

Spring 7-18-2013

An Investigation into Teacher Support of Science Explanation in High School Science Inquiry Units

Rebecca Sue Hoffenberg
Portland State University

Follow this and additional works at: https://pdxscholar.library.pdx.edu/open_access_etds



Part of the [Educational Methods Commons](#), [Science and Mathematics Education Commons](#), and the [Secondary Education and Teaching Commons](#)

Let us know how access to this document benefits you.

Recommended Citation

Hoffenberg, Rebecca Sue, "An Investigation into Teacher Support of Science Explanation in High School Science Inquiry Units" (2013). *Dissertations and Theses*. Paper 1103.
<https://doi.org/10.15760/etd.1103>

This Thesis is brought to you for free and open access. It has been accepted for inclusion in Dissertations and Theses by an authorized administrator of PDXScholar. Please contact us if we can make this document more accessible: pdxscholar@pdx.edu.

An Investigation into Teacher Support of Scientific Explanation in High School Science

Inquiry Units

by

Rebecca Hoffenberg

A thesis submitted in partial fulfillment of the
requirements for the degree of

Master of Science Teaching
in
General Science

Thesis Committee:
William Becker, Chair
Emily Saxton
Cary Sneider

Portland State University
2013

© 2013 Rebecca Hoffenberg

Abstract

The Framework for K-12 Science Education, the foundation for the Next Generation Science Standards, identifies scientific explanation as one of the eight practices “essential for learning science.” In order to design professional development to help teachers implement these new standards, we need to assess students’ current skill level in explanation construction, characterize current teacher practice surrounding it, and identify best practices for supporting students in explanation construction. This multiple-case study investigated teacher practice in eight high school science inquiry units in the Portland metro area and the scientific explanations the students produced in their work samples.

Teacher Instructional Portfolios (TIPs) were analyzed with a TIP rubric based on best practices in teaching science inquiry and a qualitative coding scheme. Written scientific explanations were analyzed with an explanation rubric and qualitative codes. Relationships between instructional practices and explanation quality were examined.

The study found that students struggle to produce high quality explanations. They have the most difficulty including adequate reasoning with science content. Also, teachers need to be familiar with the components of explanation and use a variety of pedagogical techniques to support students’ explanation construction. Finally, the topic of the science inquiry activity should be strongly connected to the content in the unit, and students need a firm grasp of the scientific theory or model on which their research questions are based to adequately explain their inquiry results.

Table of Contents

Abstract.....	i
Introduction	1
Literature Review	5
The Importance of Scientific Explanation	5
Student Difficulties with Scientific Explanations	8
Instructional Strategies to Support Scientific Explanations.....	13
Summary.....	17
Research Question.....	18
Methods.....	19
Overview	19
Participants	20
Study Design	23
Instruments.....	25
Procedure.....	31
Results.....	33
Case Descriptions.....	33
Abigail.....	33
Cathy	34
Felicia	36
Joe	39
Margot	40
Penny	42
Sonia.....	43
Valerie	45
Teacher Instructional Portfolios (TIPs).....	47
Quantitative Analysis	47
Qualitative Analysis.....	48

Student Work Samples	50
Explanation Scoring Examples	50
Quantitative Analysis	53
Qualitative Analysis.....	55
SWS and TIP Relationships.....	58
SWS Categories vs. TIP Composite Scores.	59
TIP Categories vs. SWS composite scores	60
Discussion	62
Characteristics of Teacher Instruction	62
Classroom Roles	62
Content/Cognitive skills	65
Assessment for Learning.....	66
Explicit Support for Explanation.....	68
Summary	69
Characteristics of Students' Scientific Explanations	70
Composite Explanation	70
Reasoning.....	71
Defense	72
Templates.....	73
Data Displays.....	74
SWS and TIP comparisons.....	74
High TIP – High SWS – Abigail and Sonia	75
High TIP – Low SWS – Felicia and Joe	78
Low TIP – High SWS – Margot and Penny.....	84
Low TIP – Low SWS – Cathy	90
Discussion Summary	91
Limitations of the study	95
Implications for practice	96
Implications for professional development.....	98
Questions for further research	99

References	101
Appendix A.....	104
Appendix B.....	106
Appendix C.....	109
Appendix D.....	111

Introduction

Science educators aim to teach students how to think about the world in a rational way, and to demonstrate the value of a scientific approach in many areas of life. A scientific perspective does not merely value a command of facts for its own sake, but the ability to apply scientific concepts to explain observations, bring evidence to bear in decision making, and to build a clear and convincing case to support one's proposition. The recently-released Next Generation Science Standards (Achieve, Inc., 2013) (NGSS) support this approach, emphasizing scientific explanation construction and arguing from evidence as two of the eight essential skills for students learning science. With the expected adoption of the NGSS, science educators will need adapt their curricula to incorporate the practices, crosscutting concepts and core ideas on which the standards are based (National Research Council, 2012). Teachers will need considerable support from professional development providers to become familiar with the new standards and to implement them effectively. To lay the foundation for continued development, it is necessary to characterize students' skill levels and current teacher practices surrounding these essential skills.

According to *A Framework for K-12 Science Education: Practices, Crosscutting Concepts and Core Ideas* (National Research Council, 2012) (hereafter referred to as "the Framework"), the foundational document for the NGSS (Achieve, Inc., 2013):

The goal of science is the construction of theories that can provide explanatory accounts of the features of the world. ...Scientific explanations are explicit applications of a theory to a specific situation or phenomenon....

Current Oregon state standards also require students to be proficient in constructing scientific explanations, though explanation is combined with other inquiry skills and not identified as a separate practice. The Oregon state standards adopted in 2009 define scientific inquiry as:

...the investigation of the natural world by a systematic process that includes proposing a testable question or hypothesis and developing procedures for questioning, collecting, analyzing, and interpreting multiple forms of accurate and relevant data to produce justifiable evidence-based explanations and new explorations” (Oregon Department of Education, 2009).

Standard H.3S.3 for high school science inquiry states that students should be able to:

Analyze data and identify uncertainties. Draw a valid conclusion, explain how it is supported by the evidence, and communicate the findings of a scientific investigation (Oregon Department of Education, 2009).

Both the definition of science inquiry and the cited standard incorporate the concept of scientific explanation, but are not as explicit as the NGSS in emphasizing explanation construction as an essential skill. Because Oregon science educators have this foundation, it is especially important to characterize their current practices surrounding science inquiry in general and explanation in particular to implement effective professional development.

Scientific explanation in the context of science inquiry activities is grounded in the learning theory of cognitive apprenticeship as developed by Brown, Collins and Duguid (1989). This theory holds that true learning is inseparable from the culture and

conditions in which it is acquired. Just as tradespeople learn through apprenticeship, assisting master tradespeople in activities that are useful and relevant to success in the trade, so students gain authentic knowledge of an academic field by engaging in activities similar to those of practitioners of the discipline. Likewise, students learn how scientific knowledge is constructed by engaging in open-ended inquiry activities in a social context with the goal of producing scientific explanations (Sandoval, 2003).

Inquiry-based instruction in science classrooms can be differentiated for diverse learners. Rosebery et al. (1992) demonstrated that teaching science with open-ended inquiry activities helps English language learners engage in science classes and reason more like scientists. Fradd and Lee (1999) argue that explicit instruction in science inquiry classrooms benefits students from diverse language backgrounds. These two studies, taken together, argue for a balanced approach between student-centered and teacher-centered instruction around science inquiry to maximize benefits for diverse learners.

Scientific explanation is one of the goals of science inquiry, and students will be more successful in inquiry activities if they understand this goal (Sandoval, 2003). However, the education literature demonstrates that students have difficulty in producing high quality scientific explanations. Sandoval (2003) showed that students do not understand the importance of citing data to support a claim. Kuhn and Reiser (2005) demonstrated that students need support in defending claims and arguments, and Ruiz-Primo, Li, Tsai, and Schneider (2010) found that students' written explanations

often lacked the three crucial elements of a full explanation: claim, evidence and reasoning. Studies have also found that the teaching of scientific explanation is inconsistent, and that teachers need support in their instructional practices around this skill (Ruiz-Primo et al., 2010; McNeill and Krajcik, 2008).

In the fall of 2010, science teachers in the Portland, OR, metropolitan area attended a professional development workshop in which they collaborated to plan and develop science inquiry (SI) or engineering design (ED) units. After implementing the units in their classrooms, the participants submitted Teacher Instructional Portfolios (TIPs) and all students' work samples (SWSs) from the unit for further analysis. Forty-five teachers completed the program. Eight of them provided TIPs and work samples for high school SI units and consented to their use for research.

This multiple-case study employed qualitative and quantitative methods to investigate how these eight teachers supported students in constructing scientific explanations in SI activities and how their students performed. We explored connections between teacher practice and student explanation quality to help teachers understand which practices are most effective in supporting students in developing this important skill.

Literature Review

The science education literature documents the importance of scientific explanation skills in science inquiry. There are also investigations into students' explanation abilities and the areas in which they struggle; and teacher practice around supporting explanation construction.

The Importance of Scientific Explanation

As noted in the introduction, the Framework (National Research Council, 2012) is explicit about the importance of scientific explanation in science inquiry, identifying it as one of the eight practices deemed “essential elements of the K-12 science and engineering curriculum.” The Framework asserts that explanations are the means by which scientists link observations of phenomena with scientific theories and propose causal mechanisms that can “illuminate the nature of observed phenomena, predict future events, or make inferences about past events.” In learning how to construct explanations, students will develop an understanding of the scientific meanings of “theory,” “hypothesis” and “model” and how they compare to the everyday usage, a source of considerable confusion and obfuscation in current policy debates involving scientific issues. In constructing explanations, students will also gain deeper knowledge of the major theories and models underlying current scientific knowledge, such as the

theory of evolution, kinetic-molecular theory and the greenhouse-gas model of climate change; and how these models can explain data patterns or observed phenomena.

Competing explanations for phenomena such as climate change can lead to rich discussions in the classroom about how well each explanation is supported by the data, is parsimonious and satisfies other scientific values. Constructing model-based or data-based explanations is a powerful skill that gives students opportunities to learn science content more deeply and gain profound insight into the nature of science and the practices and values of professional scientists.

In their study of epistemic and conceptual scaffolds for science inquiry activities, Sandoval and Reiser (2004) argue that helping students develop an understanding of how scientific knowledge is created is one of the primary goals of science inquiry activities. These researchers focus on scientific explanation as one of the most important aspects of scientific epistemology. Convincing scientific explanations, they say, have a clear causal claim and data to support the claim. The type of causal reasoning needed varies according to the theory that frames the investigation and the scientific discipline to which it belongs. For example, a claim that natural selection has acted to change a trait in a species requires a certain type of argument framed by the theory of natural selection, which would necessarily include evidence of an environmental stressor and a variation in traits in a population. A claim that the results of an experiment support a theory or model in a physics inquiry would require a theoretical prediction of an experimental result derived from that model and

verification of the prediction with experimental data. Students will learn more about the nature of science and the mental habits of scientists if they practice developing scientific explanations that conform to these criteria.

Sandoval and Reiser (2004) extensively develop the rationale for why supporting students in creating scientific explanations teaches them how scientific knowledge is built by working scientists, and why this is crucial in disabusing students of the idea that science is a collection of facts to be memorized and not questioned. If students understand how scientific knowledge is built, they can become more sophisticated consumers of science knowledge and more critical of scientific claims in popular culture.

In another study, Kuhn and Reiser (2005) describe scientific explanation as the culmination of the cognitive processes of “sense making,” or constructing personal explanations for natural phenomena; articulating explanations, and defending them. Their study of student difficulties with these processes, described in detail below, found that the defense criterion was where many students stumbled, and concludes with ideas about why claim defense is important. Scientific knowledge is based on consensus in the scientific community, where claims are “critiqued, debated and revised” (Kuhn & Reiser, 2005). Teaching students how to defend scientific claims, and giving them opportunities to practice these skills among their peers, is engaging and helps emphasize the importance of this scientific practice.

Zohar and Nemet (2002) investigated the effects of explicit instruction in argumentation in ninth grade biology classrooms. The type of argumentation they

studied involved moral dilemmas related to biology content knowledge. Their treatment, described in more detail below, produced significant gains in students' content knowledge compared with the control group. They reasoned that gains in content knowledge were due to engagement of students' higher-order thinking while learning content, connection to interesting social/moral dilemmas, more student-centered instruction (group work, discussion in pairs, etc.) and the social context of knowledge construction. Given that scientific explanation is a complex skill requiring critical thinking and complex reasoning, it is not surprising that students struggle to master it.

Student Difficulties with Scientific Explanations

Sandoval (2003) probed student difficulties with scientific explanation using a software program he developed to scaffold explanation construction of evolutionary phenomena using natural selection. He explored students' "epistemologies of science: beliefs about the nature of science and scientific knowledge." (Sandoval, 2003)

Sandoval (2003) went on to clarify that "...students' epistemologies of science include their ideas about what scientific theories and explanations are, how they are generated, and how they are evaluated as knowledge claims." The criteria he evaluated in the students' explanations were causal coherence and evidentiary support. He judged causal coherence by whether students supported their claims with causal mechanisms and whether the chain of cause and effect was sensible. He judged evidentiary support

by whether the students used evidence to support their claims, and whether they stated evidence explicitly in their explanations.

The author analyzed how the students used the software package he developed and the quality of the explanations they produced with it. The software prompted the students to consider four specific aspects needed to support a claim about natural selection, and provided students with data of various types related to a real-world biological phenomenon. The software also contained many kinds of data that may be generated in a natural selection investigation, and tracked which data the students looked at while developing their explanations. The 69 students in his sample, working in 19 groups, were in three introductory high school biology classes with the same teacher. Sandoval evaluated their explanations based on articulation of the causal claims in each of the four aspects, causal coherence, whether the groups looked at enough data to support their claims (warrant), and whether they explicitly cited data as part of their explanation.

He found that 74% of the groups articulated claims in all four of the categories required for a natural selection argument. The software prompts were likely responsible for this success. Almost all of the groups looked at data that would support their claims in the first two categories, but the percentage of warranted claims dropped for the second two categories. The author conjectured that the data supporting the first two links in the causal chain were easy to understand, but the increased complexity and different types of data needed to warrant the claims related to the second two aspects

confused the students and caused the decrease. The percentage of groups who explicitly cited data to support their claims ranged from 75% in the first claim category to 33% in the last two categories. The mean causal coherence score was 0.7, on a scale of 0-1. The most common problems with causal coherence were “lack of clear causal language and unconnected causal claims” (Sandoval, 2003). Sandoval (2003) concluded that students understood the importance of showing causal mechanisms and causal links in a scientific explanation, but viewed data as “something to be explained, but not necessarily as a necessary component of an argument” (Sandoval, 2003). They also did not view lack of data as an argument against a claim.

Kuhn and Reiser (2005) also studied student difficulties with producing scientific explanations in the context of a middle school science inquiry unit that the authors helped design. When the researchers analyzed the written explanations for two of the science inquiry activities in the curriculum, they noticed that students generally were able to make sense of the phenomena and articulate explanations, but they struggled with the “defense” criterion. The authors identified two aspects of explanations that supported explanation defense: differentiation between evidence and inference, and inclusion of overtly persuasive statements.

The authors collected students’ written explanations from three classes that were using the same biology inquiry unit. Two were suburban classrooms, one in a middle school and the other in a K-8 magnet school, and the third was in a 7-12 urban school. The researchers analyzed 92 explanations produced by 53 students. One of the

inquiry activities required students to produce explanations individually and the other had them working in pairs.

The best defenses, Kuhn and Reiser (2005) argue, have a clear differentiation between data and interpretation, so that the audience can discern which statements come directly from data sources and which are the scientists' interpretation of the data. The authors found that 45% of the explanations they analyzed lacked a clear distinction between data and inference. Though overt statements of correctness ("My hypothesis is correct because...") were not required as part of an explanation, the researchers evaluated them as a clear attempt to persuade, thus showing that the students were attending to the defense requirements. Only 29% of the explanations contained persuasive statements.

The authors suggest that they may change the curriculum to give students more opportunities to engage in defense discourse among their peers. They may also revise their "claims-evidence-reasoning" model of teaching explanation to: "What is the answer? Why does that make sense scientifically?" and "How do you know you're right?" to make the defense criterion more explicit. While this re-wording of the components of explanation may help focus students' attention on the persuasive elements of explanation, the claim-evidence-reasoning terminology is more aligned with the "Engaging in Argument from Evidence" practice in the Framework (National Research Council, 2012) which focuses more on the persuasive aspects of scientific discourse.

Ruiz-Primo et al. (2010) documented difficulties students had in writing scientific explanations in science inquiry (SI) activities. The authors developed a rubric, described in detail in the Methods section below, to evaluate the quality of scientific explanations in students' science notebooks in an SI unit on density. Using their rubric, the authors found that most of the explanations lacked important elements.

The authors evaluated the scientific explanations from the science notebooks of students from eight middle-school classrooms that were using the SI density unit. The authors chose nine notebooks from each classroom, for a sample size of 72. They scored the scientific explanations in the notebooks based on three required elements: claim, evidence and reasoning. They also tested the correlation between a composite explanation score and summative assessment scores.

The study found that only 18% of the notebooks contained scientific explanations with the three required elements of claim, evidence and reasoning; 40% of the explanations were claims without any supporting evidence. The correlations between composite explanation quality scores and summative assessment scores were positive among all types of assessment questions, though the highest correlation was with the performance questions. The performance question asked students to design and conduct an experiment to investigate a problem with provided materials.

The authors concluded that the ability to construct high-quality scientific explanations might be linked to student learning as measured by summative assessments. They also characterized some aspects of students' science notebooks and

teachers' guidance in using the notebooks that make them more useful in developing and assessing students' understanding of science inquiry activities.

Instructional Strategies to Support Scientific Explanations

Because of the importance of scientific explanation to students' learning and their struggles in producing them, a number of researchers have designed interventions to support students in this activity. Zohar and Nemet (2002) studied the effects of explicit instruction in argumentation using moral dilemmas in ninth grade biology classrooms. Their study design consisted of an experimental group of five classes with a total of 99 students, and a comparison group of four classes with a total of 87 students. In the experimental groups, the authors presented a unit that gave students numerous opportunities to engage in argumentation, and included a lesson specifically on how to formulate an argument. Participating teachers taught the experimental groups that a complete argument includes a stated opinion and reasoning to back it up. At least some of the reasoning should be related to science content. The dilemmas in this unit were related to genetic diseases, and the content knowledge included whether a condition is dominant or recessive, and the probability that a person will have a disease or be a carrier. In the comparison groups, the researchers provided a booklet that addressed the same content in a "traditional textbook approach" (Zohar & Nemet, 2002) and had the students solve standard genetics problems, without the lesson in argumentation and the moral dilemma curriculum.

By analyzing written pretests and posttests and audio taped discussions, the authors found that students who received explicit instruction on how to engage in argumentation showed significant improvement in both their argumentation skills and content knowledge in relation to the comparison group. Students in the experimental group showed large gains in their argumentation and content scores over just one lesson. The authors reasoned that most students already possess basic argumentation skills, but that traditional instruction rarely calls upon students to exercise or improve these skills. They also pointed out that their striking results were different from other studies that involved scientific explanations of causal relationships, likely because moral dilemmas are more engaging for students and involve opinions and values, which are not necessarily right or wrong. Students are also more accustomed to discussing and defending their own opinions than supporting scientific causal relationships with data. Though argumentation in a moral dilemma context may not target exactly the same construct as scientific explanation defending a causal claim, it may be a useful scaffold to introduce students to the process of scientific argumentation in an engaging way that draws on existing skills and interests.

In another study, McNeill and Krajcik (2008) investigated instructional practices related to teaching middle school students how to write scientific explanations. They videotaped thirteen teachers presenting a lesson in a chemistry unit that focused on how to construct complete scientific explanations. The researchers rated the teachers on four aspects of instructional practice in the lesson: defining, modeling, and explaining

the rationale for scientific explanations, and making connections between scientific and everyday explanations. The authors also collected pretest and posttest data to evaluate students' progress toward the learning goals, and calculated correlations among the quality of the implementation of the four practices and student achievement. The pre- and posttests contained multiple choice and free-response items. Three of the four free-response questions required students to provide scientific explanations for concepts covered in the unit.

The study found that the largest positive effect on student achievement, as measured by posttest scores on the three open-ended explanation items, came from explaining the rationale for science explanations. Connecting scientific explanation to everyday experience had a negative effect on learning. Defining scientific explanation in isolation also had a negative effect on learning, but when teachers combined the defining practice with providing an explicit rationale for scientific explanations, there was a positive effect. Modeling scientific explanation did not have a significant effect on student learning.

The authors pointed out that small sample size may have affected their analysis of the practices of providing rationale and everyday examples. Since the curriculum materials provided by the researchers did not instruct teachers to state an explicit rationale for scientific explanation, only two of the thirteen teachers did so. Only three of the teachers connected science explanation to everyday explanation. Because of the

surprising results surrounding these two practices, the authors point to these as important areas of future research.

In the study by Ruiz-Primo et al. (2010) of scientific explanations in an SI unit on density, described above, the authors found that students from the same classroom tended to produce similar patterns of explanation quality, suggesting that the teachers' practice influenced the students' ability to produce them. The authors concluded that despite emphasis in the standards on constructing scientific explanations in science inquiry, teachers are not consistently teaching this skill, or not stressing all the required components.

Both Ruiz-Primo et al. (2010) and Sandoval (2003) commented on how the degree of scaffolding or the use of teacher-provided templates influences students' explanations. Sandoval (2003) showed that scaffolding and prompting can help students develop coherent causal chains that can plausibly explain data, but teachers need to use the results from science inquiry activities to drive classroom discussions about using data and citing it explicitly to support scientific claims. Ruiz-Primo et al. (2010) commented that instructions limited to sections headings such as "Conclusions" or "My Claim" were insufficient to focus students' responses, whereas a template with a very high level of guidance tended to discourage students from relying on their own thinking.

These studies researchers suggest some important questions: What are teachers doing in classrooms to support students' skills in scientific explanation construction?

How effective are current teaching practices in helping students meet the current science inquiry standards, and do they address the explanation standards articulated in the Framework? And, how can we support teachers and students in achieving the standard of producing complete scientific explanations as part of their science inquiry experiences? The current study attempts to answer these questions.

Summary

Scientific explanation is a central goal in science education, but teachers are not consistently supporting students in achieving it. The Framework details the importance of explanation construction in science education and justifies its place among the eight practices students must learn (National Research Council, 2012). Sandoval and Reiser (2004), Ruiz-Primo et al. (2010) and Zohar and Nemet (2002) argue convincingly that learning how to produce quality scientific explanations is crucial to advancing students' scientific understanding. Explanation engages students' higher order thinking skills, and supports content understanding and model-based reasoning.

Sandoval (2003), Kuhn and Reiser (2005) and Ruiz-Primo et al. (2010) document specific difficulties that students have with this important skill. Sandoval (2003) demonstrated that students need support in developing coherent causal claims. He also showed that students have difficulty deciding which data to cite to defend their claims and how much data is sufficient. Kuhn and Reiser (2005) concluded that students need support in learning how to defend their claims, an important aspect of knowledge-

building in science. Ruiz-Primo et al. (2010) showed that few scientific explanations found in middle school science notebooks contained complete explanations as defined by including claims, evidence and reasoning.

Researchers have also investigated interventions and instructional practices related to this problem. Sandoval (2003) tested a software program that scaffolded how to produce a complete causal claim for a natural selection problem. Zohar and Nemet (2002) showed that presenting a genetics unit in the context of moral dilemmas, combined with explicit instruction in argumentation, produced large gains in the quality of students' arguments. McNeill and Krajcik (2008) demonstrated that when teachers combined explicit definitions of explanation elements with a rationale for why this type of argumentation is important, there was a positive effect on student learning. Ruiz-Primo et al. (2010) concluded that despite the importance of scientific explanation skills for student learning and the emphasis on explanation in SI standards, teachers are not consistently stressing explanation in their instruction.

Research Question

This study investigated what instructional strategies were evident in teacher instructional portfolios for supporting students in constructing scientific explanations in high school science inquiry activities, evaluated the quality of the explanations the students produced, and identified possible connections between the teachers' instructional strategies and their students' work samples.

Methods

Overview

The goals of this multiple case study were to:

1. Investigate what instructional strategies were evident in Teacher Instructional Portfolios (TIPs) for supporting students in high school science inquiry units,
2. Evaluate the quality of the scientific explanations present in the students' science inquiry (SI) work samples, and
3. Identify possible connections between the teachers' instructional strategies and their students' scientific explanations.

The cases in this study were the eight high school classrooms that submitted SI work samples and TIPs for a professional development workshop (described below) with teacher consent to make these materials available for research analysis. This study used a mixed-methods analysis to achieve its goals. For goal 1, the TIP for each case was scored with a rubric developed by Saxton and Rigelman (unpublished) based on best practices in the research literature for teaching science inquiry; qualitatively coded with categories specific to supporting explanation adapted from the research literature; and holistically described in case descriptions using information from the TIPs, teacher and class demographics. For goal 2, the student scientific explanations present in eight science inquiry work samples chosen at random from each class were scored with an

explanation rubric adapted from Ruiz-Primo et al. (2010) and qualitatively coded with categories adapted from the research literature. For goal 3, relationships between the teacher instructional practices and the students' scientific explanation scores were explored to identify possible links between teacher instruction and the quality of students' explanations.

Saxton and Rigelman (unpublished) collected TIPs and student work samples from eight high school teachers' science inquiry units. The TIPs consisted of three sections containing knowledge, skills and experience (KSE) outcomes, assessment practices and pedagogical strategy reflections. See the "Instruments" section below for a detailed description of the TIP contents. The student work samples were the SI reports the students submitted for grading after completing the science inquiry units.

Participants

Context. The teachers in this study were participating in a professional development (PD) workshop at Portland State University's Center for Science Education (CSE) in the fall of 2010 in which they collaborated to design units that culminated in an SI or ED work sample. After implementing the units, the participating teachers submitted TIPs and students' work samples for further analysis. The teachers were asked if they would be willing to make these materials available for research. Only those who consented to research use were included in this study. Most of the teachers

who participated in the workshop were recruited based on their attendance at other PD workshops conducted by the CSE.

Number of cases. Eight teachers from public high schools in the Portland, Oregon metropolitan area submitted the requested materials for science inquiry units and consented to research analysis of the materials. One of the TIPs submitted did not conform to the expected criteria in that it seemed to contain materials for several different units. The work samples for this unit were included in the SWS analysis, but the TIP was not scored.

Number of student participants. The classes analyzed for this study varied in size, and not all students in every class participated. Work samples from students who did not consent were removed from the research. The number of students who submitted work samples for the study was 196. Eight work samples from each class were chosen at random for analysis.

Class Demographics. Table 1 shows the size range, ethnic composition and percent with limited English proficiency of the classes in this study. The school districts provided aggregate demographics for the classes. Some of the district data appeared to be combined from multiple classes taught by the instructor. The individual classroom teachers filled out Surveys of Enacted Curriculum (SEC) (Wisconsin Center for Education Research, 2010) from which the number of students in the participating classes and rates of English proficiency were extracted. The ethnic compositions were obtained

from the combined data provided by the district because this was the only source of information for this category.

Table 1: Class demographics

Class-room*	Number of students (s)	% female (d)	% White (d)	% Black (d)	% Latino(a) (d)	% Asian / Pacific Islander (d)	% Indian / Alaska Native (d)	% Multi-ethnic (d)	% ELL (S)
Abigail	≤ 15	70	38	0	8	31	23	0	≤ 10
Cathy	≥ 26	49	70	3	11	16	0	0	≤ 10
Felicia	16 – 25	49	54	12	0	18	11	6	11-30
Joe	≥ 26	37	66	9	11	11	3	0	≤ 10
Margot	≥ 26	54	50	4	8	38	0	0	≤ 10
Penny	≥ 26	50	59	3	24	12	3	0	≤ 10
Sonia	16 – 25	82	57	0	8	31	23	0	≤ 10
Valerie	No data	53	23	17	20	33	2	5	11-30

*The names assigned to the classroom teachers are pseudonyms, here and throughout the study.

(s) - Data were obtained from SEC.

(d) - Data were provided by the districts.

Teacher participants. Teachers participating in the professional development workshop described above agreed to submit TIPs and SWSs for the upcoming year as part of the workshop evaluation. As also noted above, the teacher participants were asked for their consent to use the TIPs and SWSs for research purposes. Those who agreed and submitted materials for high school SI units were included in this study.

The participating teachers provided some information about their own demographics, education and teaching experience through the SEC. These data are summarized in table 2.

Table 2: Demographics, experience and education of participating teachers from SEC.

Class-room	Gender	Ethnicity	Teaching experience (years)	Years teaching at current school	Highest degree	Major field of study for bachelor's degree	Major field for highest degree
Abigail	F	White	6-8	3-5	Master's	Science	Sci Ed
Cathy	F	White	6-8	<1	Master's	Science	Other
Felicia	F	White	1-2	1-2	Multiple Master's	Science	Sci Ed
Joe	M	White	<1	<1	Multiple Master's	Science	Sci Ed
Margot	F	White	6-8	6-8	Multiple Master's	Science	Sci Ed and Sci
Penny	F	White	<1	<1	Multiple Master's	Science	Sci Ed
Sonia	F	White	6-8	6-8	Master's	Science	Sci Ed
Valerie	F	Mixed	>15	12-15	Multiple Master's	Science	Sci

Study Design

In this mixed-methods, multiple case study, the teachers' instructional practices in science inquiry units were assessed using the TIP rubric, a qualitative coding scheme focused on practices related to scientific explanation and a holistic case description. Teaching methods, grade levels, content areas and inquiry topics varied across the eight classrooms studied without any influence by the researchers (Table 3).

Table 3 – Unit information for participating classrooms

Class-room	Grade Level	Content Area	Unit topic	SI Activity
Abigail	11	Biology (IB*)	Cell theory	Osmosis Lab- concentration of solute in a stick of celery
Cathy	10	Chemistry**	Ecology	The effect of changing conditions on the sprouting and growth of turnips
Felicia	9	Physical Science	Chemical Reactions and Reaction Rates	Alka-seltzer / water reaction time with various conditions
Joe	10	Chemistry	Chemical Reactions and Reaction Rates	Alka-seltzer / water reaction time with various conditions
Margot	11	Physics	2D motion	Predict motion of a ball rolling off of a countertop
Penny	9	Life Science	Ecology	The effect of changing conditions on the sprouting and growth of turnips
Sonia	12	Biology (IB*)	Anatomy and physiology	Relationship between changes in the cardiovascular system relative to changes in body position or stimuli.
Valerie	10 & 11	Biology	Relationships among living things	The effects of pH on macronutrient uptake in plants.

*IB stand for “International Baccalaureate”

**The teacher reported her content area as chemistry, though the unit submitted was more consistent with a life sciences focus.

SWS sampling method. Onwuegbuzie and Leech (2007) recommend that in a parallel sampling design such as ours, at least three members of each subgroup should be sampled. We selected as many work samples as was practical to increase the representativeness of our data. To this end, eight student work samples from each classroom were chosen. There was no demographic information available on individual students, so a simple random sampling scheme was used. If a work sample appeared incomplete, then an alternative SWS was chosen at random for analysis.

The portion of each work sample where the explanation would be expected (typically labeled “Discussion”) was analyzed quantitatively using an explanation rubric

and mean scores were calculated for each rubric category and for the sum of the category score. The characteristics of the explanations were also analyzed qualitatively using a coding scheme, and frequencies of various characteristics were tabulated.

Instruments

The instruments used for this study were the TIPs, the qualitative codes for assessing the TIPs and the student work samples.

TIP. The TIP instrument for measuring teacher practice has three parts: the binder with the instructions and prompts for documenting the units' knowledge-skills-experience (KSE) goals, assessment practices and pedagogical reflections; the teachers' documentation of their practices therein, and the rubric used to assess those practices. The TIP rubric is based on best practices in science inquiry instruction identified from the literature: student-centered classroom activity, emphasis on activities combining higher-order thinking skills with content learning, and assessment for learning (Saxton & Rigelman, Unpublished). The TIP rubric is shown in Appendix A.

The qualitative codes for TIP assessment included teacher practices identified by McNeill and Krajcik (2008), along with some additional codes added for this study based on trends in our data set. The TIP qualitative coding scheme is in Appendix B.

SWS. The student work sample instruments were the explanation rubric and the qualitative codes for assessing the explanations. The explanation rubric utilized the

same categories as Ruiz-Primo et al. (2010) (described below); however, the parts of the rubric specific to the density inquiry for which it was developed were generalized so that it was applicable to the wide range of content across the cases (Table 3). For this study, a score of at least 3 on each aspect of the explanation rubric was considered 'adequate' for high school science students. The Next Generation Science Standards (Achieve, Inc., 2013) state that high school students should be able to "make and defend a claim about the natural world that ... reflects scientific knowledge, and student-generated evidence," "make quantitative and qualitative claims regarding the relationships between dependent and independent variables," and "apply scientific reasoning, theory and models to link evidence to claims to assess the extent to which the reasoning and data support the explanation or conclusion." The '3' level of each aspect of the explanation rubric is consistent with adequate performance on these aspects for students in grades 9-12. See Appendix C for the explanation rubric used in this study.

The qualitative coding scheme for the student explanations was adapted from Ruiz-Primo et al. (2010), Kuhn and Reiser (2005) and some additional codes developed for this study based on trends in our data set. The Ruiz-Primo et al. (2010) codes characterize the claims, evidence and reasoning in the explanations in more detail than is captured by the quantitative scores. The Kuhn and Reiser (2005) codes characterize the degree to which the students attend to the defense of their claims in their explanations. The additional codes characterized some trends of interest to the

researchers in this study. See Tables 7-10 below for the qualitative explanation codes and Appendix D for the code definitions.

Quantitative TIP analysis. As noted above, the TIP instrument developed by Saxton and Rigelman (unpublished) includes a TIP rubric used in this study for the quantitative TIP assessment. The TIP is based on best practices for inquiry-based science and mathematics instruction supported in the research literature. The practices identified are: a shift from teacher-centered instruction to student-centered instruction, in which the teacher acts more as a guide to help students construct their own knowledge rather than as a transmitter of knowledge from a position of authority; an integration of higher-order cognitive skill practice with science and mathematics content knowledge; and assessment processes that are intended to gather data to improve teaching and learning (as opposed to measuring learning solely for the purposes of grading). The TIP rubric developed by Saxton and Rigelman (unpublished) measures to what degree the teaching practices documented in the TIP align with those which support successful inquiry-based learning.

Qualitative TIP analysis. The qualitative codes for assessing the TIP were focused largely on those practices identified by McNeill and Krajcik (2008) to support scientific explanation. As already noted, the practices analyzed by the researchers were: defining the three aspects of explanation, modeling each aspect, stating an explicit rationale for explanation and relating scientific explanation to “everyday”

argumentation. McNeill and Krajcik (2008) rated the instructor's level of achievement on each of these goals by reviewing coding videotaped lessons focused on explanation construction, and investigated students' achievement on a performance-related posttest. The current study rated each TIP for evidence in these four categories.

SWS quantitative and qualitative. Ruiz-Primo et al. (2010) developed a method for assessing scientific explanations in middle-school students' science notebooks for an inquiry unit on density. If there was an explanation present, the rater went on to assess whether it contained a claim, evidence, and reasoning to show how the evidence supported the claim, yielding a score from zero to three (0 for no explanation, 3 if an explanation contained all three components). The rater then scored for claim focus, or the degree to which the claim addressed the important issues in the inquiry. The researchers identified five major concepts important to the density inquiry, so the score ranged from zero to five.

After assessing the claim, Ruiz-Primo et al. (2010) address the evidence quality cited in the explanation in terms of its type, nature and sufficiency. They coded evidence type as to whether the evidence was anecdotal, had concrete examples or was based on the student's investigation. The "nature" of the evidence captured whether the evidence was presented as isolated examples or patterns of data. The sufficiency score rated whether there was enough evidence provided to support the claim.

Finally, the researchers evaluated the quality of the reasoning evident in the explanations. They considered alignment, or whether the evidence was connected to

the claim or whether they were presented as unconnected parts. They also scored the “quality of the link,” or how strongly the language in the explanation established the connection between the claim and the evidence. The rubric included three extended qualities to further characterize the students’ explanations: consideration of the quality of the evidence or reasoning, alternative explanations for the data or observations and broader implications of the findings. The explanation rubric and most of the qualitative explanation codes in this study were based on the study by Ruiz-Primo et al. (2010) (Appendix D). However, the explanation rubric for this study was generalized to apply to the range of contents across the classes.

As noted above, Kuhn and Reiser (2005) analyzed written explanations in a middle school science classroom and found that students had difficulty attending to the “defense criterion” of a successful scientific explanation. They identified two aspects of explanations that bolstered the quality of an explanation defense: differentiation between evidence and inference, and overtly persuasive statements. They noted five techniques for differentiating evidence from inference. These five techniques and overtly persuasive statements were included in the qualitative coding scheme for this study (Appendix D).

Reliability and validity of each instrument.

The TIP for investigating teacher practice is currently being evaluated for reliability and validity. To determine the inter-rater reliability of the TIP rubric for this

study, one of the rubric's authors trained three members of the research group on interpretation of the rubric categories. The four raters then practiced independently with other TIPs not included in this study, and met to discuss scoring and interpretation. After several practice sessions, three raters worked independently to score the TIPs for this study. We calculated the intraclass correlation coefficients in SPSS using the two-way random, single measure and absolute agreement options (Landers, 2011) for each of the TIP categories and the composite (the sum of the three category scores). We did not achieve a 0.7 intraclass correlation coefficient on any of the measures, so we resolved our scoring discrepancies by discussion and used the agreed-upon scores in our analysis. The consensus TIP scores were used in all data analysis for this study.

The SWS instrument is based on the study by Ruiz-Primo et al. (2010). In that study, the authors based their explanation construct on Toulmin's (1958) argumentation structure, a construct adapted widely in the education research field. The authors found positive correlations with diverse assessments, bolstering the validity of the instrument. The inter-rater reliability of the explanation rubric in the current study was determined in a similar manner as that of the TIP rubric described above. Four members of the research group met for training and practice on student work samples not included in this analysis. During the discussions, the explanation rubric was refined and clarified to increase its reliability. The intraclass correlation coefficients for the three explanation rubric categories and the sum were calculated as described above. On the final version of the explanation rubric, the intraclass correlation coefficients

were 0.325 for claim, 0.710 for evidence, 0.964 for reasoning, and 0.893 for composite score (the sum of all three categories). Claim was the only category in which the cutoff score of 0.700 was not achieved. The reasoning score reliability was likely inflated due to the considerable number of explanations in the practice set which contained no reasoning at all and scored zero. Thus, the high degree of reliability for reasoning does not guarantee that the measure reliably distinguishes explanations with differing qualities of reasoning present. It is likely that with continued practice we would have achieved adequate inter-rater reliability in all categories, but due to practical considerations the author proceeded with assessment of the selected SWSs, with this reliability weakness noted.

Procedure

TIPs and student work samples were collected for eight high school science inquiry units implemented in the fall of the 2010 school year in the Portland, Oregon metropolitan area. The units were in three content areas (biology/life science, chemistry/physical science and physics) and covered various topics (Table 3). The inquiry units were implemented by the students' science teachers. The TIPs were evaluated with the TIP rubric, with scores determined by research group agreement, to determine the extent to which documented instruction reflected best practices in inquiry classrooms from the research literature. A random selection of eight student work samples was chosen from each of the eight units and the quality of the scientific

explanations therein was scored with the explanation rubric by a single author, the author of this study. The qualitative characteristics of the TIPs and SWSs were coded by the author alone, and either analyzed by frequency (SWS) or tabulated (TIPs). The case descriptions were developed from the group's TIP review notes and teacher and class demographics.

The TIP rubric generated scores in three categories of teacher practice in a science inquiry unit, as well as a composite score for each class. The explanation rubric generated scores in three aspects of science explanation, as well as composite scores (the sum of the three category scores). The explanation results were averaged across the eight work samples in each class. The results were organized into 2x2 contingency tables, which sorted the results for each class into high and low ratings for each rubric category and for the composite scores. The qualitative and quantitative results and case descriptions were used to characterize the instructional strategies to support explanation, the explanations the students produced, and to determine whether any of the instructional strategies were associated with student outcomes in scientific explanation skills.

Results

Case Descriptions

For goal 1, the case descriptions were developed from TIP review to develop a holistic view of the classroom context, unit implementation and pedagogical strategies. These case descriptions are presented below.

Abigail

Abigail's class was an 11th-grade International Baccalaureate (IB) biology class. The teacher was a white female with 6-8 years of teaching experience, with 3-5 of those years at the school where she implemented the unit in this study. She holds a bachelor's degree in a science field, and master's degree in science education. She was certified to teach middle school and secondary science (Table 2).

The class had 15 or fewer students in it, with 10% or fewer classified as having limited English proficiency. Its gender composition was 69% female and 39% male. Ethnically, 38% of the students identified as white, 8% as Hispanic, 31% as Asian/Pacific Islander, and 23% as Indian/Alaskan Native (Table 1).

The teacher's title for the unit in this study was *IB Biology: Cells*. The content was primarily focused on cell theory including cell structure, function in unicellular organisms such as *E. coli*, and cellular processes such as mitosis and passive transport. This was a five-week unit, with five days devoted to hands-on exploration activities and inquiry related to different content, i.e. one day per activity on unicellular organisms, a

“cell lab” microscope activity, osmosis, cells surface area and mitosis. The students were allowed one day outside of regular class to complete the activity for the SWS.

On the other ten days of class, the lessons were devoted mostly to content-focused instruction in the form of handouts and worksheets on topics like cell theory, characteristics of life, type of living things, cell sizes and plasma membranes. There was also evidence in the TIP of some class discussions, group activities like acting out mitosis and test review. There was also some evidence of skill practice, such as modeling diffusion and osmosis, and inquiry skills like observation and controlling variables.

The topic of the students’ SWSs was osmosis, which they investigated by determining the concentration of solutes in the cells of a celery stick. The students placed celery sticks in salt water of various concentrations, compared the celery sticks’ weights before and after to determine how much water the celery sticks gained or lost, and used the data to determine the salt concentration at which the celery’s mass did not change (or would not have changed). This would be the concentration at which the celery cells’ solute concentration and the surrounding salt water’s concentration were equal.

The teacher of this class did not respond to the section of the TIP asking for percent of time spent on pedagogical strategies.

Cathy

Cathy’s class was a 10th-grade chemistry class. Cathy was a white female with 6-8 years of teaching experience, with less than one of those years at the school where

she implemented the unit in this study. She holds a bachelor's degree in a science field, and master's degree in a field other than science or science education. She was certified to teach middle school and secondary science. (Table 2)

The class had 26 or more students in it, with 10% or fewer classified as having limited English proficiency. Its gender composition was 49% female and 51% male. Ethnically, 70% of the students identified as white, 3% as black, 11% as Hispanic and 16% as Asian/Pacific Islander (Table 1).

The topic for this 17-day unit in this study was ecology. The knowledge learning targets related to energy transfer through ecosystems, ecosystem relationships, matter cycles and the impacts of human activities on ecosystems. The teacher listed SI skills as the targeted skills for the unit. The experiences in the unit were a predator-prey activity, a station activity culminating in students constructing a nitrogen cycle, a plant growth inquiry and a carbon footprint extra credit activity.

There was evidence in the TIP that the teacher used some "warm-up" questions to stimulate discussions about matter conservation, such as where water in wet clothes goes as the clothes dry, where the matter in a growing tree comes from, and whether the mass of a jar of seeds would change if the seeds sprouted. There were six sets of PowerPoint slides in the TIP, suggesting that lecture was used on six of the unit days. The topics of the lectures were ecology, ecological interactions, populations, cycles of matter, human impacts and declining polar ice.

To target SI skills, there was a presentation sheet for the SWS that instructed students that the discussion section should include a “scientific explanation” with “at least four reasons for your results based on science” (emphasis in TIP). The sheet also included a note stating, “You may need to do additional background research to understand why you got the results you did,” cuing students that the reasoning in the explanation should include some science content.

The topic of the inquiry for the SWSs was plant growth; specifically, the effects of various substances such as lemon juice, baking soda, hydrogen peroxide and vinegar on the sprouting and subsequent growth of turnip seeds in a petri dish. The students were allowed to choose which substance they wanted to test against a control group without the substance. They spent 6 days (or parts of each day) of the unit working on their SWSs.

In the Pedagogical Reflection section of the TIP, the teacher reported that the class spent 40% of the time on group work, 40% on science inquiry, 15% on lecture and 5% on class discussion.

Felicia

Felicia’s class was a 9th-grade physical science class. Felicia was a white female with 1-2 years of teaching experience, with 1-2 of those years at the school where she implemented the unit in this study. She holds a bachelor’s degree in a science field, and multiple master’s degrees. One of her master’s degrees is in science education. She

had one emergency, provisional or temporary certification and a secondary science certification. (Table 2)

The class had 16 to 25 students in it, with 11-30% classified as having limited English proficiency. Based on the number of students reported for this teacher and subject by the district, it is likely that the gender and ethnic composition provided is for all of her physical science classes combined. The gender composition was 49% female and 51% male. Ethnically, 54% of the students identified as white, 12% as black, 18% as Asian/Pacific Islander, 11% as Indian/Alaska Native and 5% as multi-ethnic (Table 1).

The content topic for the five-week unit was chemical reactions. For the first three weeks, the content instruction focused on balancing chemical reaction equations. There was a lab activity entitled "Introduction to Reactions" that asked the students to observe chemical reactions and describe the products. In another lab exercise entitled "Conservation of Mass Lab" the instructions for the conclusion section explicitly directs students to state whether or not their hypotheses were supported, to use data examples to support their claims, and to explain how the results related to the law of conservation of mass. The class started talking about reaction rates in the fourth week, using a demonstration and an inquiry into the reaction rate of magnesium ribbon with hydrochloric acid. The Mg-HCl reaction lab template was highly scaffolded. In the TIP, Felicia commented that it did not work well because the acid was not strong enough; however, she instructed students to make a "judgment call" about the reaction rate based on the number of bubbles being formed. She noted, "They have to support their

claims with observations.” The objective, materials and procedures sections were written for the students, and the data table was already constructed. The instructions for the “Conclusion” section said,

Write a conclusion for your lab. Begin by stating your hypothesis. Next decide if your hypothesis was supported by the results of your experiment. Use data to support your conclusion. Explain why you think you got these results (see section 5.4 for ideas). Finally, describe one thing that could be improved upon in your procedure.

The students began planning their SI at the end of the fourth week, and began collecting data at the beginning of the fifth week. The SWSs investigated the effects of various conditions on the reaction rate of Alka-Seltzer with water. The SWSs in this study all varied the temperature of the water and/or the surface area of the Alka-Seltzer (by crushing it, grinding it or leaving it as a whole tablet). The template for the SWS had detailed instructions on what to include, along with learning progressions in each section. The section describing what should be in the conclusion section included the instructions:

Report the results (summarize what happened). Use ACTUAL NUMBERS from your data to support your conclusion. Identify any patterns in the data and/or graphs. Try to explain why things behaved as they did (emphasis in original).

In the TIP the teacher noted that she was pleased with the overall quality of the SWSs, and she thought the unit “adequately prepared students for this lab.” The teacher’s pedagogical reflection indicated that she spent 5% of the class time on direct instruction, 60% on group work, 20% on independent work (practice and classwork) and 5% on homework.

Joe

Joe's class was a 10th-grade chemistry class. Joe was a white male with less than one year of teaching experience. He holds a bachelor's degree in a science field and multiple master's degrees. One of his master's degrees is in science education. He had a secondary science certification. (Table 2)

The class had 26 or more students in it, with 10% or fewer classified as having limited English proficiency. Its gender composition was 49% female and 51% male. Ethnically, 54% of the students identified as white, 12% as black, 18% as Asian/Pacific Islander, 11% as Indian/Alaska Native and 5% as multi-ethnic (Table 1).

The content topic for the five-week unit was chemical reactions, including stoichiometry, limiting reagents and factors affecting reaction rates. The unit began with instruction in SI skills, including a cookie lab in which the teacher modeled knowledge claims with "good" and "bad" examples. The example of a good knowledge claim included an assertion with experimental evidence to support it. There was no differentiation between claim and evidence in this model, and no reasoning. In the lab exercise about limiting reagents, the students received a handout with the procedures and data table prepared by the teacher and list of questions at the end that did not require any scientific explanation. Students conducted peer assessments of each other's knowledge claims, results and conclusion sections in the cookie lab, and the conclusion sections from the limiting reagent labs.

To prepare students for their SI activity, there was evidence of a class discussion around knowledge claims and a handout for students to follow in preparing their SWS which defined a knowledge claim as “a generalizable assertion explaining a scientific phenomenon that is based upon supporting experimental evidence.” The students had one day of lecture and demonstrations regarding kinetic molecular theory (KMT) before beginning their work samples.

The SWSs investigated the effects of various conditions on the reaction rate of Alka-Seltzer with water. The SWSs in this study all varied the temperature of the water, the degree of agitation and/or the surface area of the Alka-Seltzer (by crushing it, grinding it or leaving it as a whole tablet). After the SI, the students saw a 30-minute video on KMT.

In his pedagogical reflection, Joe wrote that the demonstrations with pair-share discussion of scientific explanations produced the largest learning gains for his students. The teacher estimated that he spent 10% of the class time on lecture, 30% on group discussions, 20% on group work, 10% on group data collection, 20% on demonstrations, 5% on peer-peer reviews and 5% on videos.

Margot

Margot’s class was an 11th-grade physics class. Margot was a white female with 6-8 years of teaching experience, with 6-8 of those years at the school where she implemented the unit in this study. She holds a bachelor’s degree in a science field, and

multiple master's degrees. Two of her master's degrees are in science and science education. She had a secondary science certification. (Table 2)

The class had 26 or more students in it, with 10% or fewer classified as having limited English proficiency. Its gender composition was 54% female and 46% male. Ethnically, 50% of the students identified as white, 4% as black, 8% as Hispanic and 38% as Asian/Pacific Islander (Table 1).

The content topic for the 23-day unit was two-dimensional motion, including calculations with Newton's laws applied to objects under the force of gravity. According to the unit plan, 14 of those days were spent in direct instruction with lecture, video and or watching online simulations; one day was set aside for independent problem-solving practice and four days were spent on completing practice tests and unit tests. The class spent two days on a graph-matching activity with Vernier motion sensors, and two days on their SI work samples.

The SWSs all involved rolling a ball bearing off of a countertop and measuring the horizontal distance it landed from the end of the counter. The students compared the measured distance with the distance predicted from theoretical calculations.

In her pedagogical reflection, Margot estimated that she spent 50% of the class time on lecture, 25% on independent practice, 12% on group work, 8% on labs and 4% on simulations. This is roughly consistent with the unit plan percentages above, if the teacher characterized the Vernier activity as "group work" and combined the practice test with the independent problem-solving practice as "independent practice."

Penny

Penny's class was a 9th-grade life science class. Penny was a white female with <1 year of teaching experience. She holds a bachelor's degree in a science field, and multiple master's degrees. At least one of her master's degrees is in science education. She had a middle school and secondary science certification. (Table 2)

The class had 26 or more students in it, with 10% or fewer classified as having limited English proficiency. Its gender composition was 50% female and 50% male. Ethnically, 59% of the students identified as white, 3% as black, 24% as Hispanic, 12% as Asian/Pacific Islander and 3% as Indian/Alaska Native (percentages add to 101 due to rounding effects) (Table 1).

The content topic for the unit was ecology, including predator-prey relationships, food webs, matter cycles and human influences. The activities included "warm ups" to stimulate class discussion, a predator-prey game, a food-web worksheet, a "traveling nitrogen" station activity in which the students had to construct their own nitrogen cycle, and an exercise in which students calculated their carbon footprints.

The topic of the science inquiry SWSs was the effect of various substances such as ammonia, sugar and hydrogen peroxide on the spouting and growth of turnip seeds in a petri dish. In a PowerPoint presentation entitled *Inquiry Lab Part III and IV:*

Collecting Data/ Analyzing and Interpreting Results, a slide includes the instructions:

"Give a scientific explanation for why you got the results you did – try to give at least 4 possible reasons for your results ***based on science and not just your opinion**. Did your results support your hypothesis? Say here, 'The evidence I

found in my experiment did (or did not) support my hypothesis because _____." (emphasis in original)

In the pedagogical reflection, Penny estimated that she spent 40% of the class time on direct instruction, 30% on inquiry, 10% on class discussion, 10% on cooperative learning and 10% on pair work.

Sonia

Sonia's class was a 12th-grade IB biology class. Sonia was a white female with 6-8 years of teaching experience, with 6-8 years at the school where she implemented the unit in this study. She holds a bachelor's degree in a science field, and a master's in science education. She had a secondary science certification. (Table 2)

The class had 16-25 students in it, with 10% or fewer classified as having limited English proficiency. Based on the number of students reported for this teacher and subject by the district, it is likely that the gender and ethnic composition provided is for all of her biology classes combined. Its gender composition was 83% female and 17% male. Ethnically, 56% of the students identified as white, 9% as Hispanic and 35% as Asian/Pacific Islander (Table 1).

The content topic for the unit was anatomy and physiology, including digestive, circulatory and respiratory systems. Out of the 28 days of the unit, 11 were spent on lab activities, and 16 were spent doing lecture, taking notes, watching videos and working problems. One of the labs investigated the differences in taste sensitivity among a sample of the student population. The instructions for this lab had a section labeled

“Discussion of Results” with specific questions for students to answer, primarily about data interpretation. However, some of the questions did prompt students for answers that would allow students to practice some aspects of scientific explanation, such as “Were your thresholds for all three flavors the same? Why or why not?” This question prompts for a claim and some reasoning, but does not ask students to include evidence in the answer. Two of the four questions ask about taste variation with age, though it is unclear how students would substantiate any claims made in this regard based on data from the experiment, which only appeared to collect taste thresholds on the student and others close to the same age.

There was also a calorimetry lab that was mostly procedural, in that the student followed prescribed procedures and had no opportunities to practice explanation. In addition, they completed a “lung lab” activity investigating the different lung volumes (inspiratory volume, tidal volume, etc.) in which the procedures were prescribed. This activity included several questions that required students to apply their learning in various hypothetical situations, but did little to engage students’ explanation skills. The question that came closest was:

Did you notice a difference between the lung capacities of the two people who were tested? How could you account for this?

When answered fully, this question could prompt students to include claim and reasoning, though this was not presented at the beginning of the lab instructions as the question students were trying to answer. Instead, the stated purpose of the activity was

to “determine [the volume of inhaled and exhaled air with each breath] as well as some reserve volumes.”

The topics of the science inquiry SWSs were the effects of various historical factors, physiological factors or stimuli (e.g. gender, type of athletics participation, age) on physiological measurements such as heart rate after exercise and blood pressure. The students chose their own topics within these parameters for their SWSs. Some examples of topics in the study sample were the differences in heart rate changes after physical activity in athletes vs. non-athletes, and the relationship of swimming experience on swimmers’ heart rates after a measured sprint.

There was no evidence in the TIP that Sonia mentioned scientific explanations during this unit, but she did note that the IB Biology students had had a great deal of practice with science inquiry in the past. In her pedagogical reflection, she estimated that the class spent 22% of the time watching PowerPoint presentations, 26% on pair work for labs and poster projects, 23% on videos and class discussion, 25% of inquiry activities, and 4% on cooperative learning for test review.

Valerie

Valerie’s class was a biology class with grade level unspecified, though the work samples analyzed for this study had cover sheets listing the grade levels of the students as 10th and 11th. Valerie was a white female with more than 15 years of teaching experience, with 12-15 years at the school where she implemented the unit in this

study. She holds a bachelor's degree and multiple master's degrees, with at least one of the master's in a science field. She had a secondary science certification. (Table 2)

Valerie did not specify the class size (all the work samples analyzed were from the same class period, according to the SWS cover sheets), but did state that 11-30% of students were classified as having limited English proficiency. Based on the data from the school district, it is likely that the demographics data were from multiple classes combined. Their combined gender composition was 53% female and 47% male. Ethnically, 23% of the students identified as white, 17% as black, 20% as Hispanic, 33% as Asian/Pacific Islander, 2% as Indian/Alaska Native and 5% as multiethnic (Table 1).

Valerie's title for the unit was *Relationships Among Living Things*, though the TIP appeared to combine materials from several units, making it difficult to determine what specific content was associated with the unit in which the SWSs for this study were implemented.

All of the SWSs analyzed for this study were entitled *The Effects of pH on Macronutrient Uptake – Plant Lab*. The experiments involved adding bone meal or blood meal fertilizer to an experimental group of fava bean seeds planted in soil and comparing their growth to that of a control group, and measuring the soil pH and potassium, phosphorous and nitrogen levels with detector strips. Despite the title, it was unclear how the pH levels were controlled in these experiments. It was more properly another dependent variable, while the independent variable was fertilizer vs. no fertilizer.

Because the TIP seemed to have content and activities from several different units submitted together, we were not able to distinguish which activities and pedagogical techniques were germane to the SI unit for this study. Thus, this case was excluded from the TIP analysis.

Teacher Instructional Portfolios (TIPs)

Another aspect of characterizing the instruction in the TIPs for goal 1 was to score them with the TIP rubric and qualitative codes. The results are presented below.

Quantitative Analysis

Table 4 summarizes the TIP scores as evaluated with the TIP rubric.

Table 4: Summary of Teacher Instructional Portfolio Scores by Class

Class ID	Abigail	Cathy	Felicia	Joe	Margot	Penny	Sonia	Valerie*
Classroom roles	3	3	3	3	2	2	2	
Content/Cognitive	2	2	3	3	2	3	3	
Assessment	3	2	3	3	2	2	3	
Sum	8	7	9	9	6	7	8	

* Valerie's TIP was not included in the TIP analysis because the materials submitted in the TIP seemed to be from several different units, making the TIP unscorable.

Table 4 shows that four of the seven TIPs scored a '3' in Classroom Roles, indicating that the instruction was generally well-balanced between student-centered and teacher-centered learning. Three of the seven TIPs scored a '2' in this category, showing that these classrooms were tipped farther towards teacher-centered

instruction. Likewise, four of seven TIPs earned a 3 in the Content/Cognitive category, suggesting a good balance between content knowledge and higher-order cognitive skills in the instruction. In classes that earned ‘2’ in Content/Cognitive, there was evidence in the TIP that either content or cognitive skills were emphasized in the unit at the expense of the other. Four of the seven TIPs also scored ‘3’s’ in Assessment, implying that the teachers mostly employed best practices in assessment such as using pre-assessments, formative and summative assessments that were linked to learning objectives and appropriately challenging; and incorporating self- and peer-assessment to encourage metacognition and students’ ownership of their own learning.

The TIP sum scores reveal that four of the seven TIPs were at ‘8’ or above, showing they earned at least an “adequate” score of ‘3’ in most of the categories. The rest were below ‘8’ indicating an “inadequate” score of ‘2’ or below in at least two of the three domains.

Qualitative Analysis

The qualitative analysis of the TIPs was based on codes adapted from McNeill and Krajcik’s (2008) analysis of teacher practices that support scientific explanation, with some codes added based on our observations. Table 5 shows the qualitative results.

Table 5: Qualitative Analysis of TIPs

Class	Abigail	Cathy	Felicia	Joe	Margot	Penny	Sonia	Valerie **
Defining explanation (1)								
Defined claim*	0	0	5	5	0	0	0	
Defined evidence*	0	0	5	3	0	0	0	

Table 5: Qualitative Analysis of TIPs

Class	Abigail	Cathy	Felicia	Joe	Margot	Penny	Sonia	Valerie **
Defined reasoning*	0	4	3	0	0	4	0	
Making rationale explicit (1)	0	0	0	0	0	0	0	
Modeling explanation (1)								
Modeling claim	0	0	0	2	0	0	0	
Modeling evidence	0	0	0	1	0	0	0	
Modeling reasoning	0	0	0	0	0	0	0	
Connecting science explanation to everyday explanation (1)	0	0	0	0	0	0	0	
Explanation mentioned in TIP	Yes	Yes	Yes	Yes	No	Yes	No	
Degree of scaffolding evident in TIP	Low	High	High	Low	Med	High	Low	

(1) Codes from McNeill and Krajcik (2008)

*For these scores, we did not require that the instruction use the words "claim," "evidence" or "reasoning," just that an equivalent of the component was considered.

** Valerie's TIP was not included in the TIP analysis because the materials submitted in the TIP seemed to be from several different units, making the TIP unscorable.

Four of the teachers in our sample provided evidence that they defined one or more aspects of scientific explanation, and two of them modeled one or more aspects, to some degree. Only Felicia's and Joe's TIPs had evidence that the teachers provided students with definitions of claim and evidence. Cathy's, Felicia's and Joe's TIPs had evidence of guidance to the definition of reasoning. Only Joe's TIP demonstrated modeling of claims and evidence. In the examples of a good and bad knowledge claim, the instructor combined claim with evidence under the heading "knowledge claim," resulting in '2' score on 'Modeling Claim' for "identifying too much" and a '1' score on 'Modeling Evidence' for "does not identify" (Appendix B). There was no evidence in the TIPs that any instructors stated an explicit rationale for scientific explanation nor connected it to everyday examples.

Student Work Samples

Goal 2 of this study was to evaluate the quality and characteristics of the scientific explanations present in the students' (SI) work samples using a mixed methods analysis. The results are presented below.

The student work samples were evaluated with the explanation rubric (Appendix C). Two explanations from work samples in Abigail's class demonstrate how this rubric was applied.

Explanation Scoring Examples

It should be noted that, in this study, the claims were not evaluated for accuracy. Unlike the investigation by Ruiz-Primo et al. (2010), the SI topics in this research covered a range of content areas, and it was beyond the scope of the study to assess all the topics for accuracy. We chose to focus on whether the claims addressed the research question or hypothesis as described in the explanation rubric (Appendix C).

The SI topic for Abigail's unit was osmosis. The students explored this topic by determining the concentration of solutes in the cells of celery stalks. They placed the celery in various concentrations of salt-water solution and measured the weight change before and after the soaking to determine whether water was absorbed by or removed from the celery. The rationale was that plant cells will absorb water by osmosis when the concentration of solutes in the cells is higher than in the surrounding solution. Likewise, the cells will lose water if the surrounding medium has a higher concentration of solutes than the cells do. The salt solution that produces no gain or loss of water

weight in the celery is isotonic (equal in concentration) with the plant cells, thus yielding the desired result. One student's Conclusion section read (note: the grammar and usage errors in the quoted SWS sections were copied directly from the source and not amended):

From the results gathered in this experiment it was concluded that the concentration of the solutes of a piece of celery was 0.1M (10%) because the difference between the before and after measurements of the celery in the 10% salt solution are the closest to the weight difference in the distilled water sample. Celery is made up of 90% water and 10% [sic] These results supports the idea that a piece of celery will contract or expand depending on the salt concentration of the solution it's put in. The results somewhat support the hypothesis. Although the results were not tested by any statistical test there is very little probability that these results happened by chance. There may have been better results obtained if there was a bigger sample size and different salt solutions that ranged between 2% and 30%.

This explanation scored a 2 in claim. The claim clearly stated the student's result for the concentration of solutes in the celery, but it was not supported by the data presented, and demonstrates a misconception about how the celery solute concentration is related to the concentration of the surrounding solution.

The explanation also scored a 2 in the "Evidence" category. While two data points, the weight differences for the 10% solution and the distilled water, were referred to, no data were specifically stated. While the student compared the results of the distilled water and 10% conditions, the other data points were ignored and no pattern was discussed.

The "Reasoning" category score for the explanation was 1. The student's reasoning, that the concentration of solutes in the celery stalk was 10% because the

celery's weight difference in the 10% salt solution was closest to the weight difference in the distilled water, demonstrates a lack of understanding of the process of osmosis. While the explanation did bring in some background knowledge related to the proportion of the celery's weight due to water, it stopped short of specifying any other component of the celery. The statement about expansion and contraction is an observation, but its connection to osmosis is unstated. The explanation also makes some unsupported claims about statistical significance.

This student understood that the purpose of the lab was to determine the concentration of solutes in the celery stalk, but the explanation demonstrated the student's lack of understanding of the underlying mechanism, osmosis, that was the real content learning target of the SI. The student also did not understand the role of patterns in data for supporting a claim.

There were several examples of high-quality explanation in this class. One was this:

The results of the lab indicated that celery has a solute concentration of about 5% ($\pm 1\%$). This corresponds to the hypothesis, which stated that the solute concentration in celery was 5%. The results of the lab shows this because the small change of weight for the 5% solution means that there was little movement of water in and out of celery, meaning that the celery was placed in a near isotonic (equal concentration of solute and water) solution. Whereas for the other solutions, there was either significant rise or fall in the weight. In the distilled water solution, for example, had a final weight of 2.90g, .07 grams heavier than the initial. The 25% solution also had a huge change. The celery's final weight was 2.71, a drop of about .13 grams.

The experiment was useful for examining the process of osmosis. It was useful because it illustrated how isotonic, hypotonic, or hypertonic solutions affect a substance that is placed in it. We see that if a substance was placed in a hypotonic solution, the substance will expand (water is coming into the

substance). It placed in a hypertonic solution, it will shrink (water leaves the substance). And if it is placed in an isotonic solution, the size should stay the same (water enters and leaves equally).

The explanation scored 4s (the highest level) in all three rubric categories. The claim was clearly stated and its relationship to the hypothesis was explicit. As evidence, the student provided two data points and stated the relationships in the data (the 5% solution had a “small change of weight,” the others had “significant rise or fall in the weight”). The explanation connected the evidence to the claim by reasoning with scientific content (“because the small change of weight ... means the celery was placed in a near isotonic ... solution.” The second paragraph further develops the concept of osmosis and how it affects the size of the celery. This explanation demonstrates that the student understood the concept of osmosis and how it was used to answer a scientific question.

Quantitative Analysis

For this study a score of at least 3 on each part of the rubric was considered ‘adequate.’ Table 6 summarizes the student work sample results.

Table 6: Summary of Student Work Sample Explanation Scores by Class*

Class ID	Abigail	Cathy	Felicia	Joe	Margot	Penny	Sonia	Valerie
Claim mean	3	2	3	3	2	3	3	3
Evidence mean	3	2	2	2	3	3	4	3
Reasoning mean	3	2	1	1	3	2	3	2
Composite mean	9	6	7	6	9	8	9	7

*Because of rounding, the composite mean displayed is not necessarily the sum of the displayed component scores.

For the ‘Claim’ category (hereafter referred to as “Claim”) table 6 shows that six of the eight classes’ average Claim scores rounded to a ‘3,’ considered ‘adequate,’

indicating that in general students included a claim that was explicitly stated in the explanation section, related to the research question and supported by the data. Very few students achieved a '4,' which would have indicated that the claim's relevance to the research question was explicitly stated and that the claim was proximate to the evidence and reasoning as part of a cohesive argument. Two of the six classes had average Claim scores closer to a '2,' suggesting that, on average, the students' claims were incomplete, implicit, incorrectly addressed the research question or were not supported by the data.

In the 'Evidence' category ("Evidence"), five classes scored an average of '3' or above when rounded. This shows that most of the students included language describing a pattern when presenting their evidence. In three of the classes the students, on average, either neglected to identify a pattern in the data to adequately support a claim, or used data in a flawed or unscientific manner.

SWS scores fell precipitously in the 'Reasoning' category ("Reasoning"). Only three classes of the eight had average reasoning scores that rounded to 3. Three classes had scores that rounded to 2, and two classes' Reasoning scores were closer to 1 on average. In these classes, there were two or more students in the sample who brought no reasoning nor science content at all into their explanations, thus earning a '0' for that category. Most of the SWSs in the study were below the adequate level in Reasoning.

Half the classes earned a rounded composite score of 8 or above, showing that the explanations earned 'adequate' scores in at least two of the three categories. Half

also fell short of this standard, earning an ‘inadequate’ score of 2 or below in at least two of the three categories.

Across all eight classes, 78% (50 out of 64) of the explanations were “complete,” defined as having a score greater than zero in all three categories; 17% (11 out of 64) had non-zero scores in Claim and Evidence but contained no reasoning; 3% (2 out of 64) had a claim without evidence or reasoning, and 2% (1 out of 64) had no claim.

Qualitative Analysis

The qualitative analysis of the student work samples was based on codes from Ruiz Primo et al. (2010), Kuhn and Reiser (2005), and our own observations. Tables 7 - 10 show the frequencies of elements related to students’ scientific explanations in their work samples.

Table 7: Frequencies of Claim Attributes by Class in Percentages (1)

	Abigail	Cathy	Felicia	Joe	Margot	Penny	Sonia	Valerie
All elements addressed	50	25	63	63	38	63	88	75
Some elements addressed	50	63	38	38	50	13	13	25
Did not address any element	0	0	0	0	0	25	0	0
No claims only evidence	0	13	0	0	13	0	0	0
No explanation	0	0	0	0	0	0	0	0

(1) Codes adapted from Ruiz-Primo et al. (2010).

Table 8: Frequencies of Evidence Attributes by Class in Percentages (1)

	Abigail	Cathy	Felicia	Joe	Margot	Penny	Sonia	Valerie
<u>Type of evidence provided</u>								
Investigation data	100	88	75	75	100	88	100	75
Investigation and artificial data	0	0	0	0	0	13	0	0
Artificial data	0	13	0	0	0	0	0	13
Anecdotal data	0	0	0	0	0	0	0	0
Word "data" only mentioned (2)	0	0	13	25	0	0	0	13
No evidence	0	0	13	0	0	0	0	0
<u>Nature of evidence provided</u>								
Data pattern only	13	13	25	38	13	0	38	38
Data pattern and examples	50	75	50	25	75	63	63	50
Data examples only	38	13	0	25	13	38	0	0
<u>Sufficiency of evidence</u>								
Sufficient	75	63	63	63	88	100	100	75
Insufficient	25	38	25	38	13	0	0	25

(1) Codes adapted from Ruiz-Primo et al. (2010).

(2) When students used the words "data" or "graph" or "table" or referred to data without instead of presenting or describing data, this category was coded.

Table 9: Frequencies of Reasoning Attributes by Class in Percentages (1)

	Abigail	Cathy	Felicia	Joe	Margot	Penny	Sonia	Valerie
<u>Alignment of claim and evidence</u>								
Complete	50	13	38	13	50	13	88	13
Partial	38	50	0	38	38	75	13	75
No	13	25	0	0	0	13	0	0
Not applicable (2)	0	13	63	50	13	0	0	13
<u>Type of link - connection of evidence to claim</u>								
Elaborated connection	63	0	25	13	63	13	88	0
Simple connection	25	75	13	25	0	50	13	38
No connection	0	13	0	13	13	38	0	50

(1) Codes adapted from Ruiz-Primo et al. (2010)

(2) No reasoning present

Table 10: Frequencies of Other Explanation Attributes by Class in Percentages (1)

	Abigail	Cathy	Felicia	Joe	Margot	Penny	Sonia	Valerie
<u>Differentiation between claim and evidence (2)</u>								
Referencing evidence source directly	63	88	50	50	75	100	88	63
Referencing evidence source vaguely	25	0	25	50	13	0	13	38
Bounding evidence by time or context	0	0	0	0	0	0	0	0
Attributing confidence	0	0	13	0	0	0	13	0
'I know' for evidence vs. 'I think' for inference	13	50	0	0	0	13	0	0
<u>Overtly persuasive statements (2)</u>								
Asserting accuracy	50	0	75	88	13	50	75	38
Presenting counterargument	0	13	0	0	0	0	0	0
Consideration of quality of evidence or reasoning (1)	100	75	63	100	88	75	100	100
Consideration of alternative explanations (1)	13	13	0	0	0	13	50	13
Consideration of broader implications (1)	0	63	0	63	88	63	50	1
<u>Format (3)</u>								
Provided template used	0	0	50	0	0	0	0	0
Uniformity of format suggests template followed	0	88	0	0	75	100	63	100
No template evident	100	13	50	100	25	0	25	0
References tables or graphs in explanation section (3)	38	0	25	25	25	25	50	13

(1) Codes adapted from Ruiz-Primo et al. (2010)

(2) Codes from Kuhn and Reiser (2005)

(3) Added upon review of SWS

Of note in the qualitative results, only two students in the study did not include a claim as part of their explanations (Table 7). Felicia, Joe and Valerie had one or more students leave supporting data out of the explanation completely (Table 8). Felicia's and Joe's classes also had the lowest frequency of students in the sample who referenced their sources of evidence directly in their explanations (Table 10).

A majority of the samples in all the classes provided sufficient evidence to support their claims, meaning that their data were collected in their experiments and either described as a pattern or enough examples were provided to establish a pattern (Table 8). In two of the classes, Felicia's and Joe's, half or more of the students provided no reasoning linking the evidence to the claim and incorporating science content (Table 9). In all of the classes except Sonia's, no more than three of the eight students studied referenced their data tables or graphs in their explanation section (Table 10).

Overall, 94% of the explanations (61 out of 64) exhibited some differentiation of data and inference, primarily in the form of citing data examples or patterns. Half of the explanations (32 out of 64) contained overtly persuasive statements.

The classes with the highest frequencies of template use (as determined by how strongly the uniformity of SWS structure across samples suggested use of a provided template), Cathy's, Penny's and Valerie's, also had the highest frequencies of students producing explanations with either a simple connection or no connection between claims and evidence (Tables 9 and 10).

SWS and TIP Relationships

The third goal of the study was to examine the relationships between teaching practices and SWS explanation quality. Contingency tables were constructed for this analysis. In the tables below, each class' results were sorted into high and low

categories for each explanation aspect and instructional practice examined by the rubrics.

SWS Categories vs. TIP Composite Scores.

Tables 11-14 demonstrate how the explanation rubric category scores compare with the TIP composite scores.

Table 11: Composite results

TIP \ SWS	High (8-12)	Low (3-7)
High (8-12)	Abigail, Sonia	Felicia, Joe
Low (3-7)	Margot, Penny	Cathy

Table 12: Class results by SWS Claim vs. TIP composite

TIP \ SWS	High (3-4)	Low (1-2)
High (8-12)	Abigail, Felicia, Joe, Sonia	
Low (3-7)	Penny	Cathy, Margot

Table 13: Class results by SWS Evidence vs. TIP composite

TIP \ SWS	High (3-4)	Low (1-2)
High (8-12)	Abigail, Sonia	Felicia, Joe
Low (3-7)	Margot, Penny	Cathy

Table 14: Class results by SWS Reasoning vs. TIP composite

TIP \ SWS	High (3-4)	Low (1-2)
High (8-12)	Abigail, Sonia	Felicia, Joe
Low (3-7)	Margot	Cathy, Penny

For the samples in this study, no predictive pattern emerged between composite TIP scores and SWS scores (Table 11). However, high TIP scores were associated with high Claim scores and low TIP scores aligned with low Claim scores for all the classes but Penny's (Table 12). There was no pattern between composite TIP scores and explanation Evidence or Reasoning scores (Tables 13 and 14).

Abigail's and Sonia's classes scored high in all categories on the explanation rubric and on TIP composite. Felicia and Joe had high composite TIP scores but their students were low in all aspects of SWS explanations except Claim. Cathy's class scored

low on all aspects of SWS explanation and low on the TIP composite. Margot’s class scored high on composite explanation, Evidence and Reasoning but low on Claim, with a low TIP score. Penny had a low TIP composite, while her class scored high on explanation composite, Claim and Evidence but low on Reasoning.

TIP Categories vs. SWS composite scores

Tables 15-17 examine the association between the individual teacher practices scored with the TIP rubric and the composite SWS scores.

Table 15: Class results by TIP Classroom Roles vs. SWS composite

TIP \ SWS	High (8-12)	Low (3-7)
High (3-4)	Abigail	Cathy, Felicia, Joe
Low (1-2)	Margot, Sonia, Penny	

Table 16: Class results by TIP Content/Cognitive vs. SWS composite

TIP \ SWS	High (8-12)	Low (3-7)
High (3-4)	Sonia, Penny	Felicia, Joe,
Low (1-2)	Abigail, Margot	Cathy

Table 17: Class results by TIP Assessment vs. SWS composite

SWS \ TIP	High (8-12)	Low (3-7)
High (3-4)	Abigail, Sonia	Felicia, Joe
Low (1-2)	Margot, Penny	Cathy

Table 18: Direct instruction, Classroom Roles and SWS composites

Classroom	Abigail	Cathy	Felicia	Joe	Margot	Penny	Sonia
% Direct instruction / lecture	High	15	5	10	50	40	45
TIP Classroom Roles	High	High	High	High	Low	Low	Low
SWS Reasoning	3	2	1	1	3	2	3
SWS Evidence	3	2	2	2	3	3	4
SWS composite	High	Low	Low	Low	High	High	High

*Teacher 4H did not provide an estimate of time spent on various pedagogical strategies, but the analysis of the TIP suggested a considerable amount of direct instruction in the form of informational packets.

High scores in the Classroom Roles TIP rubric category were associated with low SWS composite scores, and low Classroom Roles scores were associated with high SWS scores (Table 15). On a related note, the classes in which the pedagogical reflection estimated over 40% of class time in teacher-led activities (lecture, direct instruction, videos) had a lower score in Classroom Roles and a higher composite SWS score and mostly higher Evidence and Reasoning scores (Table 18). There were no patterns in this sample between the TIP Content/Cognitive and Assessment categories and the overall SWS scores (Tables 16 and 17).

Discussion

As noted by Ruiz-Primo et al. (2010), students need explicit support in producing scientific explanations, but teachers are not consistently providing them with this support. Our results confirm those findings.

Characteristics of Teacher Instruction

One goal of this study was to examine the teacher practices supporting explanation. The TIP instrument was designed, in part, to assess instruction for supporting students in science and mathematics inquiry-based classrooms. The TIP targets best pedagogical practices across a broad range of contexts including content instruction and units focused on SI or ED. As such, it does not focus exclusively on scientific explanation skills. The TIPs revealed that three of the seven teachers who submitted scorable TIPs were using methods that rated high in at least two of the three categories measured by the TIP rubric.

Classroom Roles.

Four of the seven TIPs earned threes on Classroom Roles (Table 4), suggesting that many science teachers in this study are implementing student-centered activities in their classrooms. For example, in Abigail's class, an 11th-grade IB Biology class, there was a good balance between student-centered instruction and instruction delivered

directly from authoritative sources such as the teacher, textbook and informational packets. Some examples of the latter were worksheet packets and handouts with questions and activities with primarily “right answers” found directly from the reading. Evidence of student-centered instruction included multiple opportunities for students to be active learners and to make connections to prior learning. For example, there was a kinesthetic activity in which the students cooperated to “act out” the process of mitosis by joining and letting go of hands, and an activity that prompted students to investigate properties of membranes using soap bubbles as a model.. Abigail noted there was “little teacher input” for the SI activity on osmosis, making it somewhat student-centered; however, the topic and method of using celery sticks were not chosen by the students. There were other hands-on activities to complement the content such as viewing cells through microscopes and an activity related to why cells don’t grow indefinitely. However, it should be noted that not all hands-on activities are student-centered. The latter activity was highly scripted; students followed a series of prescribed steps designed to lead them to a predetermined conclusion. These confirmatory activities, while giving students opportunities to practice some inquiry skills, do little to engage higher-order thinking or elicit explanations (Katchevich, Hofstein, & Mamlok-Naaman, 2013).

The other TIPs earning a 3 in this category utilized group work and group problem solving often in their classes. The TIPs had little evidence of the teacher

pressing for higher-order thinking while students were working in groups, though this is difficult to measure without direct observation of teacher activity during these lessons.

Three of the seven TIPs earned a 2 in this category (Table 4). The three teachers who scored a 2 in Classroom Roles reported that at least 40% of the class time was used for teacher-led activities such as lecture and videos.

For example, in Margot's class, the majority of the activities were teacher-centered or passive, with students listening to and watching PowerPoint presentations, videos, and computer simulations, with few opportunities to construct meaning. There were days set aside for independent practice in solving problems using teacher-provided algorithms, and for completing practice tests. The SI activity was a confirmatory lab with the topic chosen by the teacher, and students either asserted that their results "came out right" or tried to explain why they did not. As Concannon and Brown (2008) point out, students required to perform confirmation labs "are denied the opportunity to think and reason for themselves." Katchevich, Hofstein and Mamlok-Naaman (2013) demonstrated that classroom discourse and written explanations in confirmation labs are lower in argumentation and explanation quality compared with open-ended inquiries, and do not engage higher-order thinking to the extent that open-ended inquiry activities do. The lab did give students the opportunity to apply their calculation and measurement skills to a physical situation, and the students in Margot's class produced explanations with higher-quality evidence and reasoning than most, but did not make adequate claims, suggesting that they did not have a clear idea of the theory

or model they were testing or the hypothesis they were confirming (see below). To make lab activities more student-centered and to engage higher-order thinking, teachers should have students engage in more open-ended inquiry activities (Concannon & Brown, 2008).

Content/Cognitive skills

In the Content/Cognitive category, four of the teachers provided multiple opportunities for students to learn and practice content and higher-order cognitive skills, frequently in the same activity or lesson (Table 4). They accomplished this by, for example, including discussion questions that prompted for content application during hands-on activities and using group problem solving during classroom demonstrations using a predict-observe-explain format.

For example, in Felicia's class, the teacher incorporated multiple pedagogical strategies to teach the same content. For chemical reaction rates, she used structured observations of chemical reactions, allowing students to practice observation skills while learning about factors affecting reaction rates. The students also had lectures, movies, and practice labs that scaffolded their content knowledge and inquiry skills in preparation for the SI activity. The practice labs offered students opportunities to practice their explanations skills; the topics, procedures, and even data tables were given to the students, along with explicit instructions on how to structure the Conclusion sections. However, as already noted, when topics and procedure are laid

out for students, they are less engaged and less likely to employ the higher-order thinking skills necessary for scientific explanation (Concannon & Brown, 2008).

Three of the teachers earned a 2 in this category (Table 4). As a demonstration, Cathy's class had some activities that allowed students to construct meaning (e.g. the "Traveling Nitrogen" activity) and some "warm-up" questions that promoted application of prior knowledge to explain phenomena such as wet clothes drying on the clothesline. However, the crowning activity of this unit, the SI on turnip growth, had little to do with ecology, the content focus of the unit, and represented a missed opportunity to combine content knowledge with inquiry skills to achieve a deeper understanding of the content.

Assessment for Learning

Four of the TIPs earned an adequate score in Assessment for Learning (Table 4). As an exemplar, Joe used multiple forms of assessment throughout the implementation the unit. He used a lab about cookies to allow students to practice their explanation skills, focusing on constructing adequate claims and providing evidence. The students received feedback on their efforts using peer review and editing. As formative assessments, Joe documented that he used practice problems from the textbook, informal lab write-ups, collective class problem solving with teacher observations, a mid-unit stoichiometry quiz, calling on students and class polling. Overall, the SI summative assessments in these units were connected to the content learning targets, thus providing evidence of students' progress towards these goals.

Penny's class was among the three that earned a 2 in this category (Table 4). There was evidence of formative assessments planned such as warm-up discussions, homework, quizzes, class discussion, group work and video analysis; however, there were no lesson plans provided so it was unclear whether and how any of these were implemented and whether the teacher used the information from them to modify her teaching. One summative assessment was a unit test with several open-ended questions, none of which required students to use higher-order thinking by, for example, applying their knowledge to novel situations. The other summative assessment was the SI. While related to plants, this SI topic had little to do with the unit content of ecology and thus did not provide evidence that students had mastered the unit content. The only opportunities for students to self-assess were the presentation of learning targets at the beginning of the lectures.

The other TIPs that earned 2s in this category were Cathy's and Margot's. Cathy's SI summative assessment had the same weakness as Penny's, in that it was not closely aligned with the content learning targets. Both of these classes had activities that could have been used as informal assessment but provided no evidence that the data from them were used to inform instruction. In all three classes, opportunities for self-assessment were limited to the periodic presentation of learning targets or taking practice tests. Margot provided little evidence of formative assessment; homework was stamped but not reviewed, and she circulated during group work to "help or give hints."

Explicit Support for Explanation.

The codes from McNeill and Krajcik's (2008) study on teacher practices supporting scientific explanation were created for a curriculum provided to the participating teachers which included a focal lesson on defining, modeling and connecting scientific explanation to everyday explanation, and the researchers observed the lessons firsthand. The participating teachers in our study were not enacting a provided curriculum and had not received training from us regarding the emphasis of scientific explanation as the goal of SI nor the elements expected in a scientific explanation. As such, this study documents cases of teachers' current practices around scientific explanation without explicit intervention from researchers or professional development providers. Nevertheless, we found evidence that some teachers used some of the aforementioned techniques for teaching explanation.

Table 5 demonstrates that while four of the seven classes provided some explicit support of scientific explanation, most of this instruction was limited to defining explanation. Only one of the teachers provided definitions, however incomplete, for all three aspects of explanation. One of the teachers modeled claim and evidence, though weakly. None of the teachers provided a rationale for explanation, the one strategy McNeill and Krajcik (2008) identified as the most important for supporting student learning in explanation construction. In fact, McNeill and Krajcik (2008) found that defining explanation without providing an explicit rationale actually had a negative impact on student learning, and modeling explanation made no difference.

Summary

The TIP analysis demonstrates that about half of the high school science teachers in this study adequately employed pedagogical techniques that research suggests support learning in inquiry-based science classrooms. However, there was limited evidence of explicit teaching of explanation, and no evidence that teachers helped students understand the reasons explanations are important, a crucial component in supporting students' explanation skills. While all the students had opportunities to participate in science inquiry, the topics and procedures were often chosen by the teacher.

There were also classrooms in which students were given opportunities to practice content and higher order cognitive skills together in the same activities, but there was room for improvement in structuring the SI activities to meet this criterion. Most of the activities that supported SI skills were highly scripted, limiting opportunities for higher-order thinking (Katchevich et al., 2013). SI topics were chosen by the teachers, and some were only marginally related to the content covered in the unit.

Most of the teachers used assessment practices that included some opportunities for formative assessment, though often it was unclear whether the teachers used the data from the assessment to alter instruction. There were also some classrooms that included peer- and self-evaluation, including some peer evaluation of knowledge claims. On the whole, the teachers in this study employed best practices in SI classrooms, but these practices were not geared towards supporting scientific

explanation. It should be noted that teachers who are interested in supporting inquiry may be overrepresented in this sample, as they were recruited from past professional development workshops offered by Portland State University's Center for Science Education, chose to attend the workshop to design an inquiry unit, and allowed their units to be analyzed for research.

Characteristics of Students' Scientific Explanations

Composite Explanation

Only four of the eight classes received adequate scores on the explanation composite. However, the high school students in this study fared better than the middle school students studied by Ruiz-Primo et al. (2010). In their study, only 18% of the explanations contained claim, evidence and reasoning. In this study, 80% of the students did so. This included all students who got a non-zero score in all of the three categories, regardless of their level of performance. In the Ruiz-Primo et al. (2010) study, about 12% had claims and evidence but not reasoning; in this study, 16% did so. Only 3% of the explanations in this study had claims with no evidence or reasoning, whereas 40% of the students in the Ruiz-Primo et al. (2010) study did so.

There are some possible explanations for these discrepancies. The students in this study were older and presumably had more experience writing up labs and SI projects, so they may have been more aware of the necessity of providing at least some

evidence and reasoning in their explanations. A slightly higher percentage of students in this study (16% vs. 12%) had claim and evidence but not reasoning. This could represent some students whose explanation skills were advanced enough over those in Ruiz-Primo et al. (2010) to include evidence but stopped short of using reasoning. Another difference was that any SWS that lacked an explanation section was eliminated from our analysis, whereas the sample from Ruiz-Primo et al. (2010) included 19% with no explanation. We excluded SWSs with no explanations because we did not analyze science notebooks as the other study did, so there was no way to know whether the part of the SWS containing the explanation was missing or misplaced due to a clerical error. Two classes, Felicia's and Joe's, contained the majority of students with no reasoning, suggesting, as in Ruiz-Primo et al. (2010), that teacher practice is a significant factor affecting students' explanations. Finally, this study assessed explanations in SWSs submitted as summative assessments, rather than explanations in science notebooks. As noted in the introduction, the state of Oregon's standards adopted in 2009 require students to "draw a valid conclusion [and] explain how it is supported by the evidence" in their SI work samples. Because teachers are required to assess these qualities related to scientific explanation, both the teachers' and students' expectations for explanation quality in this study may have been higher than those in the science notebooks.

Reasoning

As in the Ruiz-Primo et al. (2010) study, the rubric category with which students struggled the most was Reasoning (Table 6). Ruiz-Primo et al. (2010) noted, "Regarding

how well students reasoned through from the evidence to claims, nearly most [sic] students were unable to elaborate their reasoning by describing or interpreting the data to make the link more explicit.” As previously noted, 19% of SWSs analyzed for this study included no reasoning or science content at all in their explanations. In Felicia’s class, 63% of the students included no reasoning in their explanations (table 9), even though she explicitly instructed them to do so and gave them opportunities to practice (see case description). This suggests that students may leave reasoning out of explanations even when they know it should be included. Perhaps the students need more practice or did not have sufficient understanding of the underlying scientific theory or content to explain their results.

Defense

The high school students in this study also improved upon the middle school students in Kuhn and Reiser’s (2005) defense criterion. In the latter study, 45% of the explanations lacked a clear distinction between data and inference, a technique the author’s reasoned bolstered the defensibility of the explanation. In the current study, only 6% lacked this differentiation (Table 10). The most likely reason for this is that the students in this study worked exclusively from first-hand data; that is, data collected by themselves during their investigations. As Hug and McNeill (2008) found, students working with first-hand data are more likely to identify their data source when discussing their data, an important technique for differentiating between evidence and inference (Kuhn & Reiser, 2005). In contrast, the students in Kuhn and Reiser’s (2005)

study were arguing from second-hand data; that is, data collected by others and provided to them.

In Kuhn and Reiser's (2005) study, only 29% of the explanations contained overtly persuasive statements, which the authors regard as evidence that the student is attending to the persuasive function of scientific explanations. Fifty percent of the students in this study included overtly persuasive statements (Table 10) such as "...which confirms my assumption..." and "This experiment proved my hypothesis was correct." There was evidence in the TIPs that some of the teachers provided explicit instructions to their students to state whether their hypothesis was supported or not. This cued students to include persuasive statements in their explanations.

Templates

The qualitative results also suggested that templates for structuring student work samples should be constructed with care. The classes with the highest frequencies of template use, Cathy's, Penny's and Valerie's, also had the highest frequencies of students producing explanations with either a simple connection or no connection between claims and evidence (Tables 9 and 10). This hints that templates can prompt students to include all the parts of an explanation while ignoring the connections. Indeed, as Ruiz Primo et al. (2010) reported, notebook prompts with "a high level of guidance" resulted in explanations that "promoted copying from the board and did not allow students to incorporate their own thinking." Those authors also found that

prompts with a very low level of guidance produced explanations with a lack of focus (Ruiz-Primo et al., 2010).

Data Displays

Another interesting observation is that while most students included tables, charts and/or graphs in their Results sections, only 16 students across the entire sample, or 25%, referenced these in their explanations (Table 10). This suggests that while students were aware that this type of data display is expected when reporting scientific results, they were not aware that the purpose of these displays was to expose patterns useful in constructing a scientific argument or explanation. If the display is not to be used to bolster a claim or to raise a further question to pursue (both topics appropriate for the Discussion/Conclusion sections), then what is the rationale for including it?

SWS and TIP comparisons

When we undertook this study, we expected to find that classes with high-quality instruction, as measured by the TIP rubric based on best practices for SI-based instruction from the research literature, would be associated with high explanation scores from the students' work samples. However, as noted above, the TIP was designed to evaluate instructional practices that support a wide range of content and SI/ED skills, and is not focused on explanation skills exclusively. We found that

pedagogical practices emphasized in the TIP supported students in some parts of explanation but not others.

High TIP – High SWS – Abigail and Sonia

Abigail's and Sonia's classes had both high overall TIP scores and high overall SWS scores (Table 11). Both of these classes were part of the International Baccalaureate program. This program requires additional assessments and advanced coursework, so students who are high achievers may have self-selected to enroll in these courses. There was no evidence in the TIPs that the teachers specifically addressed scientific explanation in their instruction for the unit studied in this investigation, (Table 5), but it is possible they received such instruction earlier in the year or in previous years. Sonia's TIP noted that the students had a great deal of SI experience in the past. These two classes scored high in all the categories related to SWS explanation (Tables 6).

Abigail's class. Abigail's class was the smallest class in our study, with fifteen or fewer students enrolled, so there may have been more opportunities compared to other cases for students to get personalized assistance from the instructor. Abigail did not provide a pedagogical reflection with estimates of the amount of time spent on various pedagogical techniques, but the unit calendar revealed that a significant portion of the time was spent on content instruction in the form of lecture and lengthy information packets with reading and questions. Some examples of explanations from Abigail's class were included in the Results section to demonstrate scoring procedures.

Sonia's class. The SI topic in Sonia's class was the most open-ended in the study. Students were given a broad area of inquiry: physiological changes that can be easily measured in the classroom (such as heart rate and blood pressure) under various conditions. As already noted, studies have found that open-ended inquiries tend to engage students in higher-order thinking and produce better explanations (Concannon & Brown, 2008; Katchevich, et al., 2013). An example typical of Sonia's class is a work sample addressing the research question: "Is the change in heart rate in athletes and non-athletes caused by physical activity statistically significant?" The conclusion section included the following passage:

When looking at the data, it can be concluded the difference between the changes in heart rates of the athletes compared to the non-athletes was not statistically significant. This statistical insignificance can be attributed to the varying levels of effort exerted between the athletes and the non-athletes. Over the period of 30 seconds when the subjects were asked to jump rope the athletes tended not only to exert more energy as shown by faster jumping and less stumbling and more general agility. While watching the non-athlete test subjects jumping rope it was acknowledged there was slower jumping even when prompted to go faster and less agility as demonstrated by more tripping, stumbling and starting over. Another possible explanation for the lack of difference between the changes in heartbeats for the two classifications is the age group. Because tall [*sic*] of the subjects were between the ages of 16-18, and all of the students were in high school that could have affected the results. This would have affected the results because of the young age, the subject have more energy than people of older ages. This makes it harder to achieve higher heart rates. In table 2, it can be seen that with an average of an increase of 60.9 beats per minute athletes had a greater increase in heart rate than the non-athletes who on average raised their heart beats by 58.5 beats per minute. These results yielded a p value of .6924, which means that the null hypothesis is accepted, and the slight difference in change in heart rate can be attributed to coincidence caused by random sampling, this would mean that the difference between the changes in hear rates before and after physical activity do not demonstrate a big enough difference to be significant to society.

In this explanation, the claim directly addressed the research question, but the reader had to wait until the end of the paragraph to discover the p-value, the focus of the research question. The evidence presented was strong; the student included both observations and quantitative data. The average values for the two groups were presented and there was a convincing discussion of the meaning of the p-value. The student offered some plausible reasons why no difference was found between the two groups, but the explanation lacked grounding in physiological reasoning as to why a difference might have been expected. A reference to some academic literature or a textbook source to present what others have found and why this experiment had different results would have made it a richer explanation and may not have led the student to conclude that fitness does not make a difference in heart rate recovery in general.

Sonia estimated that the class spent about 22% of its time with PowerPoint presentations, 23% on videos and class discussions around content, and 4% on test review, for a total of 49% of the time on content instruction. The rest of the time was spent on labs, poster projects and SI activities.

Summary. Because strong explanations use scientific knowledge to connect claims and evidence, and these SI topics were closely connected to the content in the units, these students had the tools to reason their way from claims to evidence using scientific knowledge. As McNeill and Krajcik (2007) found,

Students with stronger content understanding constructed stronger explanations.... This suggests that strong content knowledge is important to

appropriately take part in scientific inquiry practices such as accurately constructing scientific explanations. Students may be unable to apply their understanding of a scientific inquiry practice to a context without an understanding of the particular science content.

However, the overall scores in these classes failed to reach the top score. It is notable that while students in both of these classes had opportunities to practice inquiry skills during the unit, these did not include practice in formulating scientific explanations in an SI context.

High TIP – Low SWS – Felicia and Joe

Among the more surprising findings of this study were the results for Felicia's and Joe's classes. These classes with high TIP scores in all the TIP categories (Table 4) had low Evidence and Reasoning scores, leading to low total explanation scores (Tables 6 and 11-14). Both classes' SWS topics involved chemical reaction rates, specifically, the change in the rate Alka Seltzer dissolves in water with different water temperature, agitation and/or surface area (Table 3).

The explanations scored high on Claim, but they were the only classes to have multiple students leave evidence out of their explanations completely. Also, the majority of students with no reasoning in their explanations came from these two classes (Table 9).

Except for Claim, the students in these classes on average produced poor explanations. There was evidence in the TIPs that both teachers provided accurate and complete definitions of Claim (Table 5). Felicia also provided an accurate and complete

definition of evidence and a vague definition of reasoning (Table 5), but this did not result in higher explanation scores in these areas. This supports the finding by McNeill and Krajcik (2008) that instruction in the definitions of the three aspects of scientific explanation do not help students if an explicit rationale for producing explanations is not also offered.

Felicia. The Teacher Reflection section of Felicia's TIP revealed that she spent three days on instruction around reaction rate. The first day included some lecture, group discussions and sharing out. There was also some evidence of demonstrations of reaction rates. It is possible that the students did not learn enough content from these activities to be able to make sense of their experimental results. The second two days of reaction rate instruction involved a highly structured lab activity in which the procedures and data table were provided for the students, along with some explicit instructions for the "Conclusion" section to write what amounted to a scientific explanation. These instructions pointed students to some text ("section 5.4" – presumably in the textbook), possibly cuing them to use some science content in their conclusions. Felicia's comment that "they will have to support their claims with observations" indicates that students had practice supporting claims with data. This lab and the "conservation of mass" lab earlier in the unit should have been good opportunities to scaffold the SI skills, including explanation, required in the SWS.

Despite these multiple opportunities to practice using claim, evidence and reasoning to explain their results, five of the eight SWSs samples for this study included no reasoning in their Discussion sections. Three sample explanation sections were:

The higher the temp. the faster it melts, the lower the temp the slower it melts.

When we timed the reaction the beaker that was boiling desolved the alka-saltzer faster than the beaker with cold water.

According to my data my hypothesis was correct. The Alka seltzers melted way faster when heat was applied and way slower when they were cooled which means heat will affect a chemical reaction faster then the reactant being cooled.

The students who produced these explanations made connections between water temperature and reaction rate. However, the first explanation was a claim with no evidence or reasoning, nor any reference to the purpose of the lab or the substance being tested. The second explanation referred vaguely to evidence by referencing the timing of the reaction, but contained no reasoning as to why the reaction rate was faster in the boiling water. The third attempted a general statement about heat and reaction rate but offered no reasoning as to why this might be so. Also, two of the three explanations above described the reaction as “melting” rather than “dissolving.”

Felicia estimated that she spent 15% of the class time on direct instruction and 60% on group work. However, it is possible that if the teacher was not monitoring group work closely to check for learning, the heavy reliance on this pedagogical technique would fail some students. As Webb (2009) pointed out, “most researchers agree that simply placing students in small groups does not guarantee that learning will take place.”

This was one of the smaller classes in the study, with 16-25 students; however, 11-30% of the students were classified as having limited English proficiency. As Fradd and Lee (1999) have pointed out, students new to our culture may be ...

...unaware of the culturally-based rules for engaging in exploratory activities. Because the indirect nature of exploratory activities makes it difficult for students to acquire participation rules on their own, exploration may limit, rather than enhance, students' opportunities to learn.

Thus, it is possible that the group work favored by this instructor was actually detrimental to some of the English language learners in the class.

Finally, as noted in the case description, Felicia wrote in her TIP that she was pleased with the quality of the SWSs her class produced and thought the unit's activities adequately prepared her students for the SI. This class had 16 - 25 students (Table 1), and five of the samples in this study, composing 20% to 31% of the class, contained no reasoning in their explanations. This may suggest that despite her explicit instructions to include scientific reasoning in their work sample discussion sections, Felicia did not hold reasoning in explanations as an essential component of a SWS, and did not emphasize this skill in her instruction.

Joe. The unit plan in the TIP for Joe's class showed that the class spent only one day with instruction related to kinetic molecular theory (KMT), the accepted theory to explain the phenomena observed in the SI activity. This instruction involved some lecture and some demonstrations related to KMT with an observe-explain instructional technique. Although there were many opportunities for students to practice skills in the unit, the skills were not typically focused on explanation in SI. In the limiting reagent

lab, the students practiced SI skills such as labeling, measuring and data display, but not explanation. The cookie lab had students practice claim and evidence, but not reasoning. The demonstrations of KMT had an observe-explain format that could have given students an opportunity to practice reasoning, but not in the context of SI.

Four out of the eight SWS explanations in this unit contained no reasoning, and two only earned a 1 in Reasoning. Some examples follow:

From the data collected, it can be concluded that the hypothesis was correct. Heat speeds up the reaction time of a chemical reaction. At 70 Degrees Celsius, the highest temperature used in this experiment, the average rate of reaction was 17.64 seconds. At 55 Degrees Celsius, the average rate of reaction was 20.59 seconds. At 40 Degrees Celsius, the average rate of reaction was 25.05 seconds. And at Room Temperature, the average rate of reaction was 116.80 seconds. Because of this data, it can be concluded that heat increases the rate of a reaction.

By looking at the average speed of the reaction it can be concluded that the temperature of a reaction is a cofactor and greatly influences a reaction both positively and negatively. With these findings, it can be inferred that the time necessary to complete the reaction will increase given that the temperature decreases.

Based on the data that was collected, the faster the agitation, the faster the reaction time was. The data also supported the hypothesis.

The first example was a student that made an explicit claim (though it did not specify the reaction studied) and gave adequate evidence in the form of examples with a pattern implied, but attempted no reasoning. The other two examples referred to data collected in the experiment but did not bring any of it into the explanations, and also contained no reasoning.

Overall, Joe estimated that 10% of the class time was spent in direct instruction and 5% on videos. Group activities such as group work, group discussions, group data collecting and peer review took 65% of the class time. Demonstrations took 20%. Joe opined that the biggest learning gains came from pair-share discussions of scientific explanation, and stated that these were “all done with firsthand guidance and feedback from the instructor.”

Despite the statement suggesting that the pair work around scientific explanation was closely monitored by the teacher, it is unclear whether the group work around the science content was also closely supervised. Regardless, only one day of instruction around KMT may have been insufficient to prepare students to explain their SI results using this theory. As noted above, strong content knowledge is important for students to produce adequate explanations (McNeill & Krajcik, 2007). Also, the definitions and modeling that Joe provided around explanation did not include reasoning, and there were no opportunities in the unit for students to practice reasoning skills in the context of SI until the summative SWS. Joe’s was the only TIP that showed modeling of some inquiry skills, including claim and evidence, but this did not result in higher scores. McNeill and Krajcik (2008) also found that modeling components of scientific explanation had no significant effect on student performance.

Summary. In both of these units, the students were expected to develop scientific explanations using KMT to interpret results about factors affecting reaction rates. As noted above, the Next Generation Science Standards call for students to

“apply scientific reasoning, theory and models to link evidence to claims to assess the extent to which the reasoning and data support the explanation or conclusion” (Achieve, Inc., 2013). However, there is evidence that the students may not have received sufficient instruction in this theory to apply it in this manner. Felicia offered three days of instruction on KMT, but relied heavily on group work in her pedagogy. Joe included only one day of instruction on KMT before beginning the SI and also relied heavily on group work throughout the unit. These commonalities suggest that students struggle to produce high quality explanations without a firm grounding in the scientific theory and model which is being tested, a finding supported by McNeill and Krajcik (2007). While group work can be an effective pedagogical tool, teachers need to ensure that students have a firm grip on the science, and for some students, as Fradd and Lee (1999) pointed out, direct instruction can be useful.

Low TIP – High SWS – Margot and Penny

Two of the classes in this study, Margot’s and Penny’s, had low TIP composite scores associated with high student explanation scores. Both of these instructors spent a considerable proportion of their time on lecture or direct instruction according to their pedagogical reflections. Margot’s class’ explanation scores were boosted by high scores on Evidence and Reasoning but fell on Claim, whereas Penny’s students had high Claim and Evidence scores but were low in Reasoning. Although these classes covered very different content areas, a comparison of these cases yields some interesting connections between unit structure and explanation quality.

Margot. Margot received low TIP scores in all categories (Table 4) but her class received high SWS explanation scores in Evidence, Reasoning and overall (Tables 6). Her students, on average, had low explanation scores for Claim (Table 6). The instruction in this unit leaned heavily toward teacher-centered instruction and content practice. Margot estimated that she spent 50% of class time on lecture. The class also spent 25% of its time on independent practice and 4% on simulations. The TIP suggested that the students watched these simulations rather than performing them on their own. Thus, students spent approximately 79% of their time on passive instruction and independent seat work, and only about 20% of the time on group work and labs.

The SI in this unit was a confirmatory experiment on two-dimensional motion (Table 3). The goal was to make a theoretical prediction of where a ball would land when it rolled off a table and compare it with experimental results. Students typically did use their experimental values and mathematical reasoning to test whether the experimental results matched the theoretical results, using formulas and manipulations they had practiced heavily in class.

An example of an explanation that got adequate or higher scores in evidence and reasoning but fell on claim follows. The experimental setup had the ball bearing rolling off a ramp onto a flat surface for 25 cm in one condition and 50 cm for the other before launching off the end of the counter.

The actual ranges for 50 cm and 25 cm were 23.3 and 23.67 cm respectively. The mathematical ranges were 22.7 cm for a launch distance of 50 cm and 24.29 cm for a launch distance of 25 cm. Each mathematical range was within 1 cm of the actual range. It makes sense that the longer distance on the table would lead to

shorter range because more time on the table would increase the friction, slowing the ball bearing down and limiting its range. ...*[Paragraph here about sources of error and suggested improvements in the method.]*... In conclusion, our actual range and our mathematical range were within 1 cm of each other. Therefore, for a given initial velocity and vertical distance h the range R can be predicted.

This explanation included multiple data examples, explicitly stated the relationship between the theoretical and experimental results, and included some science content to explain the results. The reasoning could have been improved with some reference to Newton's laws and gravitation as the theoretical framework from which the results were predicted. However it is difficult to pinpoint the actual claim in this explanation. The reader has to go through the entire section, including a long paragraph about experimental errors and potential improvements, to locate the claim that addresses the research question, which is the last sentence in the section. Until the reader gets to the end, it is difficult to follow what the student is arguing.

The majority of the SWSs examined in this class had incomplete claims, suggesting that they did not clearly understand the "big picture" goal of the experiment: to test a model (Newton's laws applied to gravitation) by making a prediction using it, and to verify (or refute) that prediction with an experiment. Despite this lack of context, students were able to take measurements and carry out much-practiced operations. While students were successful in attending to the requirements of providing evidence and content-based reasoning in their scientific explanations, their lack of strong claims demonstrates that this SI was a missed opportunity for students to engage in model-based theoretical reasoning and learn about the nature of science, aspects the

Framework asserts are some of the most important rationales for engaging in explanation in science education (National Research Council, 2012).

Penny. This life science unit on ecology also had low TIP scores associated with high SWS scores (Table 11). In this case, the TIP scored low on Classroom Roles and Assessment for Learning and high on Content/Cognitive Skills (Table 4). The students' SWS explanations scored high on Claim and Evidence but low on Reasoning. While some of the activities in the unit, such as "traveling nitrogen," offered students opportunities to construct meaning, and the unit was well- balanced between content and cognitive skills instruction, the skills were more related to graph interpretation and application of knowledge to new situations. The graph interpretation practice may have bolstered student's grasp of data patterns as evidence, but there was no opportunity to practice scientific reasoning.

There was evidence in the TIP that Penny included some instruction on scientific explanation, specifically instructing students to include possible reasons for their results "based on science." There was also evidence that students received instruction on the sections of the ODE rubric, which includes some criteria related to explanation; however, there was no indication that claim and evidence were singled out for emphasis, as reasoning was. Nevertheless, these students received adequate scores on Claim and Evidence in the explanation rubric, and their lowest scores on Reasoning (Table 6). Aside from the fact that the TIP contained no instruction related to the rationale for scientific explanation, another possible reason for the poor performance in

Reasoning is the questionable relationship of the SWS topic to the rest of the content in the unit. Though the ecology unit does contain some content related to matter cycles, it may be difficult for students to make the connection between the “big picture” idea of matter cycling through an ecosystem to the demonstration of the effects of various chemicals on seeds sprouted in a petri dish.

An exemplar explanation from this unit follows (the measurements refer to the lengths of the turnip spouts):

Analysis: The average for each solution was pretty equal, for the control (water) I got a 3.59 cm average. For the solution (sugar) I got a 3.1 cm average. I personally wouldn't rely on these averages because the solution was the lowest average but it also had the biggest stem, and the control, which had the highest average, also tied with the shortest stem. The p-value I got was .4962 meaning that there was about a 50 percent chance that the plant would grow more with the solution or the control.

Interpreting Results: I believe the results turned out this way, because of the exact reason I specified earlier, there were no nutrients sugar that help or hinder the growth of turnips. I don't believe that I made any big data changing mistakes in my experiment, but there is the possibility. Other than that I can't find any more inferences about my results, it just didn't do anything. My results show that, not everything has to effect the environment, if something gets in the water supply, some plants or animals might not be affected by it.

The claim required some piecing-together but the discussion around p-values makes it clear that the student is claiming that there was no difference between the results in the experimental and control conditions. The evidence not only included data examples and the relationship between the two groups, but also an observation about the masking of important results when only relying on means. However, the reasoning was limited to an assertion that sugar contains no nutrients that plants need, without any

discussion of what nutrients plants do need nor any resolution of the fact that plants use the process of photosynthesis to make glucose, one of the constituents of a sugar solution.

In the SWSs, many of the students included some connection to ecosystem issues, but the connections were facile. For example, in another work sample in which a student added sugar to the seeds' environment, the analysis section included the statement, "The results of the experiment was [sic] similar to pollution because it caused harm to the turnips rather than helping them. The sugar restricted water flow to plants preventing some water from getting in soaked into the turnips." While it is true that pollution can harm plants, it is unlikely that sugar would be an environmental pollutant, and the effect of sugar on the seed sprouting and plant growth is better explained with some chemistry ideas, as this student was attempting to do. The block-quoted explanation above exhibits this same difficulty with relating the results to ecology, stating that "not everything has to affect the environment."

Among the eight work samples evaluated in this unit, the substances tested were ammonia, sugar, hydrogen peroxide, salt water and vinegar. There did not appear to be any content in the unit that would help students explain why these substances had the effects on the plants that they did.

On a final note, Penny reported that 40% of the class time was spent on direct instruction. This is in contrast to Cathy's class, which did the same topic and activities as Penny's with different results (see below).

Low TIP – Low SWS – Cathy

Cathy's unit, another ecology unit with a plant growth SI, scored low on both the TIP and the SWS explanations. The TIP in this case was high in Classroom Roles but low in Content/Cognitive Skills and Assessment. The SWS explanations were low in all three categories. An example explanation from this unit follows:

Analyzing and interpreting results: My average for my control was 2.964 and for my experimental it was 1.953. This shows a significant difference between the two plants average growth, so lemon juice I believe stopped the growth of the plant.

Interpreting results: Lemon juice is an acidic base and acids do usually kill plants and stop their growth, which in my case happened. Acid rain tends to fall in certain parts of the earth, so the plants that get hit by it tend to die. In my research I found that some plants can tolerate the acid and can still live after being showered in it. The turnip plant does not tolerate the lemon juice and its acidity contents, because other wise we would have seen a lot more observations but the plant died instead.

This explanation included an explicit claim about the effect of lemon juice on turnip sprouts; the evidence included two data examples with an assertion about significance that was not supported by a p-value. The student's struggle with the reasoning about the acidity of the lemon juice exemplifies the struggles many students had in constructing explanations requiring chemistry knowledge in a unit about ecology.

The unit had virtually the same activities as Penny's: the predator-prey activity, the "Traveling Nitrogen" exercise, the carbon footprint calculation, and the plant growth inquiry with turnip seeds in petri dishes. As with Penny's unit, there was some instruction in scientific reasoning, with the materials instructing students to include a "scientific explanation" with "at least 4 reasons based on science" along with a note that

students may need to do some research to understand why the various additives had the observed effects. The main difference between Cathy's and Penny's TIPs was the breakdown of pedagogical techniques used by the teacher. Cathy spent 40% of the class time on group work and only 15% on lecture, almost the reverse of Penny.

This is consistent with a trend seen across the cases in this study. The teachers who used direct instruction or lecture the most tended to have students who produced explanations with better evidence and reasoning in their work samples (Table 18).

Discussion Summary

This was a multiple case study that analyzed aspects of instruction and student performance related to scientific explanation production in student science inquiry work samples in eight high school science classrooms in the Portland, OR, metropolitan area. The instructional techniques were documented by the teachers with the TIP instrument and analyzed by the researchers using case descriptions, the TIP rubric and qualitative codes. The student explanations were analyzed using an explanation rubric and a qualitative coding scheme.

The TIP analyses showed that most of the teachers in this study were using some instructional techniques identified in the research literature as best practices in inquiry-based classrooms, though only about half were employing them to an adequate level. All of the TIPs scored at least 2 in every category, though three of the seven TIPs earned inadequate scores in two or more of the TIP categories. This may be an overestimate of

the proportion of secondary science teachers using such techniques in the general population, because it is likely that teachers interested in supporting SI were overrepresented in our sample.

The TIPs did not document best practices for teaching explanation. The qualitative TIP analysis showed that half of the teachers offered some instruction focused specifically on explanation, mostly defining one or more of the aspects. McNeill and Krajcik (2008) found that defining without providing the rationale did not help students develop explanation skills on performance assessments. One teacher used modeling to help students differentiate between good and poor claims and evidence, though the models provided did not clearly distinguish claims from evidence. None of the teachers provided students with an explicit rationale for producing scientific explanations, which McNeill and Krajcik (2008) identified as the most powerful technique for supporting explanation among the four examined.

The explanation analyses demonstrated that the high school students in this study were more successful than the middle school students studied in the cited research literature at including claim, evidence and reasoning in their scientific explanations. Eighty percent of students included claim, evidence and reasoning in their explanation, versus 18% in Ruiz-Primo et al. (2010). Reasons for this could include that we examined stand-alone work samples submitted as summative assessments rather than science notebooks, that high school students have more experience with inquiry activities and the expectations for explanation sections, and that we did not include

work samples that contained no discussion or analysis sections. However, high school students still struggle with explanations, especially the reasoning aspect, confirming the findings of Ruiz-Primo et al. (2010) for middle school students. About one out of every five explanations in this study did not include reasoning with science content in their explanations. The Framework for K-12 Science Education (National Research Council, 2012) explains that students should be able to produce explanations that incorporate “their current understanding of science, or a model that represents it.” Students cannot construct adequate explanations without content-based reasoning.

This study also found that while teaching techniques evaluated by the TIP do help students formulate claims, an indicator of a student’s understanding of the overall purpose of the investigation (Ruiz-Primo et al., 2010), direct instruction is most effective for helping students include evidence, reasoning and related science content. The comparison between instruction and student explanation showed that high TIP scores were associated with high Claim scores, but Evidence, Reasoning and composite explanation scores were better in classrooms that scored low in the TIP Classroom Roles category. This finding is consistent with Stephens, McRobbie and Lucas’ (1999) study that found that students’ ability to engage in higher-level model-based reasoning to explain the results of SI investigations required “considerably more guidance” than students in their study received, despite students’ having received instruction in the relevant model prior to their investigations and having carried out careful, successful

investigations that helped them develop correct empirical relationships regarding the phenomena of interest.

This study also confirmed McNeill and Krajcik's (2008) finding that defining aspects of explanation in isolation did not appear to support students in producing strong overall explanations. Felicia defined claim and evidence well, and her students produced adequate claims but were low in evidence and reasoning. Joe also defined claim well and evidence fairly well, and his students wrote adequate claims but included inadequate evidence and reasoning. Cathy and Penny both defined reasoning and had students who performed low in explanation reasoning. The teachers that did not define any aspects of explanation in their units, Abigail, Margot and Sonia, had classes with high explanation scores. Thus, based on the TIPs for these units, defining aspects of explanation did not have a predictable effect on student performance. The one instance of modeling claim and evidence was imprecise in its examples, and was not associated with higher explanation scores. We were not able to find examples of providing rationale for explanation nor for providing everyday examples.

Finally, the study found that students need a solid foundation in the science content required to explain their SI results. The classrooms in which the students scored high in all three aspects of scientific explanation were those in which the SI topic was strongly tied to the content in the unit and in which there was a considerable amount of direct content instruction. This confirms the findings of Stephens et al. (1999) and McNeill and Krajcik (2007) that students need strong content knowledge and

considerable teacher guidance to construct high quality explanations from their SI activities.

Limitations of the study

Although this study provided a snapshot of teacher practice in high school science inquiry classrooms related to science inquiry and scientific explanations, and student performance in producing scientific explanations, this was not a random sample. Because the participating teachers were recruited to attend a workshop to develop inquiry units, and consented to participate in research related to science inquiry, teacher interest in SI was likely overrepresented in the sample.

We were not able to reach an adequate level of independent inter-rater reliability on either the TIP rubric or the Claim category of the explanation rubric. However, the research group was able to come to consensus on the TIP scores. This suggests that a different researcher performing the same analysis alone may produce different results, though a group approach can increase reliability.

Although the TIP is a rich source of information for studying teacher practice, we were not able to observe the teachers' classroom instruction directly, which would have given us a more complete picture of the classrooms and aided our analysis. For example, teachers may have conducted their class discussions or lecture activities in unique ways that were not captured in the TIPs. Teachers also may have provided information to students, such as a rationale for explanation, that was not documented in the TIP because it was verbalized spontaneously or covered earlier in the year, for

example. We also had no way to gauge student engagement or other aspects of classroom culture that may have helped explain the results.

We also did not have any student-level data that may have been valuable in explaining some of the results we found, because this data set was originally collected for a research proposal that did not require student-level detail. For instance, we had no way of knowing how many of the work samples in our data set were produced by students who were English language learners, who had disabilities, were academically gifted, etc. We also did not know about instruction that had taken place earlier in the school year or in previous years surrounding explanation. All of these additional sources of data could have helped us explain our results, and may have reduced the chances that the results were heavily affected by sampling error.

Implications for practice

Science inquiry is an important aspect of science education when implemented effectively. It has the potential to engage students in activity similar to what professional scientists perform when they design experiments to answer open-ended questions about the world. It merges science content with creative thinking as students formulate questions and create acceptable methods to answer them. They must also engage in observation and critical thinking as they consider how to address their research questions, argue from evidence, describe the limitations of their studies and consider alternative explanations.

To implement science inquiry effectively, teachers must help students understand the goal of the project. The instruction must emphasize that the goal is to create a scientific explanation, a cohesive argument that uses evidence to support a scientific claim and contains scientific reasoning that connects the claim to the evidence. However, merely defining claim, evidence and reasoning is not enough. In this study high TIP scores, indicating best teaching practices in inquiry-based classrooms, were associated with adequate scientific claims in the students' work samples. Direct instruction was associated with adequate evidence and reasoning in explanations. This argues for a balanced approach that brings varied pedagogical techniques to bear; an over-reliance on any one pedagogical approach will hamper students' performance in explanation construction.

The results also suggest that moderation is also called for in how much guidance the teacher gives in the formatting of the work sample. Sectioned templates can hamper students' ability to conceive the explanation as a cohesive argument and instead render it an algorithmic exercise in which all the parts must be present while the relationship among them is obscured.

Finally, science inquiry in the classroom is most powerful when the topic of the inquiry is clearly connected to the content in the unit. Students need a firm grounding in the science content to be successful in scientific explanation. The highest-performing classrooms had inquiries in which the content knowledge from the unit was necessary for students to ask appropriate research questions and make sense of their results.

Among the lowest-performing classes in the Reasoning category were those in which the required content was not covered extensively in the unit or the topic of the inquiry was only tenuously related to the unit content.

Implications for professional development

With the release of the Next Generation Science Standards (Achieve, Inc., 2013) comes a strong rationale for students to learn scientific explanation, one of the practices identified in the Framework as essential in science education. This study demonstrated that some teachers are not prepared to support their students in learning this skill. One of the teachers was pleased with her student's performance in their SI work samples, at least 20% of which contained no content-based reasoning at all, hinting that this teacher was unaware of the reasoning required for an adequate explanation. Providers of professional development around the Next Generation Science Standards need to ensure that teachers know what a complete explanation entails.

Teachers also need professional development in learning best practices for teaching explanation. This study confirmed the findings of earlier studies that defining aspects of explanation was not necessarily associated with adequate student performance in writing explanations. We could not confirm the effects of providing rationale for explanation, nor for providing high-quality modeling or everyday examples, but McNeill and Krajcik (2008) suggested that providing explicit rationale for explanation along with good definitions can help.

Finally, the design of the unit culminating in a science inquiry activity influences students' performance in scientific explanation. The units in the study on ecology did not prepare the students for their plant growth inquiry, which required more chemistry and cellular biology knowledge to explain how various substances affected seed sprouting and plant growth. When students needed kinetic molecular theory to explain the results of their reaction rate inquiries, and only received one day of instruction in it (Joe's class), the students struggled to explain their results. High evidence and reasoning scores were associated with higher levels of direct instruction, while high claim scores were associated with more student-centered instruction, a balance of content and cognitive skills and assessment focused on feedback and improvement rather than just grading. These results suggest that teachers should receive support in constructing units that use a variety of pedagogical techniques, including direct instruction and student-centered instruction. It is also crucial that the reasoning needed to explain the results is one of the content focuses of the unit, and students need to spend considerable time in activities around this content before conducting their inquiries.

Questions for further research

Though this study shed some light on what teachers were doing in their classrooms to support students' scientific explanations, it would be interesting to conduct a prospective study in which teachers were asked to provide specific instruction

related to scientific explanations. In this way, researchers could extend McNeill and Krajcik's (2005) results to investigate what pedagogical techniques are most effective for teaching explanation skills.

It would also be useful to characterize the students who produce various qualities of explanations to answer questions such as: Do English language learners have more difficulty producing explanations? What are their difficulties and how can we support them? What are students' attitudes toward SI, and what do they perceive as the goals of their experiments?

Additional questions can be asked about the teachers' beliefs about the purposes and goals of science inquiry, their conceptions of the rationale for and structure of scientific explanations, how those values are communicated to the students, and the effects on student performance in science inquiry and content learning.

There is a great deal to study about teaching and learning of scientific explanation. As our knowledge increases, so may our students' participation in the age of information and scientific achievement be enriched.

References

- Achieve, Inc. (2013, January). *Appendix F - Science and engineering practices in the NGSS*. Retrieved from Next Generation Science Standards: http://www.nextgenscience.org/sites/ngss/files/Appendix%20F_Science%20%20Engineering%20Practice%20-%20FINAL.pdf
- Brown, J., Collins, A., & Duguid, P. (1989). Situated cognition and the culture of learning. *Educational Researcher, 18*(1), 32-42.
- Concannon, J., & Brown, P. L. (2008). Transforming osmosis: Labs to address standards for inquiry. *Science Activities, 45*(3), 23.
- Fradd, S., & Lee, O. (1999). Teachers' roles in promoting science inquiry with students from diverse language backgrounds. *Educational Researcher, 18*, 14–20.
- Hug, B., & McNeill, K. (2008). Use of first-hand and second-hand data in science: Does data type influence classroom conversations? *International Journal of Science Education, 30*(13), 1725–1751.
- Katchevich, D., Hofstein, A., & Mamlok-Naaman, R. (2013). Argumentation in the Chemistry Laboratory: Inquiry and Confirmatory Experiments. *Research in Science Education, 43*, 317-345. doi:10.1007/s11165-011-9267-9
- Kuhn, L., & Reiser, B. (2005). Students constructing and defending evidence-based scientific explanations. *National Associations for Research in Science Teaching*. Dallas, TX.
- Landers, R. N. (2011, November 16). *Computing intraclass correlations (ICC) as estimates of interrater reliability in SPSS*. Retrieved from NeoAcademic: Technology, education and training research from an industrial/organizational (I/O) psychologist in the ivory tower: <http://neoacademic.com/2011/11/16/computing-intraclass-correlations-icc-as-estimates-of-interrater-reliability-in-spss/>
- McNeill, K., & Krajcik, J. (2007). Middle school students' use of appropriate and inappropriate evidence in writing scientific explanations. In M. Lovett, & P. Shah (Eds.), *Thinking with data* (pp. 233-265). New York, NY: Taylor & Francis Group, LLC.

- McNeill, K., & Krajcik, J. (2008). Scientific explanations: Characterizing and evaluating the effects of teachers' instructional practices on student learning. *Journal of Research in Science Teaching*, 45(1), 53-78.
- National Research Council. (2012). *A framework for K-12 science education: Practices, crosscutting concepts and core ideas*. Washington, D.C.: The National Academies Press.
- Onwuegbuzie, A. J., & Leech, N. L. (2007). Sampling designs in qualitative research: Making the sampling process more public. *The Qualitative Report*, 12(2), 238-254.
- Oregon Department of Education. (2009). *Standards by design: High school for science*. Retrieved from <http://www.ode.state.or.us/teachlearn/real/standards/sbd.aspx>
- Rosebery, A., Warren, B., & Conant, F. (1992). Appropriating scientific discourse: Findings from language minority classrooms. *The Journal of the Learning Sciences*, 2, 61-94.
- Ruiz-Primo, M., Li, M., Tsai, S., & Schneider, J. (2010). Testing one premise of scientific inquiry in science classrooms: Examining students' scientific explanations and student learning. *Journal of Research in Science Teaching*, 47(5), 583.
- Sandoval, W. (2003). Conceptual and Epistemic Aspects of Students' Scientific Explanations. *The Journal of the Learning Sciences*, 12(1), 5-51.
- Sandoval, W., & Reiser, B. (2004). Explanation-driven inquiry: Integrating conceptual and epistemic scaffolds for scientific inquiry. *Science Education*, 88, 345-372.
- Saxton, E., & Rigelman, N. (Unpublished). Connect2Math-Connect2Science: Instructional vision and teacher practice assessment tools. *Unpublished manuscript*.
- Toulmin, S. (1958). *The uses of argument*. Cambridge, UK: Cambridge University Press.
- Webb, N. (2009). A teacher's role in promoting collaborative dialogue in the classroom. *British Journal of Educational Psychology*, 79, 21-29.
- Wisconsin Center for Education Research. (2010, November 9). *Surveys of enacted curriculum: State collaborative on assessment and student standards*. Retrieved from Welcome to SEC online: <https://secure.wceruw.org/seconline/secWebHome.htm>

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

TIP Rubric

Teacher Instructional Portfolio (TIP) Rubric

Score	<u>Classroom Roles</u>	<u>Content and Cognitive Skills</u>	Score
4	<ul style="list-style-type: none"> • Lessons and activities require students to primarily rely on their observations, reasoning and interactions with their peers to determine the accuracy of their results; the teacher facilitates student learning by pressing for higher cognitive levels of student thinking when needed. • Lessons and activities require students to engage in inquiry (e.g., asking questions, making observations, collecting data, drawing conclusions) and problem solving (e.g., looking for patterns, making conjectures, justifying those conjectures) as a means of gaining deep conceptual understanding. • Activities/investigations are cognitively challenging with students as active participants engaged in learning that is, in some respect, of their own design. 	<ul style="list-style-type: none"> • The unit is designed to target both content and higher-order cognitive skills, frequently at the same time, presenting students with opportunities to use both content knowledge and cognitive skills in the same activities, lessons, or investigation(s). • Lessons, activities, and investigation(s) in the unit, when taken together, represent multiple and diverse varieties of opportunities for students to construct meaning, apply content, and practice higher-order cognitive skills. 	4
3	<ul style="list-style-type: none"> • Lessons and activities require students mainly rely on their observations and reasoning to determine the accuracy of their results; teachers occasionally take the lead to facilitate learning by posing or answering questions or focusing attention on an important concept or a contradiction. • Lessons and activities require students to participate in inquiry and problem solving to explore concepts, after having completed tasks/activities the teacher structures to provide scaffolding of concepts and/or skills. • Activities/investigations are staged so that students are mostly active participants contributing their own ideas, applying skills previously learned to a new context, or making connection to concepts previously learned. 	<ul style="list-style-type: none"> • The unit is designed to target both content and higher-order cognitive skills with frequent opportunities for students to use both content knowledge and cognitive skills in the same activities, lessons, or investigation(s). • Lessons, activities, and investigation(s) in the unit, when taken together, represent multiple opportunities for students to construct meaning, apply content, and practice higher-order cognitive skills. 	3
2	<ul style="list-style-type: none"> • Lessons and activities require students to mainly rely on their teacher, text, or more capable peers to determine the accuracy of their results; the locus of authority for content knowledge is frequently focused on a source other than the students' reasoning. • Lessons and activities require students to engage in inquiry and problem solving by following a predetermined strategy/format or that are not a cohesive part of the larger unit, which limits the opportunity for deep content learning. • Activities/tasks/investigations are staged so that students are mostly passive participants following procedures or completing activities they know how to accomplish. 	<ul style="list-style-type: none"> • The unit is designed to target both content and higher-order cognitive skills, however, the lessons, activities, and investigations overemphasize content or cognitive skills at the expense of the other learning targets. • Lessons, activities, and investigation(s) in the unit, when taken together, provide inconsistent opportunities for students to construct meaning, apply content, and practice higher-order cognitive skills. 	2
1	<ul style="list-style-type: none"> • Lessons and activities require students to rely on their teacher, text, or other authority to determine the accuracy of their results, observations, or conceptual understanding. • Lessons and activities require students to engage in inquiry and problem solving that have no connection or fail to draw from content learning; inquiry and problem solving is treated as separate from the "regular" curriculum. • Activities/tasks/investigations are staged so that students are passive participants, with rote memorization or mastery of vocabulary being the predominate goal. 	<ul style="list-style-type: none"> • The unit is designed to target only content or higher-order cognitive skills. • Lessons, activities, and investigation(s) in the unit, when taken together, do not represent opportunities for students to construct meaning, apply content, and practice higher-order cognitive skills. 	1
0	<ul style="list-style-type: none"> • Evidence either missing or to insufficient to score. 	<ul style="list-style-type: none"> • Evidence either missing or to insufficient to score. 	0

Teacher Instructional Portfolio (TIP) Rubric

Score	Assessment for Learning	Score
4	<ul style="list-style-type: none"> • Student conceptions <u>and</u> skills are frequently assessed prior to, during, and after lessons through intentional selection of formative and summative assessments that link clearly to learning goals; misconceptions or skill deficits are addressed through intentional teacher moves. • The unit is structured with frequent opportunities for students to self-assess, reflect on their observations, understanding, or teacher/peer feedback, and monitor their progress toward learning targets. • The unit's assessments consistently have appropriate cognitive demand, represent appropriate learning progressions, are aligned with designated learning goals and standards, and generate evidence of performance criteria for meeting those goals. 	4
3	<ul style="list-style-type: none"> • Student conceptions are frequently assessed prior to, during, and after lessons through intentional selection of formative and summative assessments that link clearly to learning goals. • The unit is structured with periodic opportunities for students to self-assess, reflect on their observations, understanding, or teacher/peer feedback, and monitor their progress toward learning targets. • The unit's assessments primarily have developmentally appropriate cognitive demand, represent appropriate learning progressions, are aligned with designated learning goals and standards, and generate evidence of performance criteria for meeting those goals. 	3
2	<ul style="list-style-type: none"> • Student conceptions sometimes are assessed prior to, during, and after lessons through intentional selection of formative and summative assessments that link to some, but not all learning goals. • The unit is structured with opportunities for students to self-assess, or monitor their progress toward learning targets, but these opportunities are infrequent or superficial. • The majority of the unit's assessment are aligned with designated learning goals and standards and generate evidence of performance criteria for meeting those goals, however, a few assessments are either not appropriate in terms of cognitive demand, not representative of learning progressions, or misaligned with goals/standards. 	2
1	<ul style="list-style-type: none"> • Student conceptions are infrequently or not assessed prior to, during, and/or after lessons through intentional selection of formative and summative assessments. • The unit not is structured with opportunities for students to self-assess, or monitor their progress toward learning targets. • The unit's assessments primarily are either not aligned with designated learning goals and standards, not representative of learning progressions, fail to generate evidence of performance criteria for meeting those goals, or are not developmentally appropriate in terms of cognitive demand. 	1
0	<ul style="list-style-type: none"> • Evidence either missing or to insufficient to score. 	0

Appendix B

TIP Qualitative Codes and Definitions

Is there evidence in the TIP that the teacher's instruction included:

Practice \ Class	1	2	3	4	5	6	7	8
Defining explanation (M)								
Making rationale explicit (M)								
Modeling explanation (M)								
Connecting science explanation to everyday explanation (M)								
Explanation mentioned in TIP? (T)								
Degree of scaffolding evident in TIP (T)								

(M) Codes based on McNeill & Krajcik (2008)

(T) Codes added based on trends seen in this study

Table 1
Code for defining scientific explanation

	Level	Description of Level
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 define the component.
3	Vague definition	The teacher provided a vague definition of the component.
4	Accurate but incomplete definition	The teacher 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.

Table 2
Code for making the rationale of scientific explanation explicit

	Level	Description of Level
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.

Table 3
Code for modeling scientific explanation

	Level	Description of Level
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	The teacher 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	The teacher 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	The teacher 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 these 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 in the example or explicitly stated that the explanation did not include that component.

Table 4
Codes for connecting scientific explanation to everyday explanation

	Level	Description of Level
0	Does not mention an everyday example	The teacher does 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 component	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 all three of the components (i.e., claim, evidence, or reasoning) in relation to the example.

Note: Tables 1 – 4 in this appendix were transferred directly from McNeill and Krajcik (2008)

Appendix C

Explanation Rubric

Is there a claim ? – Statement(s) in the explanation section that addresses the hypothesis or answers the research question. Note: If no data are found in the work sample, still score claim as if there are data. If the claim is scientifically plausible, score as if supported by the data.	
If no, score zero. If yes, pick claim score (choose one):	
	1 pt - Claim does not address the research question or hypothesis, or is self-contradictory or unintelligible. (If it is unclear what is claimed, score 1). Note: Any statement akin to, “My hypothesis was correct” without any restatement of the hypothesis can only earn a 1.
	2 pt - Claim does not contain all of the relevant parts (e.g. only dependent variable trend described, or experimental result stated without reference to purpose of investigation), or is implicit. &/or Incorrectly addresses the hypothesis or research question, or is not supported by the data
	3 pt - Claim is explicit, related to the research question and is supported by the data, but may have to be pieced together, or may be found in data analysis or error analysis section.
	4 pt - Claim is explicit, and relationship to research question or hypothesis is explicitly stated. Claim is in the same section as the evidence and reasoning (if evidence and/or reasoning present), and does not have to be “pieced together” from disparate parts of the analysis/conclusion section.

Is there evidence ? – Data referenced or explicitly cited that supports or seems intended to support the claim (or could support a claim if claim is missing).	
If no, score zero. If yes, pick evidence score (choose one for highest level present):	
	1 pt - Words ‘data,’ ‘table,’ ‘graph’ or other indication that data were considered are present; &/or Evidence consists of anecdotal data – stories, past experience, or analogies &/or Data pattern or examples cited but source could not be found in SWS &/or There are serious flaws with the use of data.
	2 pt - Only examples or means explicitly cited (no pattern referenced or described) &/or data descriptors or rankings cited without language indicating a pattern.
	3 pt - Pattern of data described – qualitative or quantitative – at least one descriptor of trends in the data. Pattern language may be implicit or vague. To be a pattern, language must incorporate at least three data points and describe a trend. (Vague reference to multiple trials sufficient for at least three data points.) Data points may not be explicitly cited for this category. Pattern cannot be merely a restatement of the claim.
	4 pt - Pattern (qual or quant) + at least 1 data point to support pattern description. – Pattern language must be explicit, and data example must be proximate to pattern description and explicitly cited as support for pattern.

Is there reasoning that purports to support the claim? – Relevant science content or other logical statements that bolster the claim	
If no, score zero. If yes, pick reasoning score (choose one for highest level present):	
	1 pt - Reasoning is all or mostly irrelevant, uses mostly incorrect science content or science content is very weak or is unintelligible
	2 pt - Reasoning is partially aligned with claim, i.e. contains some incorrect, irrelevant or tangential statements but also some correct, relevant content or statements that bolster the claim. Note: For 2 pt, at least one relevant, reasonable explanation or speculation must be present, i.e. at least one reasonable “because” type statement.
	3 pt - Reasoning is aligned with claim, and all or most statements in explanation section are relevant to claim. Reasoning may presented separately from the claim, or the statements are disconnected or don’t flow together to make a convincing argument. &/or Some important pieces of content are missing.
	4 pt - Reasoning is completely aligned with claim, i.e. all or most statements are relevant and statements flow together to make a convincing argument with no important pieces missing. Some science content used correctly is required.

Appendix D

Definitions and Guidance for Qualitative Codes

- I. Claim - How well does the claim address the question posed?
 - a. All aspects of claim expected for the work sample topic were addressed.
 - b. Some parts of expected claim were left out but some were addressed.
 - c. Claim did not address research question.
 - d. No claim was present.
- II. Evidence
 - a. Type: What type of evidence was provided?
 - i. Investigation-based: Data can be tracked to tables or graphs in SWS
 - ii. Artificial – data could not be found in the SWS
 - iii. Anecdotal
 - iv. Word ‘data’ only mentioned = only words “data” “table” or “graph” without citing any actual data
 - b. Nature: Were the data presented as patterns or isolated examples?
 - i. Patterns (qualitative or quantitative) only
 - ii. Patterns and examples
 - iii. Examples only
 - c. Sufficiency: Was there enough evidence to support the claim? Data considered to be sufficient if:
 - i. Investigation-based and described as a pattern, OR
 - ii. Investigation based and at least two examples for each part of the claim (in paper, claim in two parts => at least 4 examples)
- III. Reasoning
 - a. Alignment – Is the evidence linked to the claim or are they presented as separate entities?
 - i. Alignment is complete; reasoning in explanation connects evidence to claim without tangential or irrelevant statements.
 - ii. Partial – Some reasoning statements not accurate and/or not related to claim.
 - iii. No – Reasoning statements completely incorrect and/or not related to claim.
 - b. Quality of the link – How was the evidence linked to the claim? For example, with an elaborated statement, or with only a “because” or a “for example” link?
- IV. Other
 - a. Was there differentiation between claims and evidence?

- i. Naming evidence source (“The graph of conc. vs.time showed...”)
 - ii. Generally referencing evidence (“The data showed...”)
 - iii. Referencing time or context (“When the rainfall dropped...”)
 - iv. Attributing confidence (“I know this because...”)
 - v. Implication that evidence is something they know, but inference is something they think (“Because the rainfall dropped, I think the finches regular food source died...”)
- b. Were there overtly persuasive statements?
 - i. Statements or phrases such as “I was right” or “My hypothesis was correct.”
 - ii. Refuted counterarguments such as “I first thought x to be true but my results showed it wasn’t.”
- c. Did the student consider the quality of the evidence, such as identifying sources of error, stating that evidence was missing, that sh/e ran out of time to do more trials?
- d. Did the student consider alternative explanations?
- e. Did the student consider broader implications such as:
 - i. Possible follow-up questions or experiments
 - ii. Relatedness to other topics or content studied
 - iii. Real-life applications of findings
 - iv. Limitations of the findings

Additional Guidance of Qualitative Coding of Student Explanations Argumentation Principles

For the “differentiation between claims and evidence” category, Kuhn and Reiser (2005) provide the following example:

The students in the first example have embedded their evidence with their suppositions. As a result of this rhetorical structure, it is difficult for the audience to determine which information is fact and which information is the result of inferences the students made. For example, examine the following sentence in which we italicized the facts that are available in the computer database:

“Since those plants [Chamae, Portulaca and cactus] were very scarce, there was one other plant called the Tribulus, which had harder and lengthier seeds so the best chance for survival was to adapt to the Tribulus and be able to eat the seeds....”

Without being familiar with the students’ problem context (including the instructional sequence and computer supports), it is difficult to make the distinction elucidated by the italics. This raises questions such as how do the students know that the surviving birds ate Tribulus? Thus, these students have provided little guidance to support the

reader in determining what is fact and what is inference thereby making it difficult for the reader to evaluate whether the claim is accurate and believable.

As a positive example, they describe the following:

The majority of the second example provides clearly distinguished evidence and inferences. For example, the sentence “The charts of cactus, Portulaca, and Chamae all show a major decrease to zero, from wet '73 to wet '77 except for the Tribulus plant.” labels the data source thereby helping to identify the information as fact.

Many explanations will have both. In this case, examine the entire explanation and make a holistic judgment as to whether the claims were mostly differentiated from the inferences.

Look for explicit citation of data source to differentiate, such as:

- 1) Naming the evidence source, such as “The charts of cactus, Portulaca, and Chamae all show...”
- 2) Generally referencing the evidence “The research of four birds that survived showed...” or “the graph shows that...”
- 3) *Bounding their statements by referencing the time or context* in which their evidence occurred. This is a weak data reference, however we find that statements such as “When the invasive species was put into the environment...” communicate that the observation the student is referencing did occur.
- 4) *Attributing confidence in the information presented* with statements such as “I know this [the claim] because...” As with our third category, these statements of confidence are a weak reference to data. However, statements such as this were often used to rhetorically separate the facts from the inferences.
- 5) Similarly, some students differentiated between their inferences and evidence by *implying that the evidence is something they know*. Students do this by calling the inference out as something that is not fact using phrases such as: “I think” or “I believe.” For example, examine the following response “...There was a nice size number of foxes and invasives [sic] but the rabbit and grass populations were pretty low. I believe that the rabbits couldn't have eaten all that grass if they were leaving so quickly”

(Kuhn & Reiser, 2005)

For “overtly persuasive statements, look for:

1. *Asserting the accuracy of the claim*, by using phrases such as “this [the evidence above] proves...” or, as in the example from the finch responses, “...our hypothesis is correct...”

2. *Providing a counter argument* to strengthen the claim. For example, in the following example the student states her original hypothesis and then demonstrates why it is incorrect: “The invasive species is competing for grass with bunnys [sic]. At first my group thought that the invasive species was competing with foxes, for bunnys. After more generations and more research, when bunnys die out, the invasive species lives on [so it must not eat bunnies]...” (Classroom 2, Student EJ, lesson 6). While this example does not provide evidence supporting the claim, it does attempt to persuade the reader that the alternative is incorrect.

(Kuhn & Reiser, 2005)