had difficulty coordinating their claims and evidence. Even in a classroom setting where scientific explanation is an explicit goal, students still have many difficulties. Students can have difficulty using appropriate evidence (Sandoval, 2003) and providing sufficient evidence for their claims (Sandoval & Millwood, 2005). Students also have difficulty justifying why they chose their evidence to support their claims (Bell & Linn, 2000). In our previous work, we found that students had the most difficulty justifying why their evidence supports their claiming using appropriate scientific principles (McNeill et al., under review).

We developed a scientific explanation framework for middle school teachers and students by adapting Toulmin's model of argumentation (1958) to help support students in this difficult scientific inquiry practice. The explanation framework includes three components: a claim (a conclusion about a problem), evidence (data that supports the claim), and reasoning (a justification, built from scientific principles, for why the evidence supports the claim). In other work, we discuss the development of our framework as an instructional strategy (McNeill, et al., under review; Moje, et al., 2004) and as an assessment tool to examine student work (Lizotte, Harris, McNeill, Marx & Krajcik, 2003; McNeill & Krajcik, in press). In this paper, we are interested in exploring how teachers' different uses of the explanation framework in their classrooms influenced student learning.

Teacher Practices Supporting Scientific Explanation

In terms of supporting students in the construction of scientific explanation or argument, there has been little research on teacher practices. Previous research on students' construction of explanations in science has focused on scaffolds provided in the student materials or software program (e.g. Bell & Linn, 2000; Lee & Songer, 2004; Sandoval, 2003; Zembal-Saul, et al., 2002) or on students' discussions in order to characterize their explanations (Jiménez-Aleixandre, et al., 2000; Meyer & Woodruff, 1997).

Tabak's (2004) work has been some of the only research that has looked at the role of the teacher in helping students construct evidence based explanations. She argues that the teacher plays an important role in distributed scaffolding where many aspects of the learning environment, including software and other tools, come together synergistically to support student learning. Osborne and his colleagues (2004) have also recently begun exploring pedagogical practices that support students in argumentation. They argue that argumentation does not come naturally to students and that pedagogical practices are important for enhancing the quality of students' arguments. One of their initial findings is that teacher differences in their emphasis on different components of argument may be a result of their different understandings of what counts as an argument (Erduran, Simon, and Osborne, 2004).

To further understand the role of teachers in supporting scientific explanation, we examined the literature for instructional strategies that can support student learning of scientific explanation, but also other scientific inquiry practices, such as asking questions and designing experiments. From this literature, as well as a preliminary study we conducted on teacher practices (Lizotte, McNeill, & Krajcik, 2004) we decided to examine the influence of four instructional strategies that may help students in the construction of scientific explanations.

Table 1: Instructional Strategies That May Support Scientific Explanation

Instructional Strategy	
1	Defining scientific explanation
2	Making the rationale of scientific explanation explicit
3	Modeling scientific explanation
4	Connecting scientific explanations to everyday explanations

We now describe each of these instructional strategies and provide examples for how they may support students' successful engagement in scientific explanations.

Defining scientific explanation. What is meant by various inquiry practices such as designing experiments, asking questions or constructing explanations, is not necessarily understood by students. One instructional strategy a teacher may use to help students with these inquiry practices is to explicitly make the definition of these practices clear to students. Making scientific thinking strategies explicit to students can help facilitate their understanding and use of the strategies (Herrenkohl, Palincsar, DeWater, & Kawasaki, 1999). This may be particularly helpful for minority students. Explicit instruction may benefit diverse students who are more likely to be unfamiliar with the participation rules and practices that are an essential part of scientific inquiry (Fradd & Lee, 1999). Consequently, this type of explicitness may allow minority students to more effectively participate in classroom instruction as well as be beneficial to all students in their understanding of inquiry practices.

Metz (2000) found that being explicit about scientific inquiry practices was important for helping children with the inquiry practice of formulating and refining questions. To help elementary students ask questions, Metz provided students with different heuristics for generating researchable questions. For example, one crucial heuristic was "Can you [begin] to answer this question through collecting data?" (p. 386). Other heuristics scaffolded students brainstorming of questions for the specific science phenomena of interest (crickets) in terms of types of behaviors, conditions, and types of crickets that the children might explore. Making these heuristics explicit for students helped them to formulate researchable questions. If students had created questions without this support, they might not have generated questions that they could directly answer through the collection of data. Students required support in unpacking what exactly it means to create a researchable question.

In terms of scientific explanation, students may create stronger explanations if teachers explicitly define what is mean by a scientific explanation and define the three components, claim, evidence, and reasoning. In a preliminary study (Lizotte, McNeill, & Krajcik, 2004), we found that when teachers explicitly defined scientific explanation, particularly the reasoning component, their students constructed stronger explanations.

Making the rationale of scientific explanations explicit. Instruction should both facilitate students' ability to perform inquiry practices and their understanding of the logic behind the practice (Kuhn, Black, Keselman, & Kaplan, 2000). Helping students' understand the rationale behind why a particular scientific inquiry practice is important in science may result in students being better able to complete a performance. Chen and Klahr (1999) found that providing students with the rationale behind controlling variables in science experiments resulted in greater learning of this inquiry practice relative to students who did not receive the explicit instruction. Discussing why it is important to control variables to conduct a "fair" experiment, helped students when they had to conduct their own experiments.

For scientific explanations, it may help students to construct stronger explanations if they understand why an individual may want to construct a scientific explanation and why providing evidence and reasoning results in a stronger, more convincing explanation. Students may need help understanding why someone might want to argue for a claim. Furthermore, it might not be clear to them why providing evidence and reasoning provides greater support than just providing an opinion.

Modeling scientific explanations. Modeling various inquiry practices is another instructional strategy teachers can use to support student inquiry. Crawford (2000) argues that one of the key characteristics of a teacher establishing an inquiry-based learning environment is that they model the behaviors of a scientist. For example, the teacher Crawford researched in her case study frequently modeled how to grapple with data, specifically, through the extensive questioning of both the methods and results of data collection. Tabak and Reiser (1997) also found that student learning through collaboration in inquiry settings is more effective when teachers model profitable strategies. For example, a teacher modeling how to reason from biological data can help students complete this same process of analyzing data on their own (Tabak, 2004). By thinking aloud in front of students, students are able to see how to analyze and interpret data.

By modeling scientific explanations, students can see what this particular form of inquiry looks like in practice. Modeling how to include evidence and reasons for claims can help students in their own practice (Crawford, Kelly, & Brown, 2000). Observing others perform inquiry practices can help students learn how to translate the general scientific explanation framework into a concrete example. Defining and making the rationale behind explanation explicit for students can help students understand what scientific explanation is and the importance of scientific explanation while modeling can help students understand how to create an explanation in a domain specific context. Modeling can provide students with context-specific support. Teachers can model explanations either through writing or speaking in order to provide students with concrete examples for a specific task. Providing students with examples of strong and weak arguments can help students develop an understanding of what counts as a good argument (Osborne et al, 2004).

Connecting scientific explanations to everyday explanation. Connecting scientific discourse and inquiry practices to students' everyday discourse can help support students' learning of scientific inquiry. Lemke (1990) argues that the norms in "talking science", which are prevalent throughout written scientific texts as well as in science classrooms, contribute to the mystique of science as difficult, authoritative, and objective. Consequently, scientific discourse can be difficult for all students, particularly diverse students. Since science reflects the thinking of Western society, the norms of "talking science" are more familiar to mainstream middle-class students than students from diverse languages and cultures (Lee & Fradd, 1998). The differences between diverse students' everyday discourses versus classroom or scientific discourses may be one cause of the achievement gap.

In their discussion of achieving high science standards for students with non-English-language backgrounds, Lee and Fradd (1998) propose "the notion of *instructional congruence* to indicate the process of mediating the nature of academic content with students' language and cultural experiences to make such content (e.g. science) accessible, meaningful, and relevant for diverse students" (p. 12). Moje, Collazo, Carillo & Marx (2001) build off of this concept of instructional congruence. They argue that the way students use a scientific discourse is shaped by the everyday discourses that they bring to the classroom. In order for teachers to help students develop a scientific discourse, they need to draw from students' everyday discourses, develop students' awareness of different discourses, and make connections between students' everyday discourse and the science discourse (Moje et al., 2001). This instruction helps mediate students' navigation between different discourses helping them relate their language and culture experiences to their experiences in science classrooms.

This suggests that the instructional strategy of connecting scientific discourse to everyday discourse encompasses both pointing out differences and similarities between everyday practices and scientific inquiry practices. Teachers need to provide explicit instruction about the differences between students' everyday discourses and scientific discourses (Lee, 2004). This will help students understand that the ways of talking and thinking in science are different from those in students' everyday experiences. Focusing on science as a discourse with distinct language forms and ways of knowing, such as building theories, analyzing data, and communicating their findings, can help language-minority students learn to think and talk scientifically (Rosebery, Warren, & Conant, 1992). Students need to understand how constructing an explanation in science or supporting a claim in science looks different than in everyday life. Teachers also need to draw from students' everyday discourse (Moje et al., 2001) and make connections about the similarities between scientific discourse and everyday discourse. For example, a teacher may want to discuss how "using evidence" or "constructing an explanation" is similar in science to explanations students construct in their everyday lives.

In this paper, we examine whether teachers' use of these four instructional strategies, defining scientific explanation, making the rational of scientific explanation explicit, modeling scientific explanation, and connecting scientific explanations to everyday explanations, influences students' ability to construct scientific explanations.

Method

Instructional Context

We developed a middle school chemistry unit (McNeill, Harris, Heitzman, Lizotte, Sutherland & Krajcik, 2004) by using a learning-goals-driven design model (Reiser, Krajcik, Moje, & Marx, 2003). The learning-goals-driven design model emphasizes the alignment of the materials with national standards, both content and inquiry. During the eight-week chemistry unit, students learn about substances & properties, then chemical reactions, and finally conservation of mass both at the phenomena level and the particulate level. Besides content learning goals, the unit also focuses on scientific inquiry practices, particularly modeling of molecular representations and constructing evidence based scientific explanations. Students explore these key learning goals within the context of the driving question (Krajcik, Czerniak, & Berger, 1999), "How can I make new stuff from old stuff?" The unit is contextualized in two everyday substances, soap and lard, with the students ultimately investigating how to make soap from lard. During the instructional sequence, students experience other phenomena as well, but they cycle back to soap and lard as they delve deeper into the different content learning goals.

When designing the *Stuff* curriculum materials with my colleagues, we incorporated educative components in the material including instructional strategies to support students' in constructing explanations. By educative curriculum materials, we mean teacher materials that are specifically designed to promote teacher learning and practice (Ball & Cohen, 1996). We also provided teachers with professional development where we discussed many aspects of the unit including how to support students in the construction of scientific explanations. Although we suggest that teachers use different instructional strategies, we realize that the use and enactment of those strategies will vary by teacher.

During the unit, the materials suggest that teachers introduce students to the concept of scientific explanations through a focal lesson. This lesson occurs about two weeks into the unit after students have collected data for the various properties of fat and soap (i.e. color, hardness, solubility, melting point, and density). Initially students gather data and write a scientific explanation using their prior understanding of explanation as a guide. After students write their explanations, the materials suggest that the teachers use a number of instructional strategies to help students understand scientific explanation. For example, that the teacher introduce the "scientific explanation" framework and define the different components of explanation. As we previous mentioned our explanation framework includes three components: a claim, evidence, and reasoning. The materials provide different suggestions on how to discuss the components with students. One of the suggestions is to connect the explanation framework to an everyday example, such as making the claim that an individual is the "best singer" or the "best quarterback, and discuss how a scientific explanation is similar and different from an everyday example. The instructional materials also suggest that the teacher model how to construct a scientific explanation. The materials provide three hypothetical examples of weak and strong explanations that the teacher may use to model the use of the explanation framework to evaluate the explanations for the quality of the three components. After the discussion of scientific explanations, students then critique and rewrite their own explanations.

In the focal lesson on explanation, the instructional materials explicitly discuss three of the four instructional strategies we are interested in investigating: defining, modeling, and connecting to the everyday. Unfortunately, in looking back at the lesson we found that the idea of discussing the rationale is not explicitly discussed in the curriculum materials. Yet we decided to still look for this strategy in the videotapes to see if teachers were using it and whether it influenced student learning. Any findings about this and the other practices will inform future revision of the curriculum materials.

In order for students to learn how to evaluate data, they need numerous opportunities to evaluate rich, complex models of data (Chinn & Brewer, 2001; Lehrer & Schauble, 2002). We believe students also need numerous opportunities to engage in scientific explanations. After the focal lesson, students construct approximately ten scientific explanations during the unit[1]. Students record the results of their investigations and scientific explanations on student investigation sheets, which provide written scaffolds to support students' explanation construction. These written scaffolds provide both context-specific and generic support and fade, or provide less detail, over time. Their design was informed by research we conducted during the previous enactment of the unit where we found that fading written support over time resulted in students constructing stronger explanations on the posttest, particularly for the reasoning component (McNeill et al., under review).

Participants

Participants included 13 teachers and 1197 students from schools in the Midwest. Nine of the teachers worked in a large urban area (Urban A) where the majority of these students were African American (over 90%) and from lower to lower-middle income families. Eight of these teachers worked in public schools and one teacher taught in a charter school. Two teachers were from an independent middle school in a large college town (Town B). The majority of these students were Caucasian and from middle to upper-middle income families. One teacher taught in a second large urban area (Urban C) in a public school. The student population in this school was ethnically diverse (approximately 44% Caucasian, 34% African American, 12% Hispanic and 10% Asian) and the majority of these students were from lower-middle and middle income families. The last teacher taught in a suburb of the second large urban area (Suburb D). The student population in this school was ethnically diverse (approximately 45% Caucasian, 36% African American, 16% Hispanic and 2% Asian) and the majority of these students were from lower-middle and middle income families. As supported by the range of pretest scores at the beginning of the unit, students began the unit with a diversity of prior experiences and content knowledge.

Table 2: Participants from 2003-2004 School Year

Site	Urban A	Town B	Urban C	Suburb D	Total
Schools	8	1	1	1	11
Teachers	9	2	1	1	13
Classrooms	32	4	2	3	41
Students	1026	61	51	59	1197

Scoring Teacher Practices.

For each of the 13 teachers, we videotaped their enactment of the focal lesson on scientific explanation. To characterize teacher practices, we coded the videotape for each of the four instructional strategies that we felt might influence student learning of scientific explanation. Although students' construct scientific explanations repeatedly throughout the unit, we felt that this lesson would capture a number of the instructional strategies used by teachers because of its explicit focus on explanation. We developed the coding schemes from our theoretical framework, our experiences from a preliminary study (Lizotte et al., 2004), and an iterative analysis of the data (Miles & Huberman, 1994). Two of the codes, defining scientific explanation and modeling scientific explanation, were adapted from a preliminary study of

[1] The number of explanations may vary slightly by teacher. There are optional lessons during the unit that teachers may choose to use with their students.

videotape data from the previous enactment of the *Stuff* curriculum unit, which included six teachers (Lizotte, et al., 2004). After finalizing the coding schemes, each lesson was scored by one of two raters. We randomly selected 46% of the lessons (6 out of the 13 teachers) to be coded by both independent raters. The inter-rater reliability, determined by percent agreement, was 82%. All disagreements were resolved through discussion.

To characterize how teachers defined scientific explanation, we gave each teacher a score from 0 to 5 for each of the three components of scientific explanation (claim, evidence, and reasoning). Table 3 describes these codes.

Table 3: Codes for defining scientific explanation

Code	Description of Code
0: Does not identify	The teacher did not mention the component during the focal lesson.
1: Incorrect definition	The teacher mentioned the component, but the definition of it was inaccurate.
2: No definition	The teacher mentioned the component, but did not explicitly describe or define the component.
3: Vague definition	The teacher provided a vague definition of the component.
4: Accurate but incomplete definition	Included teachers who defined the component correctly, but the definition was incomplete. The definitions of claim, evidence, and reasoning each included two parts. Teachers who received this code only discussed one of the two parts.
5: Accurate and complete definition	The teacher provided a complete and accurate definition of the component, which included both parts.

We decided to use as the accurate and complete definitions the descriptions that were provided for the teachers in the curriculum materials. A claim is described as a statement or conclusion that answers the original question. Evidence is scientific data that is both appropriate and sufficient to support the claim. The reasoning is a justification that shows why the data counts as evidence to support the claim and includes appropriate scientific principles. For each component, the curriculum materials offer different ways of discussing these definitions with students. For example, in talking about the sufficiency of the evidence the materials suggest talking to students about providing enough data and why providing more than one piece of data may be important. The teachers did not need to use identical language to the text in the curriculum, but the language did need to align with the intent of the materials in order to receive a code of accurate and complete.

In terms of providing a rationale for constructing an explanation with evidence and reasoning, we were interested in whether teachers described a purpose for engaging in this scientific practice. Specifically, we were interested in whether teachers discussed the importance of explaining phenomena and the idea of audience. The argumentation literature discusses this idea of creating an argument to persuade or convince another individual. Only providing a claim is not as convincing as supporting a claim with these other components. We gave each teacher a score of 0, 1 or 2 depending on their level of discussion. Table 4 provides a description of these codes.

Table 4: Codes for making the rationale of scientific explanation explicit

Code	Description of Code
0: Does not mention rationale	The teacher did not mention a rationale or a reason for creating an explanation during the focal lesson.
1: Vague rationale	The teacher mentioned a vague rationale, such as explanations being an important part of science or that scientists create explanations all the time.
2: Explicit rationale	The teacher explicitly mentioned the idea of constructing an explanation for an audience. Audience is discussed in terms of the purpose of an explanation is to convince or persuade someone else about the strength of a claim.

For modeling scientific explanations, we examined how teachers used the three hypothetical examples provided in curriculum materials. In order to help students understand the quality of an explanation, we felt that a teacher would need to explicitly discuss the strengths and weaknesses of each component, rather than just provide an example or make a general statement that the explanation was good. We assigned each teacher a total of nine codes: claim, evidence, and reasoning codes for each of the three examples. Each code ranged from 0 to 5. We describe each code below in Table 5. After coding each example, we then averaged across examples to assign each teacher a mean score for each explanation component (claim, evidence, and reasoning).

Table 5: Codes for modeling scientific explanation

Code	Description of Code
0: Incorrect identification	The teacher incorrectly identified the component in the explanation. For instance, a teacher might say that an example does not include a claim when in fact it did include a claim.
1: Does not identify	The teacher did not mention whether the example included the component.
2: Identifies too much	Consisted of teachers who identified more than the component in an explanation. For instance, a teacher might say that the claim in an example was "Fat and soap are different substances. Fat and soap have different colors." The second sentence is in fact part of the evidence so the teacher has identified more than the claim in this example. This score could only apply if the example included a component.
3: Vague identification	Included teachers who made a vague statement that an explanation did or did not include the component, but did not explicitly address why the example did or did not include that component. For instance, a teacher might simply say that an example includes reasoning without discussing where the reasoning is in the example or why it counts as reasoning.
4: Identifies too little	Consisted of teachers who explicitly identified only a portion of a component. For instance, an example explanation may include three pieces of evidence and a teacher only discusses two of those pieces of evidence. A teacher could only receive this code if a component included multiple parts (e.g. three pieces of evidence).
5: Accurate and complete identification	The teacher explicitly identified the component and discussed why the component was in a particular example or how that component could be improved.

In terms of connecting scientific explanations to everyday explanations, we were interested in whether teachers brought up these types of examples as well as if they discussed the three different components in relation to everyday examples. We provided each teacher with a rating of 0-4. The codes are described in table 6.

Table 6: Codes for connecting scientific explanation to everyday explanation

Code	Description of Code
0: Does not mention an everyday example	The teacher did not mention an everyday example.
1: Discusses an everyday example, but not components	The teacher talks about an everyday example, such as basketball, tennis shoes, or allowances, but does not explicitly discuss the ideas of claim, evidence or reasoning in relation to the example.
2: Discusses an everyday example, including one components	The teacher talks about an everyday example, and explicitly talks about one of the three components (i.e. claim, evidence, or reasoning) in relation to the example.
3: Discusses an everyday example, including two components	The teacher talks about an everyday example, and explicitly talks about two of the three components (i.e. claim, evidence, or reasoning) in relation to the example
4: Discusses an everyday example, including three components	The teacher talks about an everyday example, and explicitly talks about three of the three components (i.e. claim, evidence, or reasoning) in relation to the example

Although these codes are far from exhaustive in terms of capturing what strategies the teachers used to support explanation, we were interested if they captured an aspect of teacher practice that would predict student learning of scientific explanation.

Scoring Students' Explanations.

To measure student learning of scientific explanations, we collected pre and posttest data. Students completed identical pre and posttests that consisted of fifteen multiple-choice items and four open-ended items. Three of the four open-ended items asked students to write scientific explanations, which are the focus of this analysis (Appendix A). To assess students' scientific explanations we developed specific rubrics for the tasks that rated separately students claim, evidence, and reasoning scores (see McNeill & Krajcik, in press for a description of scoring). One rater scored all items. We then randomly sampled 20% of the tests, which were scored by a second independent rater. Our estimates of inter-rater reliability were calculated by percent agreements. Our inter-rater agreement was 97% for claim, 95% for evidence, and 97% for reasoning on each explanation. Only students who completed both the pretest and posttest were included in the analysis. Because of high absenteeism in the urban schools only 833 students completed both the pre and posttests. Of these students, 51% of the students were female and 49% of the students were male.

Results

Our analyses address the following three questions: 1) Did students' explanations improve from pre- to posttest and, if so, did this improvement vary by teacher? 2) What practices did teachers engage in during the focal lesson to support students' explanations? 3) Did the teacher practices measured during the focal lesson predict student learning of scientific explanations?

Students Pre and Posttest Explanation Scores

We examined whether students explanation scores improved significantly from the pre to posttest. We summed students' explanation scores across the three explanation test items (substances, chemical reactions, and conservation of mass). We then analyzed their composite explanation score, which was a sum of their claim, evidence and reasoning scores, as well as each component separately. Table 7 provides the results from this analysis.

Table 7: Overall Student Learning of Scientific Explanation (n=833)

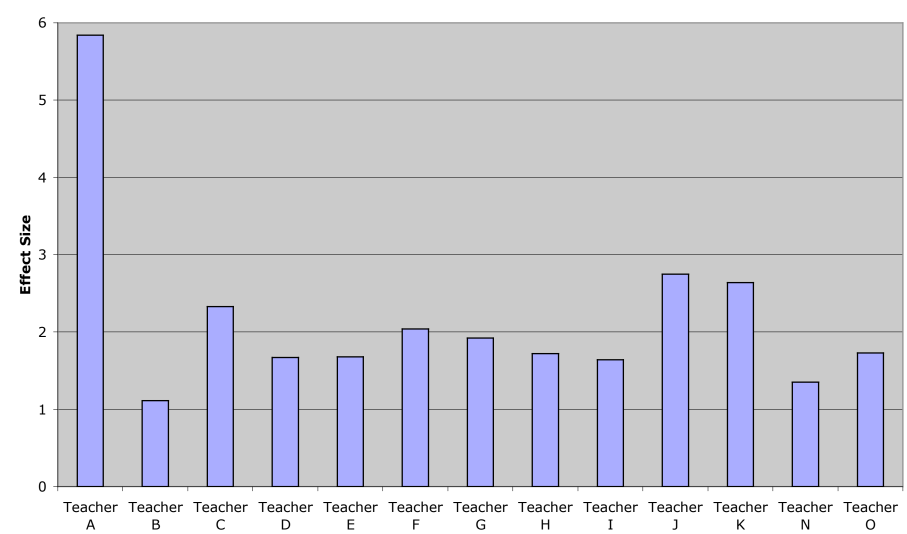
Score type	Maximum	Pretest *M* (*SD*)	Posttest *M* (*SD*)	*t* (405) [a]	Effect size [b]
Composite Score	11.25	1.37 (1.48)	4.27 (2.48)	35.16***	1.96
Component					
Claim	3.75	0.87 (1.01)	2.05 (1.18)	25.54***	1.17
Evidence	3.75	0.42 (0.57)	1.28 (0.99)	24.86***	1.51
Reasoning	3.75	0.08 (0.25)	0.94 (0.94)	26.98***	3.44

[a] One-tailed paired *t*-test
[b] Effect size is the difference between pretest *M* and posttest *M* divided by pretest *SD*.
*** $p < .001$

Students show significant learning gains on their composite explanation scores as well as on each separate component. This suggests that students are becoming more adept at constructing scientific explanations during the instructional unit. Similar to our previous research (Lizotte et al., 2004; McNeill at al, under review), we see that students have the most difficult with the reasoning component, but that the reasoning scores also demonstrate the greatest improvement from pre to post as indicated by greater effect size for reasoning compared to claim and evidence.

We also examined whether there was a significant difference in student learning between teachers. Figure 1 displays the effect sizes of the thirteen teachers for students total explanation score.[i]

Figure 1: Effect Size Across Teachers For Students' Scientific Explanations



Although each teacher's students had significant learning gains for their explanations (*ps* <.001), the effect sizes ranged from 1.11 to 5.84. We tested whether there was a significant teacher effect by performing an ANCOVA on students' posttest explanation scores with the pretest explanation scores as the covariate. There was a significant teacher effect with the student learning gains of some teachers being greater than other teachers, F (12, 820) = 16.429, $p < .001$. There was also a significant interaction between the teacher and students' pretest scores, F (12, 820) = 2.776, $p < .01$, suggesting that the effect of a students' pretest on their posttest varied by teacher. This suggests that there was differential learning of scientific explanation across teachers. Consequently, we were interested in whether our coding of the teacher practices during the focal lesson explained any of the between teacher variance.

Teacher Practices During the Focal Lesson

We were also interested in whether there was differential use of the four instructional strategies by the thirteen teachers or if all of the teachers engaged in similar practices. Table 8 displays the descriptive statistics for each of the practices. For both defining and modeling, we created an overall composite score where we summed each teacher's scores for claim, evidence, and reasoning as well as examined the scores for each component.

Table 8: Descriptive Statistics for Teacher Practices (n = 13 teachers)

	Mean (Standard Deviation)
Defining Scientific Explanation	9.85 (2.38)
Defining Claim	3.54 (1.05)
Defining Evidence	3.31 (0.75)
Defining Reasoning	3.00 (1.41)
Rationale of Scientific Explanation	0.15 (0.38)
Modeling Scientific Explanation	9.00 (2.92)
Modeling Claim	3.21 (1.13)
Modeling Evidence	3.26 (0.95)
Modeling Reasoning	2.54 (1.24)
Connecting to Everyday Explanation	0.23 (0.44)

Defining scientific explanation. In terms of defining scientific explanation, we see that all teachers do define the different components to some degree, but that the accuracy and completeness of the defining varies. Figure 2 provides the distribution for each component as well as an overall composite.

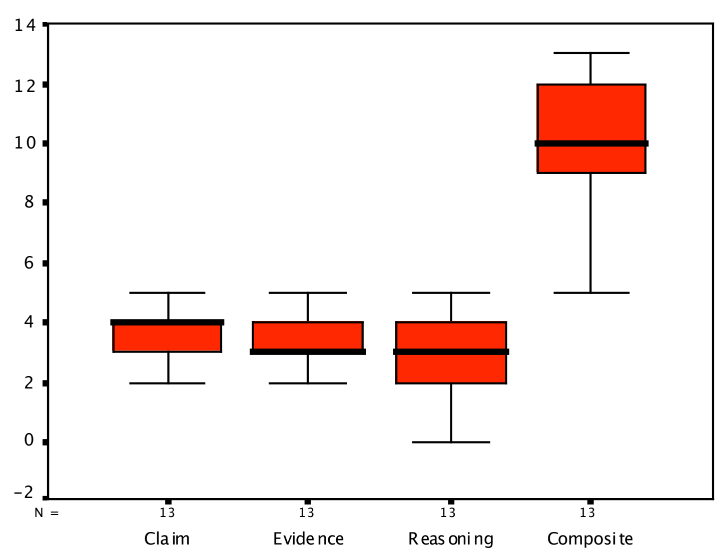


Figure 2: Defining Scientific Explanation

Interestingly, the teachers do show quite a bit of variation in terms of the degree to which they define the three components and the overall composite scores. Our coding scheme (see Table 3) included scoring a vague definition as a level 3 and an accurate and complete definition as a level 5. Of the three components, claim has the highest average score suggesting that the teachers most completely and accurately defined claim. There is the least amount of variation in terms of how the teachers discussed evidence, with most teachers simply referring to it vaguely as data with little discussion of the ideas of appropriateness or sufficiency. Finally, reasoning has the lowest average, but the greatest variation. Some teachers extensively discussed the idea of including a scientific principle to connect the claim and evidence, while one teacher did not even mention the concept of reasoning.

Making the rationale of scientific explanation explicit. Very few teachers discussed the rationale behind why an individual may want to construct a scientific explanation. Only two of the thirteen teachers discussed a rationale behind completing explanation in their classroom. Considering that the materials did not explicitly discuss this idea, it is not surprising that teachers did not include it in their classroom practice.

Modeling scientific explanation. All except one of the teachers did model how to construct a scientific explanation. Of the other twelve teachers, there was a range of discussion in terms of both the explicitness and accuracy of discussing the three different components. Figure 3 provides a distribution for each component as well as the overall composite score.

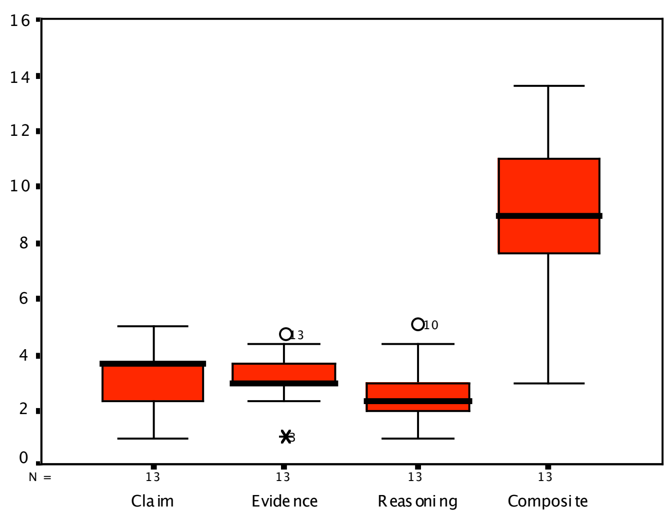


Figure 3: Modeling Scientific Explanation

Our coding scheme (see Table 5) included scoring a vague identification of a component during modeling as a level 3 and an accurate and complete identification of a component during modeling as a level 5. Similar to defining the three different components, the average claim score is again the highest suggesting that the teachers accurately modeled how to construct a claim more than the other three components. Again, evidence had the least variation while reasoning had the most variation. Teachers' discussions of reasoning both in terms of defining and modeling appeared to be the least complete and accurate of the three components though it included the greatest degree of variation.

Connecting scientific explanations to everyday explanations. Similar to discussing the rationale behind scientific explanation, connecting scientific explanations to everyday explanations rarely occurred in the classrooms that we observed. Of the thirteen teachers only three of them discussed everyday examples during the focal lesson. In two of the cases, the teachers discussed all three components in relation to an everyday example (music and basketball). In the third case, the teacher just discussed what a claim and evidence would look like in an everyday example (art).

Effect of Teacher Practices on Students' Explanations

We created a hierarchical linear regression model to determine whether there was a relationship between teachers' practices to support scientific explanation construction and student learning of scientific explanations[ii]. We z-scored the outcome variable, the explanation posttest score, so that the unstandardized regression coefficients would be equivalent to the effect size. We also z-scored the pretest in order to keep it on the same scale. The rest of the variables we left in their original metric for ease of interpretation. We wanted to be able to talk about the effect of vaguely defining the different components of a scientific explanation compared to accurately and completely defining the components, not about a change in one standard deviation. Since teachers rarely completed both making the rationale explicit and connecting to the everyday, we felt that we could not treat these as continuous variables. Rather we dummy coded both of these variables, so that the variable included in the regression model only indicates whether the teacher did (1) or did not (0) complete the practice. Since each of the thirteen teachers did receive a distinct score for both defining and modeling at the composite level, we decided to treat both of these variables as continuous. Before conducting this analysis, we also centered the two continuous variables, modeling and defining, to create the interaction terms. We calculated the interaction terms for all the predictors by multiplying each pair of predictors for a total of 15 interaction terms. The product term represents a combined effect of the two variables that is unique or goes above and beyond the separate additive effects of the two variables.

We used a hierarchical regression model because variables are grouped theoretically and then the groups are added one at a time. Model 1 includes the measures from the beginning of the unit: students' gender and students' pretest score. Model 2 includes the four measures of teacher practices: defining scientific explanation, making the rationale of scientific explanation explicit, modeling scientific explanation, and connecting scientific explanation to everyday explanation. Finally, in Model 3 we added the interaction terms using stepwise regression. The interactions are added one at a time until they no longer significantly increased the proportion of the variance in the outcome variable explained by the model at a .05 alpha level. We used step-wise regression for the interaction terms because we only wanted to include the significant interactions in the regression model.

Table 9 includes the results for the regression analysis with the explanation posttest as the outcome variable, including the unstandardized regression coefficients and significant levels for each of the predictor variables. The unstandardized regression coefficients are also in this case equivalent to the effect size because the outcome variable has a standard deviation of 1.0 since it was z-scored.

Table 9: Relationship between teacher adaptations and student learning (n=833)

Independent Variables	Model 1: Prior Knowledge	Model 2: Teacher Practices	Model 3: Interactions
Gender	0.163*	0.156**	0.211*
Pretest	0.355***	0.288***	0.326***
Defining Scientific Explanation		-0.124***	-0.184***
Rationale of Scientific Explanation		0.831***	0.572***
Modeling Scientific Explanation		0.011	0.298**
Connecting to Everyday Explanation		-0.469***	-0.455**
Rationale X Defining			0.542***
Modeling X Defining			-0.026***
Gender X Everyday			-0.327*
Pretest X Defining			0.035*
Pretest X Everyday			-0.171*
Constant	0.083	1.079***	-0.753
R^2	0.134***	0.240***	0.313***
Change in R^2		0.106***	0.073***

* p < .05; ** p < .01; *** p < .001

The first group, which included gender and the pretest score, is significant for students posttest, $F(2, 830) = 64.382$, $p < .001$. This regression model explains 13.4% of the variance in students' posttest scores. It is not surprising that students' performance on the pretest explains a large percentage of the variance on the posttest. In other words, students who score higher on the pretest are more likely to score higher on the posttest. Gender is also significant. After controlling for the pretest, females score higher on the posttest than males.

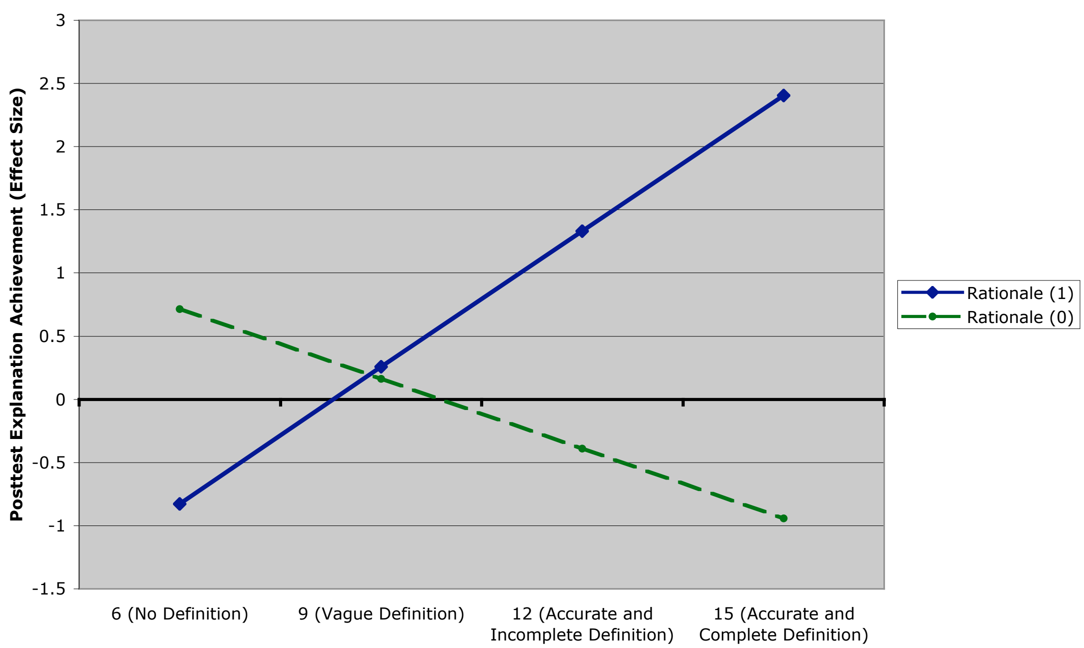
The change in the model resulting from the addition of the second group, which included the four instructional strategies or teacher practices, is significant for students' posttest scores, $F(6. 826) = 43.560$, $p < .001$. Again, the entire model is significant. Adding the teacher practices explained 10.6% more of the variance in students' posttest explanation scores. Three of the four teacher practices significantly influenced student learning of scientific explanation: defining scientific explanation, making the rationale of scientific explanation explicit, and connecting scientific explanation to everyday explanation. Whether a teacher discussed the rationale behind scientific explanation had the greatest effect. If a teacher did discuss the rationale compared to not discussing it, this resulted in significantly greater student learning of scientific explanation with an effect size of 0.831. Connecting scientific explanation to everyday explanation had a negative effect on student learning of scientific explanation. Furthermore, teachers who received a higher score for defining the components also had students with lower gains. Adding the four teacher practices also decreased the effect size of the pretest. This is important because it suggests that teacher practices can help students overcome some of their performance differences at the beginning of the unit.

The last step added five significant interactions to the regression model. Overall, this step explained 7.3% more of the variance. An interaction term suggests that the effect of one variable on student learning depends on another variable. Considering the importance of context in education, it is not surprising that the influence of one variable is going to depend on another variable. This last step is our final model and explains a total of 31.3% of the variation in students' posttest explanation scores. This final model suggests the relative importance of the variables while considering the influence of the other variables.

Adding the interaction terms altered the significance of teachers' modeling how to construct a scientific explanation. In our final model, teachers who received higher scores for modeling scientific explanation had students with higher learning gains for scientific explanation. This suggests that modeling how to construct an explanation is an important instructional strategy to help students learn how to create scientific explanations. The direction of the other three main effects for the teacher practices remained the same. Explicitly discussing the rationale behind scientific explanation resulted in greater student learning gains, while defining the components of scientific explanation and linking scientific explanation to everyday explanation resulted in lower student gains.

Besides the main effects, the interaction terms suggest some interesting relationships between the different variables. In particular, the first interaction between making the rationale explicit and defining explanation explained the greatest proportion of the increase in variance. It explained 4.5% more of the variance, which is greater than the other four interaction terms combined that add only 2.8% of the increase in variance. Furthermore, this interaction has the largest effect size. Figure 4 depicts the interaction between explicitly discussing the rationale behind scientific explanation and defining the different components of explanation.

Figure 4: Interaction Between Rationale and Defining

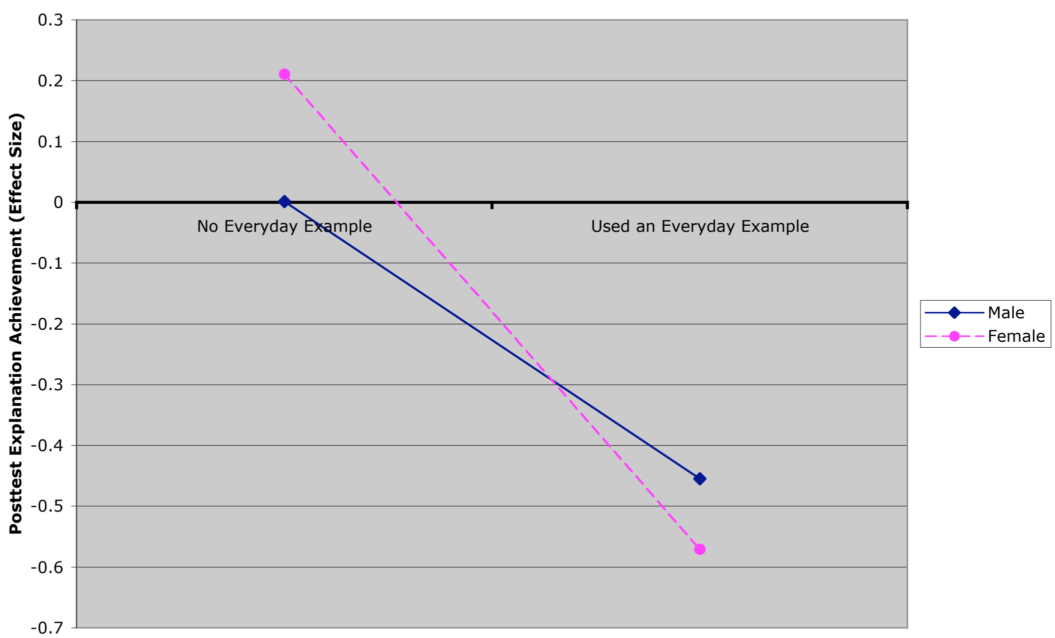


The solid line in Figure 4 represents the effect size of posttest achievement when teachers provided students with the rationale behind scientific explanation and the dotted line represents when the teachers did not provide the rationale. If a teacher discusses the rationale behind explanation, then receiving a composite definition score of above nine has a positive impact on student learning of scientific explanation. A composite score of nine could be obtained by receiving a three for each of the three components, which corresponds to providing vague definitions of all three components. A teacher who both discusses the rationale and accurately and completely defines the different components of scientific explanation has the greatest positive effect on students' posttest explanation achievement. If a teacher does not discuss the rationale behind scientific explanation, then accurately and completely defining the different components of an explanation actually has a negative impact on student learning. This suggests that the effect of one teacher practice may depend on other practices that are a part of the classroom culture.

The effect sizes of the other four interactions are smaller and explain a smaller percentage of the variance in students' posttest scores. In particular the interaction between modeling and defining and the interaction between pretest and defining have very small effects. In both these interactions, defining the components still has a negative impact on student learning though that impact is slightly less if a student's pretest score is high or if the teacher's score for modeling is low. The interaction between connecting scientific explanation to everyday explanation and students' pretest explanation scores also had a small effect. In this interaction, discussing an everyday example negatively influences student learning, but the impact is less if a student has a low pretest score.

The interaction between gender and discussing an everyday example of a scientific explanation is the only other interaction, besides the interaction between rationale and defining, which alters the direction of the effect of a variable. Figure 5 depicts the interaction between the gender of the student and whether or not their teacher discussed an everyday explanation during their discussion of scientific explanations.

Figure 5: Interaction Between Gender and Connecting to Everyday Explanations



Regardless of whether a student is female or male, discussing an everyday example has a negative impact on posttest explanation achievement. But using an everyday example had a greater negative effect on females than males. If a teacher used an everyday example, males showed greater achievement on their posttest explanations, than females. This is in the opposite direction of the main effect where female students typically showed greater posttest achievement than male students.

To summarize, the final model shows that all four teacher practices have a significant impact on student learning of scientific explanation. In terms of the main effects, discussing the rationale behind scientific explanation and modeling how to construct an explanation had a positive impact on student learning, while connecting scientific explanation to everyday explanation and defining the components of scientific explanation had a negative impact on student learning. We also found that a number of the variables in our model interacted with each other. This suggests that the influence of a variable is going to depend on the context. Particularly interesting when a teacher's discussion included both discussing the rationale and defining the components of scientific explanation, then defining the components had a positive impact on student learning. Also, while discussing an everyday example negatively impacted all students it had a greater negative impact on female students.

Discussion

The role of teachers is essential in supporting students in scientific inquiry practices (American Association for the Advancement of Science, 1993; National Research Council, 1996). Yet, like other researchers (Flick, 2000; Keys & Bryan, 2001), we argue that there have been few research studies that explicitly examine teacher instructional strategies to support students in scientific inquiry. Specifically, we are interested in supporting students in evidence based scientific explanations, which are a fundamental aspect of scientific inquiry (Duschl & Osborne, 2002). In this study, we investigated the influence of four instructional strategies on student learning: modeling scientific explanation, making the rationale of scientific explanation explicit, defining scientific explanation, and connecting scientific explanation to everyday explanation. Our results suggest that all four of these teacher practices can influence student learning of scientific explanation.

Modeling the behaviors of a scientist (Crawford, 2000) and modeling how to reason from data (Tabak, 2004) are important instructional strategies for teachers to engage in during scientific inquiry. Our results suggest that having teachers model scientific explanation can result in greater student learning of scientific explanation. Providing students with examples of both strong and weak explanations as well as critiquing the different components of a scientific explanation, may help students develop an understanding of what counts as a scientific explanation. This may allow students to construct stronger explanations when they independently construct their own explanations.

We also found that making the rationale of scientific explanation explicit for students resulted in greater student learning of scientific explanations. Instruction should help students understanding the logic behind scientific inquiry practices (Kuhn, et al., 2000). Helping students understand the rationale behind scientific explanations may help them understand why they need to include evidence and reasoning to support their claims. When teachers include this instructional strategy as part of their classroom instruction, students may obtain a stronger understanding of scientific explanation, which may help them in the construction of scientific explanations. Since so few teachers actually discussed the rationale behind scientific explanation in our study, we feel that to better understand this relationship we need to investigate more cases where the rationale is a part of classroom practice. In our model, we were only able to include the presence or absence of the instructional strategy. It would be interesting to examine the quality of discussions around the rationale to see how this influences student learning. Based on the results of this study, we will revise the instructional materials to include an explicit discussion of the rationale behind scientific explanation. Hopefully, by including the rationale in the materials more teachers will engage in this instructional strategy during their classroom instruction.

Defining the different components of scientific explanations increased student learning in some contexts, yet it decreased it in other contexts. We found that there was an interaction between providing the rationale for scientific explanation and defining the different components of scientific explanation. When a teacher provided the rationale behind scientific explanation, then defining the different components resulted in greater student learning. But when a teacher did not provide the rationale, then defining the different components of scientific explanation actually had a negative impact on student learning. Within classrooms many factors influence student learning including tools, teachers, and peers (Lampert, 2002). It is important to consider classrooms as complex systems when evaluating the effectiveness of any factor in terms of student learning. The results of our study suggest that even when looking at different teacher practices it is important to consider what other practices occurred within the classroom.

Previous research has found that being explicit about scientific inquiry practices (Herrenkohl et al., 1999) and providing students with different heuristics (Metz, 2000) can help students engage in scientific inquiry practices. Although providing students with a definition of scientific explanation and its components can help students engage in this practice, there is also the danger that explanation construction can become too algorithmic or procedural without an understanding of the inquiry practice as a whole. We conjecture that in classrooms where teachers focused on defining the parts without a discussion of the rationale behind scientific explanation as a whole, that constructing explanations became more algorithmic for students and they did not develop as deep an understanding of scientific explanation. This may explain why we see that when teachers defined the different components, but did not discuss the rationale that students had lower posttest explanation achievement.

The last instructional strategy that we examined was connecting scientific explanation to everyday explanation. We were surprised by the results of including this in our regression model. To help students develop a scientific discourse, teachers need to develop students' awareness of different discourses and make connections between students' everyday discourse and science discourse (Moje et al., 2001). Consequently, before conducting the analysis we thought that if teachers made connections between everyday explanations and scientific explanations that this would result in greater student learning. Our analysis suggests that the opposite occurred. Discussing everyday explanations in the classroom actually resulted in lower student posttest explanation achievement. Similar to our code for rationale, very few teachers engaged in this instructional strategy and we were only able to include the strategy in our model in terms of the presence or absence. It may be that this negative effect is simply a result of our small sample of teachers connected everyday explanation to scientific explanation. It is also possible that it is not the presence of this instructional strategy that is important, but rather the quality of the instructional strategy.

Our coding scheme (see Table 6) only captured whether or not teachers discussed an everyday explanation and what components they discussed. We did not code for the quality of the discussion. This is different compared to the codes we used for the other three instructional strategies. Modeling and defining both explicitly included a rating scheme from poor to high quality (0-5). For discussing the rationale, we examined the videotapes for a specific rationale, which we viewed as appropriate for engaging in scientific explanation. If a teacher provided a different rationale than the one we envisioned, we did not code that discussion as providing the rationale for scientific explanation. In order to further understand the effect of discussing everyday examples, we would need to examine more cases where teachers used everyday explanations in their classrooms and rate the quality of the examples. In reexamining the three cases where teachers discussed everyday examples, in all three instances the teachers discussed the similarities between everyday explanations and scientific explanations. In order to effectively use an everyday example, it may be more important to discuss the differences. Focusing on science as a discourse with distinct language forms and ways of knowing can help language-minority students learn to think and talk scientifically (Rosebery et al., 1992). Teachers need to discuss the differences between students' everyday discourses and scientific discourses (Lee, 2004). It may be that discussing everyday explanations is only helpful for students if it includes a discussion of the differences compared to scientific explanation, instead of only the similarities.

The small sample size of our study, thirteen teachers, may have influenced the results of our analysis, particularly for discussing the rationale behind scientific explanation and connecting scientific explanations to everyday explanations, since very few teachers used these instructional strategies. If more teachers had engaged in these practices, we may have observed that they had different effects on student learning. Furthermore, we only analyzed one lesson during the unit. The one lesson that we videotaped may not have been characteristic of the instructional strategies that teachers engaged in during the entire unit. More research needs to be conducted with a greater number of teachers, lessons, and other instructional units to further tease out the influence of each of these four instructional strategies. Yet our study provides some preliminary findings that teacher practices can play an important role in student learning of scientific inquiry practices. Even when students are engaged in the same instructional unit, differential learning occurs that can be directly linked to instructional strategies. Developing a more in depth understanding of these teacher practices is essential for supporting students in scientific inquiry practices, such as the construction of evidence based scientific explanations.

References

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

Ball, D. L., & Cohen, D. K. (1996). Reform by the book: What is—or might be—the role of curriculum materials in teacher learning and instructional reform? *Educational Researcher, 25*(9), 6-8, 14.

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.

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.

Chinn, C. A. & Brewer, W. F. (2001). Models of data: A theory of how people evaluate data. *Cognition and Instruction*, *19*, 323-393.

Chinn, 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.

Crawford, B. A. (2000). Embracing the essence of inquiry: New roles for science teachers. *Journal of Research in Science Teaching, 37*(9), 916-937.

Crawford, T, Kelly, G., & Brown, C. (2000). Ways of knowing beyond facts and laws of science: An ethnographic investigation of student engagement in scientific practices. *Journal of Research in Science Teaching, 37*(3), 237-258.

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.

Duschl, R. A. & Osborne, J. (2002). Supporting and promoting argumentation discourse in science education. *Studies in Science Education, 38*, 39-72.

Erduran, S., Simon, S. & Osborne, J. (2004) TAPing into argumenation: Developments in the application of Toulmin's argument pattern for studying science discourse. *Science Education. 88*(6), 915-933.

Flick, L. B. (1995). *Complex instruction in complex classrooms: A synthesis of research on inquiry teaching methods and explicit teaching strategies.* Paper presented at the Annual Meeting of the National Association for the Research in Science Teaching, San Francisco, CA.

Fradd, S. & Lee, O. (1999). Teachers' roles in promoting science inquiry with students from diverse language backgrounds. *Educational Researcher*, 18, 14-20.

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.

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.

Kelly, G. J., & Takao, A. (2002). Epistemic levels in argument: an analysis of university oceanography students' use of evidence in writing. *Science Education, 86*, 314-342.

Keys, C. W., & Bryan, L. A. (2001). Co-constructing inquiry-based science with teachers: Essential research for lasting reform. *Journal of Research in Science Teaching, 38*(6), 631-645.

Krajcik, J., Czerniak, C. M., & Berger, C. (1999). *Teaching children science: A project-*

Based approach. Boston, MA: McGraw-Hill.
Krajcik, J., Mamlok, R., & Hug, B. (2001). Modern content and the enterprise of science: Science education in the twentieth century. In L. Corno (Ed.), *Education across a century: the centennial volume* (pp. 205-231). Chicago: NSSE.
Kuhn, D. (1991). *The skills of argument*. Cambridge, UK: Cambridge University Press.
Kuhn, D., Black, J., Keselman, A., & Kaplan, D. (2000). The development of cognitive skills to support inquiry. *Cognition and Instruction, 18*(4), 495-523.
Lampert, M. (2002). Appreciating the complexity of teaching and learning in school: A commentary on Cobb; Froman and Ansell; McClain; Saxe; Schliemann; and Sfard. *The Journal of the Learning Sciences, 11*(2&3), 365-368.
Lee, O. (2004). Teacher change in beliefs and practices in science and literacy instruction with English language learners. *Journal of Research in Science Teaching, 41*(1), 65-93.
Lee, O. & Fradd, S. H. (1998). Science for all, including students from non english-language backgrounds. *Educational Researcher*, 27 (4), 12-21.
Lee, H.-S. & Songer, N. B. (2004, April). *Longitudinal knowledge development: Scaffolds for Inquiry.* Paper presented at the annual meeting of the American Educational Research Association, San Diego, CA.
Lehrer, R., & Schauble, L. (Eds.). (2002). *Investigating real data in the classroom: Expanding children's understanding of math and science*. New York: Teachers College Press.
Lemke, J. (1990). *Talking science: Language, learning, and values*. Norwood, NJ: Ablex Publishing Corporation.
Lizotte, D. J, Harris, C. J., McNeill, K. L., Marx, R. W., & Krajcik, J. (2003, April). *Usable assessments aligned with curriculum materials: Measuring explanation as a scientific way of knowing.* Paper presented at the annual meeting of the American Educational Research Association, Chicago, IL.
Lizotte, D. J., McNeill, K. L., & Krajcik, J. (2004). Teacher practices that support students' construction of scientific explanations in middle school classrooms. In Y. Kafai, W. Sandoval, N. Enyedy, A. Nixon & F. Herrera (eds.), *Proceedings of the sixth international conference of the learning sciences* (pp. 310-317). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
Magnusson, S., Krajcik, J., & Borko, H. (1999). Nature, sources, and development of pedagogical content knowledge for science teaching. In J. Gess-Newsome & N. Lederman (Eds.), Examining pedagogical content knowledge: The construct and its implications for science education (pp. 95-132). The Netherlands: Kluwer Academic Publishers.
Marx, R. W., Blumenfeld, P. C., Krajcik, J. S. & Soloway, E. (1997). Enacting project-based science. *The Elementary School Journal*. 97(4). 341-358.
McNeill, K. L., Harris, C. J., Heitzman, M., Lizotte, D. J., Sutherland, L. M., & Krajcik, J. (2004). How can I make new stuff from old stuff? In J. Krajcik & B. J. Reiser (Eds.), *IQWST: Investigating and questioning our world through science and technology*. Ann Arbor, MI: University of Michigan.
McNeill, K. L. & Krajcik, J. (in press). Middle school students' use of appropriate and inappropriate evidence in writing scientific explanations. In Lovet, M & Shah, P (Eds.) Thinking with Data: the Proceedings of the 33rd Carnegie Symposium on Cognition. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

McNeill, K. L., Lizotte, D. J, Krajcik, J., & Marx, R. W. (under review). Supporting students' construction of scientific explanations by fading scaffolds in instructional materials.

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.

Meyer, K., & Woodruff, E. (1997). Consensually driven explanation in science teaching. *Science Education, 80*, 173-192.

Miles, M., & Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook (2nd edition).* Thousand Oaks, CA: Sage.

Moje, E., Collazo, T., Carrillo, R., & Marx, R. (2001). "Maestro, what is quality?" language, literacy, and discourse in project-based science. *Journal of Research in Science Teaching, 38*(4), 469-498.

Moje, E.B., Peek-Brown, D., Sutherland, L.M., Marx, R.W., Blumenfeld, P., Krajcik, J. (2004). Explaining explanations: Developing scientific literacy in middle-school project-based science reforms. In D. Strickland & D. E. Alvermann, (Eds.), Bridging the Gap: Improving Literacy Learning for Preadolescent and Adolescent Learners in Grades 4-12. NY: Teachers College Press.

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

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.

Rosebery, A., Warren, B. & Conant, F. (1992). Appropriating scientific discourse: Findings from language minority classrooms. *The Journal of the Learning Sciences, 2*(1), 61-94.

Sadler, T. D. (2004). Informal reasoning regarding socioscientific issues: A critical review of research. *Journal of Research in Science Teaching*. *41*(5). 513-536.

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

Sandoval, W. A. & Millwood, K. A. (2005). The quality of students' use of evidence in written scientific explanations. *Cognition and Instruction, 23*(1), 23-55.

Tabak, I., & Reiser, B. J. (1997). Complementary roles of software-based scaffolding and teacher-student interactions in inquiry learning. In R. Hall, N. Miyake & N. Enyedy (Eds.), Proceedings of Computer Support for Collaborative Learning '97 (pp. 289-298). Toronto, Canada.

Tabak, I. (2004). Synergy: A complement to emerging patterns in distributed scaffolding. *Journal of the Learning Sciences. 13*(3), 305-335.

Toulmin, S. (1958). *The uses of argument.* Cambridge, UK: Cambridge University Press.

Van Eemeren, F. H., Grootendorst, R., Henkemans, F. S., Blair, J. A., Johnson, R. H., Krabbe, E.

C. W., Plantin, C., Walton, D. N., Willard, C. A., Woods, J., & Zarefsky, D. (1996). *Fundamentals of argumentation theory: A handbook of historical backgrounds and contemporary developments*. Mahwah, NJ: Lawrence Erlbaum Associates.

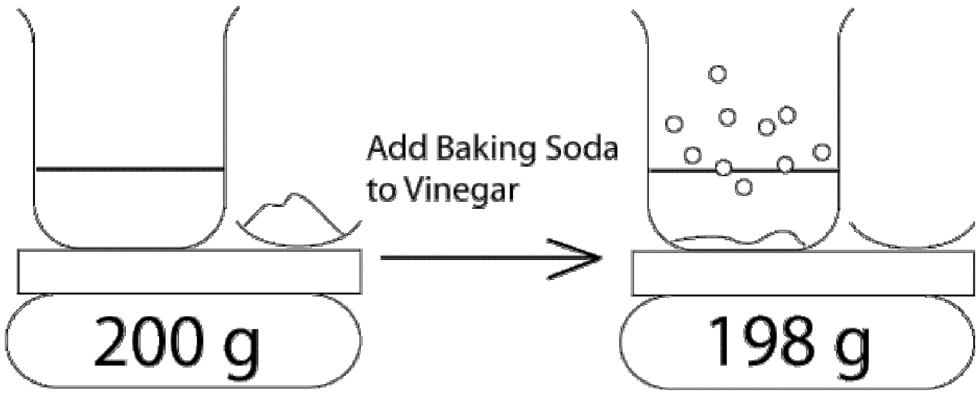
Zembal-Saul, C., Munford, D., Crawford, B., Friedrichsen, P. & Land, S. (2002). Scaffolding preservice science teachers' evidence-based arguments during an investigation of natural selection. *Research in Science Education*, *32 (4),* 437-465.

Zohar, A., & Nemet, F. (2002). Fostering Students' Knowledge and Argumentation Skills Through Dilemmas in Human Genetics. *Journal of Research in Science Teaching*. *39* (1), 35-62.

Appendix A: Scientific Explanation Test Questions

Question #1: Conservation of Mass Scientific Explanation

Dana places a beaker of vinegar and a Petri dish of baking soda on a balance. The balance reads 200 grams. Next, she pours the baking soda into the beaker with the vinegar. She does not cover the beaker. The baking soda and vinegar react and produce a gas. She places the beaker and the Petri dish back on the balance. The balance reads 198 grams.



Write a **scientific explanation** that answers the question: What is the mass of the gas produced when the baking soda reacted with the vinegar?

Question #2: Chemical Reaction Scientific Explanation

Carlos takes some measurements of two liquids — butanic acid and butanol. Then he stirs the two liquids together and heats them. After stirring and heating the liquids, they form two separate layers — layer A and layer B. Carlos uses an eyedropper to get a sample from each layer and takes some measurements of each sample. Here are his results:

		Measurements				
		Density	Melting Point	Mass	Volume	Solubility in water
Before stirring & heating	Butanic acid	0.96 g/cm³	-7.9 ˚C	9.78 g	10.18 cm³	Yes
	Butanol	0.81 g/cm³	-89.5 ˚C	8.22 g	10.15 cm³	Yes
After stirring & heating	Layer A	0.87 g/cm³	-91.5 ˚C	1.74 g	2.00 cm³	No
	Layer B	1.00 g/cm³	0.0 ˚C	2.00 g	2.00 cm³	Yes

Write a **scientific explanation** that states whether a chemical reaction occurred when Carlos stirred and heated butanic acid and butanol.

Appendix A: Scientific Explanation Test Questions

Question #3: Substance and Property Scientific Explanation

Examine the following data table:

	Density	**Color**	**Mass**	**Melting Point**
Liquid 1	0.93 g/cm^3	no color	38 g	-98 °C
Liquid 2	0.79 g/cm^3	no color	38 g	26 °C
Liquid 3	13.6 g/cm^3	silver	21 g	-39 °C
Liquid 4	0.93 g/cm^3	no color	16 g	-98 °C

Write a **scientific explanation** that states whether any of the liquids are the same substance.

Technical Notes

[i] We calculated effect size by dividing the difference between pretest and posttest means by the pretest standard deviation.

[ii] Originally, we tried to run a hierarchical linear model (HLM), because we our asking a multilevel question where students are nested in classrooms. Unfortunately, our sample size of teachers, thirteen, was not large enough to use HLM. We did not have enough variance in students' learning of explanation between teachers to model the effect of the four instructional strategies using HLM. Consequently, we decided to use regression even though we realize that the students are not independent of each other (i.e. students with the same teacher will show more similar learning gains than students with different teachers).