Curriculum implementation for scientific argumentation: Fidelity to procedure versus fidelity to goals

Katherine L. McNeill, Lisa M. Marco-Bujosa, María González-Howard & Suzanna Loper

Boston College

contact info:
Katherine L. McNeill
Lynch School of Education, Boston College
140 Commonwealth Avenue, Chestnut Hill, MA 02467
Phone: 617-552-4229
Fax: 617-552-1840
kmcneill@bc.edu

Reference as:

McNeill, K.L., Marco-Bujosa, L. M., González-Howard, M., Loper, S. (2016, April). *Curriculum implementation for scientific argumentation: fidelity to procedure versus fidelity to goals.* Paper presented at the annual meeting of the National Association for Research in Science Teaching, Baltimore, MD.

Curriculum implementation for scientific argumentation: Fidelity to procedure versus fidelity to goals

In the current standards-based accountability era, there is an increased focus on efficacy or effectiveness studies that measure how closely implementation aligns with the original curricular intervention (National Research Council, 2005). This trend, combined with an interest in the scale up of interventions, has resulted in education researchers focusing on fidelity of implementation (FOI) in education settings (O'Donnell, 2008; Lee, Penfield, & Maerten-Rivera, 2009). The goal underlying this research on implementation is to determine the impact of the intervention on outcomes. Consequently, FOI has received increased attention in calls for funding and research; however, there are numerous ways of conceptualizing and measuring this construct, ranging from a strict adherence to the procedural elements included of a curriculum (O'Donnell, 2008) to a consideration of the role of the teacher to make appropriate instructional decisions to support student learning (Brown, 2009).

We argue that the conceptualization and measurement of FOI is particularly important to consider in relation to recent science education reform efforts. The Next Generation Science Standards (NGSS) (NGSS Lead States, 2013) include a focus on science practices, which require a significant shift in science instruction away from students memorizing "final form" ideas to demonstrating "knowledge in use" as students construct explanations and develop models about the natural world (Berland, Schwarz, Krist, Kenyon, Lo, & Reiser, in press). The focus on student-driven learning is a unique aspect of the NGSS, as it encourages students actively engaging in these science practices (National Research Council, 2015), which requires the teacher to support students in ways that differ from previous science instruction.

Specifically, our work focuses on the practice of argumentation, which has typically not been a part of science classrooms (Osborne, 2010). In scientific argumentation, students engage in dialogical interactions in which they construct and critique claims using evidence (Ford, 2012). This shift in curriculum goals to include student engagement in the discursive practice of argumentation could potentially result in new challenges around FOI, which have traditionally followed the steps and actions of teachers as opposed to the actions of the students (O'Donnell, 2008). Consequently, we were interested in exploring different conceptualizations of FOI in relation to the enactment of a middle school science curriculum focused on argumentation.

Conceptual Framework

Fidelity of Implementation

FOI is a relatively new, yet growing concept in education research and implementation studies that has been defined in a variety of ways (Century, Rudnick & Freeman, 2010). Drawing upon research in public health, Lee, Penfield, and Martin (2009), defined FOI as, "the determination of how well an innovation is implemented according to its original program design or as intended" (p. 837). Alternately, FOI could be defined as the extent to which a user's practice matches the "ideal" implementation of an intervention (O'Donnell, 2008, p. 34). Defining what constitutes "ideal," in education research, however, has been less clear. For curriculum developers, the intent of FOI is often to verify whether the intervention is implemented as planned in order to appropriately attribute changes in the outcomes to an educational innovation (Lee et al., 2009). Yet, there are few studies to guide researchers on how FOI of core curriculum interventions can be measured (Century et al, 2010). The literature on curriculum implementation describes several different perspectives on FOI that have been advanced in the field to assess implementation.

In education, many researchers have defined FOI based upon *procedural* elements such as the number, order and alignment of methods prescribed in a curriculum (O'Donnell, 2008). For example, Mowbray and colleagues (2003) emphasized fidelity criteria in terms of structure (i.e., the framework for service delivery) and process (i.e., the way in which services are delivered), and in a literature review, O'Donnell (2008) emphasized a description of FOI as fidelity to structure (i.e., adherence, duration) and fidelity to process (i.e., quality of delivery, program differentiation). A number of studies also define FOI in terms of instructional quality, or, adherence and integrity (O'Donnell, 2008). For example, Lee, Penfield, and Martin (2009) operationalized FOI in terms of the quality of instructional delivery, dose or exposure, and participant responsiveness. However, Lee and colleagues' (2009) results indicate their measures of FOI, from teachers' self-reports and classroom observations, had no significant effects on students' science achievement gains, which the authors conjectured may have been due to measurement errors or their conceptualization of FOI.

These measures of FOI for curriculum use do not take into consideration the adaptive and reactive aspects of teaching practice, making these measures of instructional quality for FOI limited. As Shulman (1990) noted, "While curriculum might be the backdrop for teaching, the two are not to be confused" (p. vii). The challenge for educational researchers attempting to measure FOI is how to distinguish between good teaching and good teaching practices prompted by the curriculum (O'Donnell, 2008, p. 44). Therefore, a new conceptualization of FOI is needed in which the role of the teacher is considered (Cho, 1998).

Remillard (2005) argued that it is essential to consider a teacher's "curriculum-in-use" and the "teacher-curriculum relationship," both of which are useful to a reconceptualization of FOI in the context of the implementation of curriculum. According to Remillard (2005), "curriculum use refers to how individual teachers interact with, draw on, refer to, and are influenced by material resources designed to guide instruction," (p. 212). This view considers the teacher has an important role within a unique instructional context, as they interpret, adapt and implement the curriculum. Therefore, this perspective assumes the teacher is an active designer of curriculum rather than solely an implementer. Similarly, Buxton and colleagues (2015) extrapolated on the teacher- curriculum relationship by theorizing the variation in teacher enactment as "multiplicities of enactment" rather than inadequate implementation. They observed and described this variation in enactment, "as teachers taking ownership of the practices in ways that may be more sustainable, flexible, and responsive to ongoing changes in their classroom contexts," (Buxton, Allexsaht-Snider, Kayumova, Aghasaleh, Choi, & Cohen, 2015, p. 499).

Although this focus on curriculum-in-use has been critiqued as in conflict with FOI (O'Donnell, 2008), we argue that this perspective necessitates a different conceptualization and measurement focused on *goals* rather than procedural elements. *Fidelity to goals* focuses on the alignment with the overarching learning goals in the curriculum rather than the prescribed methods in a lesson plan. This measure of fidelity would explain the degree of variation in implementation and how it might affect or moderate outcomes. From this perspective, measuring FOI for curriculum use would involve specifying the critical components and processes of the curriculum's theory, or goals for changing teaching practice (O'Donnell, 2008). Furthermore, this view problematizes the more traditional perspective on fidelity, as it considers that the modifications made by teachers may not be a bad thing; in fact, from this perspective, teacher modifications to a given curriculum may better support student engagement in the science practices if they align with the overarching learning goals. However, focusing on fidelity to goals may not capture other important aspects of teachers' curriculum enactment. Consequently, we see two potentially very different conceptualizations of FOI in the literature – fidelity to procedure versus fidelity to goals. We are interested in how these different conceptualizations relate to science practices in recent reform efforts.

Argumentation

Considering recent reform documents in science education, the conceptualization of fidelity could have important implications for how to best support teachers and students in science practices, such as scientific argumentation. Research has shown teachers' enactment of curriculum addressing argumentation varies greatly, with teachers often making a number of different types of adaptions (Berland & Reiser, 2011; Herrenkohl & Cornelius, 2013). Thus, scientific argumentation is a productive area in which to explore FOI.

Similar to others (McNeill, González-Howard, Katsh-Singer, & Loper, 2016; Jiménez-Aleixandre & Erduran, 2008), we define argumentation in terms of both a *structural* and *dialogic* focus. The structure of an argument consists of a claim about the natural world that is supported by both evidence and scientific reasoning (McNeill, Lizotte, Krajcik, & Marx, 2006). Claims are justified using evidence, such as scientific data, and reasoning, which explains why the evidence supports the claim. In addition to its structure, argumentation involves a dialogic process in which students construct arguments through interaction with their classmates. These interactions include students questioning and critiquing competing claims (Ford, 2012). The dialogic process of argumentation emphasizes a classroom goal to collaboratively make sense of phenomena and convince others. Thus, scientific argumentation differs greatly from the typical science instruction, in which students generally interact with the teacher rather than other students (Berland, 2011).

Research has documented that teachers' enactment of scientific argumentation varies greatly (Berland & Reiser, 2011). When implementing science curriculum focused on argumentation, teachers can make adaptations to both the structural and dialogic elements of this science practice. For example, research has found teachers oversimplify the structural elements, such as removing the sensemaking focus in how they defined scientific argumentation for students, which resulted in a lack of student ability to use evidence and reasoning to explain a scientific phenomenon (McNeill, 2009). Other research documents the importance of teacher instructional moves to support student engagement in the structural aspects of argumentation, including asking questions and providing examples, to encourage students to use evidence and reasoning in classroom discourse (McNeill & Pimentel, 2010). Sampson and Blanchard (2010), focusing specifically on teacher knowledge of argument structure, found teachers had difficulty providing solid evidence and reasoning in support of a claim, indicating teachers also struggle with these concepts and may require explicit curricular supports to engage students in learning about argumentation in their science classroom.

In addition, teachers may adapt lessons focused on student-to-student interactions into more traditional teacher-led discussions. For example, in a study of curriculum supports for leading discussions in high school science, Alozie and colleagues (2010) found teachers relied on traditional "recitation" formats, and concluded that curricular supports were necessary to help teachers promote dialogic interactions in their classroom. Other research has documented the important role of the teacher, particularly the teacher's use of open-ended questions and referring to earlier comments made by students, in encouraging dialogic interactions between students (McNeill & Pimentel, 2010).

Given this variation in teacher enactment of argumentation in the classroom, and teacher ability to support student engagement in the practice, the present study was designed to explicitly explore FOI, in the context of a curriculum supporting argumentation, in two ways: fidelity to procedures in argumentation lessons and fidelity to argumentation goals.

Methods

Curricular Context. This study took place during the pilot of a digital life science curriculum for middle school students. Specifically, teachers enacted two units, *Microbiome* and *Metabolism*, which were designed to take approximately eight weeks of classroom instruction (Regents of the University of California, 2013a & 2013b). In addition to digital resources, teachers were provided with kits that included physical manipulatives for student investigations as well as student notebooks.

This was a reform-oriented educative curriculum specifically designed to support teachers' abilities to incorporate scientific argumentation into their instruction. As previously described, the curriculum addressed two goals of scientific argumentation: argument structure and dialogic interactions. The curriculum considered the structure of an argument to consist of a claim about the natural world that is supported by both evidence and scientific reasoning (McNeill, et al., 2006). The curriculum also addressed the dialogic aspects of argumentation, emphasizing scientific argumentation as a social process in which students construct, evaluate, and revise arguments through interaction with their classmates (Berland & Reiser, 2011).

The curriculum was educative in that it was designed to support teacher learning (Davis & Krajcik, 2005) about both the structural and dialogic aspects of argumentation. Educative supports were provided through both text and multimedia formats, such as videos, which provided teachers with real examples of what the structural and dialogic aspects of argumentation looked like in practice. Given the fact that teacher implementation of curriculum addressing argumentation varies widely (Berland & Reiser, 2011; Herrenkohl & Cornelius, 2013), this curriculum is a useful context to explore fidelity of implementation to the scientific practices advocated for in recent science education reform documents, notably the NGSS (NGSS Lead States, 2013).

Participants. Across the country, twenty teachers enacted the pilot curriculum. The participants in this study included five teachers selected based on their vicinity to the two research teams, which enabled the collection of videos of classroom enactment. The five teachers had a range of teaching experience from a second year teacher to over 20 years of teaching experience (Table 1).

Table 1: Teachers' Backgrounds

Teacher	School	Science Credential	Highest Degree Education	Highest Degree Science	Years Teaching Experience	# Of Argument Trainings
Ms. Majestic	School P	None	MA	BA	20 or more	1
Ms. Ransom	School S	MS/HS[1]	MA	BA	20 or more	1
Mr. McDonald	School S	MS/HS[1]	MA	BA	6 to 10	2 or 3
Ms. Newbury	School U	MS/HS[1]	MA	None	6 to 10	2 or 3
Mr. Arlington	School U	MS/HS[1]	BA	BA	2	1

[1] MS/HS = Middle School or High School Science Credential

The teachers taught in three different schools (Table 2). Ms. Majestic taught in a private school while Ms. Ransom and Mr. McDonald taught in a suburban public school. Both of these schools had a low percentage of students eligible for free or reduced lunch. Ms. Newbury and Mr. Arlington taught in an urban public school with a high percentage of students eligible for free or reduced lunch. In addition, Ms. Newbury's class was a sheltered English immersion (SEI) science classroom. SEI is an instructional model in which the teacher is responsible for teaching content and language learning objectives (Echevarria, Vogt, & Short, 2008). Her classroom consisted of 6th and 7th grade students who were all native Spanish speakers from Central or South America, who had recently immigrated to the United States.

Table 2: School Characteristics

Teacher	School	Type of School	% Of Students Eligible for Free or Reduced Lunch	% Of Students who are second language learners	Grade Level	Class Size
Ms. Majestic	School I	Private	Less than 25%	Less than 25%	7th	21-25
Ms. Ransom	School S	Public	Less than 25%	Less than 25%	7th	21-25
Mr. McDonald	School S	Public	Less than 25%	Less than 25%	7th	21-25
Ms. Newbury	School U	Public	More than 75%	25-50%	6th, 7th	15-20
Mr. Arlington	School U	Public	More than 75%	25-50%	6th	26-30

Data collection and analysis. We selected six lessons from the two units focused on argumentation to examine teachers' fidelity of implementation. Table 3 includes a summary of the six lessons. These six lessons varied with respect to the type of activities included as well as the argument goals addressed. For example, the lessons in which students read and wrote arguments (i.e Microbiome Lesson 1.9 and Metabolism Lesson 1.12) included more of a focus on the structure of an argument, whereas other lessons had more of a focus on argumentation as a dialogic process, such as Microbiome Lesson 1.10, in which students created a video argument, and Metabolism Lesson 2.10, in which students engaged in a class discussion, called a science seminar. Across the lessons, there were a range of activity structures (e.g. card sort, writing arguments, science seminar) targeting the argumentation goals.

Table 3: Summary of Argumentation Lessons

Curriculum Unit	Lesson Focus	Lesson Description
Microbiome	1.6: Identifying Claims and Evidence	• Warm-up: what do you notice happened to the patient after week 3? • Return to case study patient: review and set the purpose to investigate how antibiotics affect the microbiome. • Introduce claims and evidence in a scientific argument about the effect of antibiotics. • Observe the effect of antibiotics on agar plate, gather evidence. • Card sort to identify evidence supporting claims. • Reflect on patient and what they learned about the effect of antibiotics on the microbiome.

	1.9: Writing a Scientific Argument	• Warm-up: Review arguments read for homework and identify which is most persuasive and why. • Review and discuss two arguments; how are they similar or different, and which is most persuasive? Discuss organization and connections. • Highlight language of argumentation with sentence starters • Write scientific argument based on evidence from card sort.
	1.10: Presenting a Scientific Argument	• Warm-up: Review written argument to identify points to share with class. • Discuss claims and evidence to develop a complete story about why the fecal transplant was successful. • Students work in groups to share their argument in favor of their assigned claim. • Review video from politician and discuss the politician's argument. • Plan and create video responses. • Reflect on the unit.
Metabolism	1.12: Writing an Argument	• Warm-up: Prompt students to generate their own key concepts about what can go wrong to prevent molecules from getting to cells. • Student prepare to write their own argument by reviewing the purpose and question. A template is provided. Justification is introduced. • Students write arguments. • Introduce homework, reading about new intern role.
	2.8: Using simulation to gather evidence	• Warm-up: Think about possible interactions between activity level and growth and repair. • Use simulation to gather evidence and advise athlete about growth and repair. • Discuss simulation results as a class, projecting claims and presenting evidence in a t chart. Students identify the claim that is best supported by evidence. • Prepare for homework, using simulation to compare athletes to non-athletes.
	2.10: Science Seminar	• Warm up: Review homework, identify claims to use in seminar. • Prepare for the science seminar by writing their best ideas on the science seminar evidence sheet. • Review the purpose and structure in the science seminar. • Group 1 debates their explanations while Group 2 observes. • Group 2 debates their explanations while Group 1 observes. • Reflect on the key concepts discussed in the Science Seminar about how training to be an athlete results in changes in the body that allow it to become better at releasing energy.

All six lessons were video recorded and coded using two different FOI coding schemes. Both coding schemes were developed based on our theoretical framework and an iterative analysis of the video data (Miles & Huberman, 1994).

The first coding scheme focused on *Fidelity to Procedure* in terms of the adherence to the order and types of procedures described in the activity structures within each lesson. We identified a shift in activity structure based upon a change in what the students were doing (e.g. students writing to full class discussion). Lesson specific coding schemes were created for each of the six lessons breaking down each lesson to between 5 and 9 activity structures for a total of 40 distinct activities across the 6 lessons. For each activity, we coded the video recording for one of three codes: aligned, modified or skipped (Table 4). A teacher's enactment was coded as *aligned* when it matched the procedure in the activity structure. An activity was coded as *modified* when it aligned with some components of the description, but included an alteration (e.g. included a full class discussion, but did not use a t-chart to structure it) or different order (e.g. decided to have a discussion before students completed individual writing). Finally, an activity was coded as *skipped* if the teacher did not include any element of that activity with his/her students. Two independent raters coded each video for *Fidelity to Procedure*. Inter-rater reliability was calculated by percent agreement and was 79% across the six lessons. All disagreements were resolved through discussion.

Table 4: Coding scheme for Fidelity to Procedure.

Code	Description	Example
Aligned	The teacher's enactment aligned with the overarching activity structure and focus of the section of the lesson.	The lesson began with an independent "Do Now" activity in which students wrote a response in their lab notebook. A teacher whose enactment aligned had students work independently writing their responses to the Do Now.
Modified	The teacher modified an activity so it aligned with some components of the description, but did not include all, or follows a different activity structure, a different focus, or a new order.	For the Do Now activity described above, a teacher whose enactment was coded as modified had students discuss their answers to the Do Now activity in pairs rather than working independently.
Skipped	The teacher did not complete this activity with his/her students.	A teacher whose enactment was coded as skipped did not use the Do Now activity.

The second coding scheme focused on *Fidelity to Goals* in terms of the adherence to the overarching argumentation goals within the curriculum. We coded each lesson for the quality of argumentation instruction focusing on four goals (McNeill, et al., 2016). The first two argumentation goals focused on the structure of an argument: 1A: Use of high quality evidence (Evidence), 1B: Use of scientific ideas to explain the link between the evidence and claim (Reasoning). The second two goals focused on argumentation as a dialogic process: 2A: Students building off of and critiquing each other's ideas (Interactions) and 2B: Students critiquing competing claims (Competing Claims).

For each of the four goals, teacher enactment was coded for four elements: 1) Teachers' description of the goal, 2) Teachers rationale for the goal, 3) Teacher models and prompts, and 4) Student engagement. The first three codes focused on teacher instructional strategies while the final code focused on the role of the students. We included the code for students, because alignment with the argumentation goals often required students taking more ownership over the classroom discourse. Consequently, a teacher's silence was often a productive indicator of student engagement in argumentation. For each of the sixteen codes (4 goals each with 4 codes), we rated each lesson in terms of high quality (Level 2), low quality (Level 1) or not present (Level 0). We developed detailed coding schemes for each of the four goals. Table 6 includes part of the coding scheme, specifically for Argumentation Goal 1A: The Use of High Quality Evidence, to illustrate the high quality code (Level 2) in terms of both a description and a teacher example.

Three independent raters coded each teacher's video for *Fidelity to Goals*. The inter-rater reliability, which was calculated by percent agreement for each pair, was 77%, 78% and 80% across all of the lessons. Disagreements were resolved through discussion in which the raters revisited the video and discussed the alignment with the argumentation goals.

Table 6: Coding scheme for Fidelity to Goals for Goal 1A: The use of high quality evidence.

Category	Coding Scheme 2 – Present – High Quality	Example of High Quality Teacher Enactment
1. Teacher provides description	Teacher describes scientific evidence including these two components: • High quality evidence consists of data such as accurate measurements and observations. • Empirical evidence does not include students' opinions and personal experiences.	The teacher explains that in science evidence includes measurements and observations, not personal opinions.
2. Teacher provides rationale	Teacher provides at least two reasons why the use of high quality evidence is important. Reasons could include: 1. Scientists use evidence 2. Using evidence allows you to make sense of the natural world <u>or</u> to decide which is the strongest among claims. 3. The use of high quality evidence makes an argument more persuasive. 4. This skill is applicable to every day context or across disciplines	The teacher describes evidence is useful because it allows you to decide which is the strongest among claims and allows you to make a more persuasive argument.
3. Teacher models and prompts	The teacher models <u>and</u> prompts for evidence. Examples could include: • Teacher models by providing an example of high quality or low quality evidence. • Teacher provides models for students such as a t-chart to prompt them to use high quality evidence. • Teacher provides prompts such as – Remember to include evidence to support	The teacher models the use of high quality evidence by providing an example of high quality evidence and prompts students to their evidence by asking, "Does this evidence support your claim?"

	your claim or you can use the maps as evidence. • Teacher uses questions to prompt – What is your evidence? How would you critique the quality of this evidence?	
4. Student Engagement	Numerous students support their claims with high-quality evidence. This code is given when high-quality evidence seems to be a part of the classroom norms in terms of the students' contributions and interactions.	Many students are observed using and discussing high quality evidence during the lesson.

We used the two FOI coding schemes, *procedural* and *goals*, to examine differences in fidelity across the five teachers' enactments of the argumentation lessons. Specifically, we developed matrices and graphical representations of the analyses to look for patterns (Miles & Huberman, 1994) both across teachers and across lessons. No trends emerged in relation to the six different argumentation lessons. However, distinct patterns did exist in relation to the teachers.

Consequently, we then used the codes to develop case studies for each teacher focused on the quality of the argumentation instruction and the specific changes they made to the lessons. The first author developed the case studies to depict the complexities within each classroom and develop a narrative that captured the most important features of each classroom (Stake, 2000) around their enactment of the argumentation lessons. These case studies ranged from 9-11 pages single-spaced. Each case study was then read by the two other members of the research team, who were familiar with the classrooms, having coded each video with the two FOI coding schemes. Any discrepancies in the case studies were revised after discussion.

Results

We first provide the overall synthesis of codes to describe the key trends in the teachers' enactments for both *Fidelity to Procedure* and *Fidelity to Goal*. We then focus on one lesson, Microbiome Lesson 1.9, to illustrate the differences using examples from three case study teachers.

Fidelity to Procedure

In terms of procedure, there was variation across the five teachers' enactments of the argumentation lessons (Figure 1). As mentioned previously, the *Fidelity to Procedure* focused on the order and type of procedures described in the activity structures for each of the six argumentation lessons. Three teachers had high fidelity of procedure, with around 80% of the activities adhered closely to the curriculum. For these three teachers, they typically completed all of the activities described in the lessons and used the activities in the recommended order. Two of the teachers, Mr. Arlington and Ms. Newbury, had lower levels of alignment with about 40% of the activities closely aligning with the curriculum. Both teachers were more likely to modify and skip activities within the argumentation lessons.

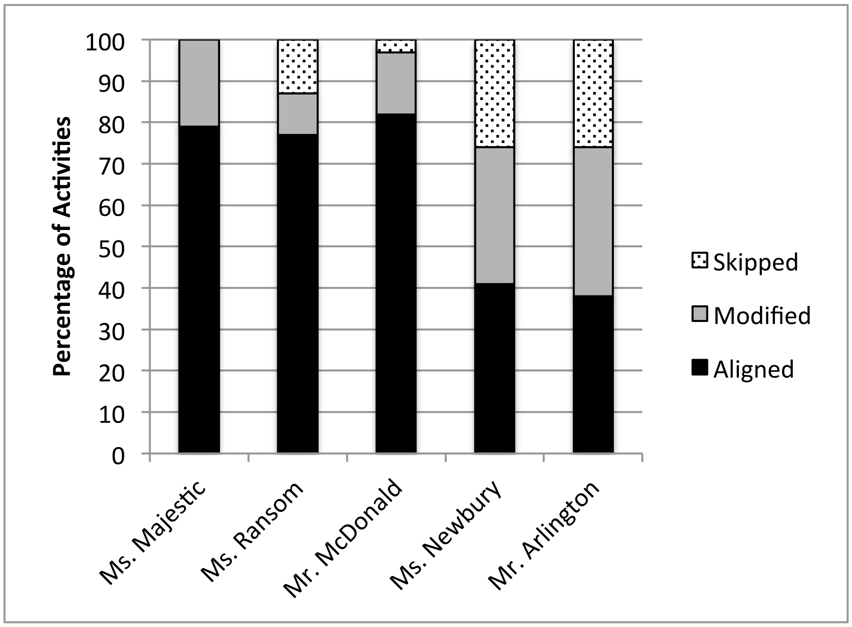

Figure 1: Fidelity to Procedure

Fidelity to Goals

For the goal of argumentation, there was again variation across the five teachers' enactments, but the pattern here was different. As described previously, this coding scheme focused on the four argumentation goals targeted within the curriculum – evidence, reasoning, interactions and competing claims. The coding scheme did not consider whether the activity or procedure aligned with the one described in the curriculum, but rather whether the instruction would support the argumentation goal. Figure 2 includes each teacher's total score for the quality of argumentation broken down by the four goals.

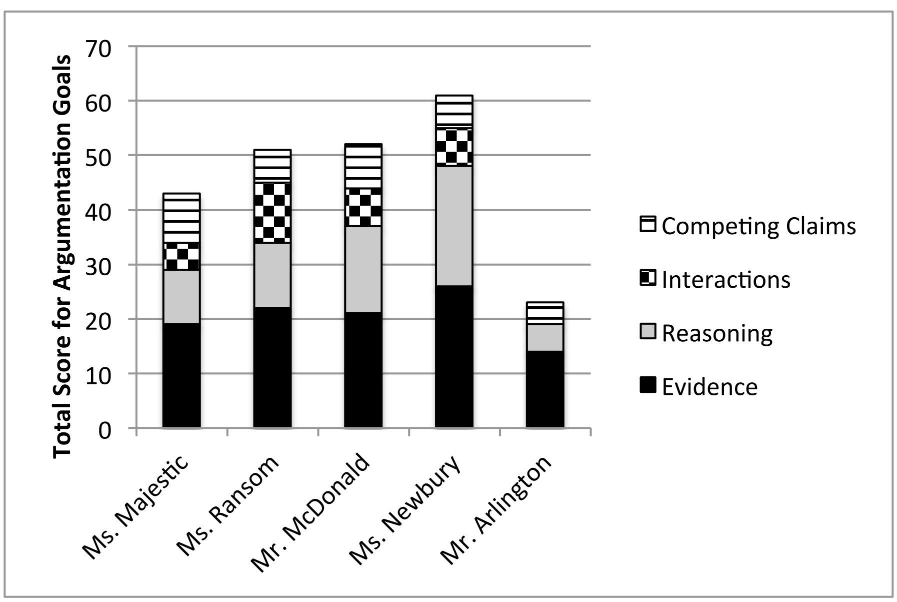

Figure 2: Fidelity to Goals

Similar to the previous coding scheme, Ms. Majestic, Ms. Ransom and Mr. McDonald had higher scores for argumentation. Each of these teachers supported all four argumentation goals, with a greater focus on the two structural goals, particularly students' use of evidence to support claims. Additionally, Mr. Arlington's score was the lowest, which was also the case for the *Fidelity to Procedure* coding. Mr. Arlington's instruction focused more on the structural elements of argumentation with the highest score for evidence; however, this was still considerably lower than the other teachers. In terms of argumentation as a dialogic process, he provided minimal support for competing claims and no support for student-to-student interactions.

The one teacher whose instruction received very different scores for *Fidelity to Procedure* versus *Fidelity to Goals* was Ms. Newbury. Ms. Newbury's score for *Fidelity to Goals* was the highest among the five teachers despite receiving a low score for *Fidelity to Procedure.* This suggests that while her enactment did not align closely with the procedures of the lessons, the changes she made did support high quality argumentation instruction. Consequently, in the next section we focus on one lesson to illustrate the differences in the changes she made to the curriculum compared to both Mr. McDonald, who represents the three teachers who aligned closely, and Mr. Arlington.

Example Microbiome Lesson 1.9: Writing a Scientific Argument

The second argumentation lesson videotaped was Lesson 1.9 in the Microbiome Unit, which was the first time the students were asked to write a scientific argument. Table 7 includes a summary of the five activities that were coded for in the lesson in terms of *Fidelity to Procedure* as well as the codes for each of the five teachers. Interestingly, the procedure for all five teachers aligned for the first activity in the lesson, Warm-Up: Student Writing, and the last activity in the lesson, Student Writing. However, the teachers' enactments differed in terms of the three activities within the middle of the lesson whose focus was on comparing two different arguments to prepare the students for their own writing. These three activities focused on making sense of the student warm-up and framing the writing task for the students.

Specifically, we will focus on the last of these activities, highlighted in Table 7. During this activity, the teacher presented and highlighted key aspects of Argument B, which was the stronger of the two arguments used in the lesson. The specific argumentation goal that this lesson targeted was Reasoning. In the curriculum, the Instructional Rationale for comparing the two arguments stated:

> *Often, students who are just beginning to learn about argumentation will simply list the evidence that supports the claim and may not include their thinking about why pieces of evidence support the claim. Modeling how to make the argument clearer will help students include this type of language in their own writing.*

Consequently, Argument A just listed the evidence while Argument B included reasoning that explained why the evidence supported the claim. The curriculum included suggestions around highlighting these differences for students before beginning their individual writing. For this section of the lesson, three of the teachers' enactments (Ms. Majestic, Ms. Ransom and Mr. McDonald) aligned with the curriculum while both Ms. Newbury and Mr. Arlington modified the lesson. However, they did so in very different ways illustrating why they received different scores for *Fidelity to Goals*.

Table 7: Fidelity to Procedure Codes for Microbiome Lesson 1.9

Activity	**Description of Activity**	**Ms. Majestic**	**Ms. Ransom**	**Mr. McDonald**	**Ms. Newbury**	**Mr. Arlington**
Warm-up: Student writing	Individual Students: Students individually write answering the question of which of two arguments (Argument A or B) they think is more persuasive	Aligned	Aligned	Aligned	Aligned	Aligned
Shared listening: Comparing arguments	Student pairs: Student pairs engage in shared listening where the first student answers – How are these arguments similar? And the second – How are these arguments different?	Modified	Aligned	Modified	Modified	Modified
Discussion: Comparing arguments	Full Class Discussion: Teacher leads a full class discussion about which of the two arguments is more persuasive.	Modified	Skipped	Aligned	Modified	Aligned
Teacher presentation: Highlight language of argument	Presentation: Teacher presentation highlights the language of argumentation pointing out the bold and underlined words in Argument B and pointing out the optional sentence starters for the student writing.	Aligned	Aligned	Aligned	Modified	Modified
Student writing	Individual Students: Students write an argument for the question - Why did a fecal transplant cure the patient who was infected with c. difficile bacteria?	Aligned	Aligned	Aligned	Aligned	Aligned

Mr. McDonald – Aligned. First we describe Mr. McDonald's enactment. He was one of the three teachers who aligned closely to the description in the curriculum in terms of the procedure. After leading a class discussion in which the students agreed that Argument B was more persuasive, he then highlighted key aspects of Argument B to focus the students on including reasoning in their arguments. He began by stressing that, "This is something you will want to think about when writing your argument today." He then projected an annotated version of Argument B from the teacher's version of the curriculum. This annotated version included bold and underlined words to highlight how Argument B included reasoning that connected the claim and evidence. Mr. McDonald pointed out the key language for the students to consider in the annotated version of the argument. He stated:

> *The other thing that it [points at Argument B] does really, really well, so much more than Argument A, is that it explains the evidence. Argument A is really just a list of things – This is why it happened – A, B, C, D E, F G. If we don't explain why, if we don't make those connections, we have a very difficult and less persuasive argument to read.*

As he discussed the reasoning in Argument B, Mr. McDonald pointed to the bold and underlined text in Argument B, which he had projected from his laptop. He talked about phrases such as "As you can see in the data," "the data show," and "Since antibiotics kill bacteria." He discussed how this type of language can help "bridge the gap between the evidence and claim…making those connections as you write." He then had students turn to page 43 in their student book which included those same sentence starters and phrases to support them in articulating their reasoning for why their evidence supported their claim. For this section of the lesson, in which Mr. McDonald highlighted key aspects of Argument B and shared the potential sentence starters, he received a score of 4 for argumentation quality around reasoning, because he provided a description (Level 1), explained a rationale (Level 1), as well as modeled and prompted using sentence starters (Level 2) to support students in including reasoning in their arguments.

Mr. Arlington – Modified with Low Argumentation Goals. In contrast, Mr. Arlington modified this section of the lesson and only received a score of 1 (i.e. low quality) for argumentation quality around reasoning because for modeling and prompting he just provided sentence starters (Level 1). Although Mr. Arlington also had the students complete the warm-up where they evaluated whether Argument A or B was more persuasive, the discussion about why Argument B was stronger did not highlight the structural differences in relation to reasoning between the two arguments. Mr. Arlington did not project the annotated version of Argument B as suggested in the curriculum nor did he highlight specific phrases or language in the argument suggested by the curriculum that made it stronger. In fact, he did not project any version of the argument, but just had students read them from their books. For example, he had one student read Argument A and then he said, "A little, a little bit jumbled I would say. You know what I mean. They are kind of just throwing stuff at us. Antibiotics kill bacteria. It is kind of just thrown in there." Although he critiques Argument A, it is not clear what aspects are lacking from the example. This differs from Mr. McDonald's discussion in which he talked about Example A as just including a list of evidence, but that it did not explain why the evidence supports the claim. Mr. Arlington then had a student read Argument B. After the student finished reading he said:

> *Good. So they [Argument B] give us a claim. They give us what they think. Ok. Where as the other one [Argument A] kind of just throws it at us. Alright. They are throwing us a bunch of facts. This one, plain and simply – It must be the case that the new infection was what made him sick. Ok. They give us an answer to the question. Ok. So we have been talking a lot about writing scientific arguments. So what I want you to do on the piece of paper I just gave you is to write a scientific argument.*

Mr. Arlington's description of Argument B suggests that the key difference is that Argument B includes a claim while Argument A did not include a claim and is "just throwing stuff at us". This is in contrast to the curriculum that describes the key difference was around reasoning. Consequently, Mr. Arlington received a lower score for *Fidelity to Procedure* because he did not project or discuss the phrases in Argument B that made it stronger. He also received a lower score for *Fidelity to Goals* because he never provided a description, rationale or model of reasoning. The one element he did follow, relevant to both conceptualizations of fidelity, is that he did refer students to the sentence starters in their student books, which provided them with a prompt for reasoning. He stated, "If you need sentence starter help you can look at page 43, 'the data shows', 'as you can see in the data.'" Consequently, he did receive a 1 for reasoning; however, he provided significantly less support for his students than Mr. McDonald. The modifications that Mr. Arlington made lowered the quality of the argumentation instruction.

Ms. Newbury – Modified with High Argumentation Goals. Finally, Ms. Newbury's discussion of the example also received a code of "Modified" because she made numerous changes. However, unlike Mr. Arlington, her alterations aligned with the overarching argumentation goals. Specifically, for the argumentation goal around reasoning she received a score of 5, which was closer to Mr. McDonald's score of 4 than Mr. Arlington's score of 1. Similar to Mr. McDonald, she used a number of strategies to support her students in reasoning including provided a description (Level 1), explained a rationale (Level 2), as well as modeled and prompted using sentence starters (Level 2). She received a higher score than Mr. McDonald because she provided a more in depth rationale than he, or the curriculum, provided. For example, in addition to talking about how including reasoning makes an argument more persuasive, she talked about how this feature of an argument was not just science specific, but an aspect that cut across disciplines. Specifically, Ms. Newbury pointed out that reasoning was similar to what the students had been learning in "Mr. Martin's" class and to what "Miss. Diaz has been teaching you in ELA or ESL."

Perhaps more interesting is that although Ms. Newbury was coded for high quality support for reasoning for modeling and using sentence starters, the examples she used were not the ones in the curriculum. As mentioned previously, she taught in an SEI classroom consisting of 6^{th} and 7^{th} graders who were all native Spanish speakers. Ms. Newbury altered the activities and supports in the curriculum and as such received a lower score for *Fidelity to Procedure* (see Table 7). For example, similar to the previous two teachers' enactments, her students decided that Argument B was more persuasive. Although she did project Argument B, she did not project the annotated version that Mr. McDonald presented from the curriculum that included bold and highlighted elements related to argument. Consequently, this activity was coded as "Modified." Instead, she projected the student version of this argument on her white board. As she discussed the example, she underlined words and phrases in the argument with a marker, words and phrases that were different than the ones targeted in the curriculum. The words in the curriculum that were in bold were phrases like “As you can see in the data”, “the data show that” and “Since antibiotics kill bacteria.” They tended to be words or phrases that were transitions or connections in the argument. Instead, Ms. Newbury underlined whole sentences and identified those sentences as reasoning for her students. For example, after underlining the sentence “*C. jejuni* causes food poisoning, so that was what was making him sick.” She stated:

> *Explains how your evidence supports your claim. Right? That’s what this does [pointing at underlined sentence on projected argument] So this is the reasoning. The reason we decided that this [Argument B]was more persuasive than the other one [Argument A] is because it has clearer reasoning.*

Her use of Argument B did include her modeling for her students strong reasoning and she focused on the connection between the claim and evidence. Consequently, although she changed the curriculum it still aligned with the intended argumentation goal.

In addition, she made another change to this lesson’s activity. Unlike Mr. McDonald and Mr. Arlington who referred the students to the sentence starters on page 43 to support their writing of an argument, Ms. Newbury developed her own sentence starters with her students. She asked her students how to start their ideas for the different structural elements of an argument – claim, evidence and reasoning. As the students shared their ideas, she typed them into a PowerPoint slide and projected the sentence starters for her students to use. These sentence starters were different than the curriculum, because they cut across the structural elements of an argument. Furthermore, although some were similar in language, others were different than in the curriculum. Consequently, this activity was coded as “Modified.” For example, for reasoning she said “For my sentences for reasoning – how can I talk about this?” As the students shared ideas, she typed some sentence starters that were the same as the curriculum such as “This means that” and “Therefore,…”, but other sentence starters were different such as “This shows” and “This makes me think that.” Overall, Ms. Newbury altered a number of procedures during the curriculum to provide her students with different linguistic supports, but her alterations maintained the argumentation goals.

Discussion

The results from this study suggest that conceptualizing and measuring FOI in various ways can present very different evaluations of a teacher’s enactment. Future research needs to explore the relationships between these FOI measurements with student outcomes. However, this work suggests that focusing on goals rather than more procedural elements of a curriculum may be more likely to predict student learning gains of science practices as our measurement of *Fidelity to Goals* offered an image of science instruction that aligns more closely with that advocated for in recent standards (Berland et al., in press). Three of the teachers received high scores for both procedure and goals while one teacher received low scores for both procedure and goals. However, the case of Ms. Newbury reveals an important distinction between the two FOI coding schemes. Her classroom context was unique in that she taught in an SEI classrooms. This suggests that different classroom contexts may offer important reasons to adapt procedural elements of curriculum.

A challenge of FOI is distinguishing between good teaching and good teaching prompted by the curriculum (O’Donnell, 2008). However, similar to Buxton and colleagues (2015), we argue that good teaching prompted by the curriculum is not going to look identical in all classrooms. Rather, good teaching is responsive to the ideas and needs of the students (Hammer, Goldberg & Fargason, 2012). Consequently, strong curriculum should provide teachers with resources that enable them to respond to and adapt to their students. We feel that FOI focused on curricular goals rather than procedures is one productive avenue for evaluating the impact of curriculum on teaching practices.

In addition, we argue that this work has implications for the future design of educative curriculum. Davis and Krajcik (2005) state that educative curriculum materials should include rationales for teachers to better understand the reasoning behind curricular recommendations. Our findings reiterate this as an essential aspect of educative curriculum. It may be less important for teachers to understand the exact procedure. Rather, the rationale may be more important in order to help them understand the goals behind that procedure to make appropriate modifications for their students, particularly around the demanding learning goals in the science practices.

Focusing on the rationale and goals could result not only in a shift in the design of the curriculum, but also a shift in how teachers use curriculum. Brown (2009) argues for the importance of considering teachers' pedagogical design capacity (PDC) or their abilities to mobilize instructional and teacher resources to better design instruction for the classroom. Teachers may need explicit learning experiences, such as through preservice courses and professional development, to support their development of PDC and to shift their views of curriculum use (Knight-Bardsley & McNeill, in press). However, our results also show that all adaptations, such as Mr. Arlington's, do not align with the goals of the curriculum. Consequently, as a community we need to think more about how to design educative curriculum that support teachers in responsive teaching where they respond to the needs of the students, but also meet the overarching goals of the curriculum.

References

Alozie, N. M., Moje, E. B. & Krajcik, J. S. (2010). An analysis of the supports and constraints for scientific discussion in high school project-based science. *Science Education*, 94, 395-427.

Berland, L. K. (2011). Explaining variations in how classroom communities adapt the practice of scientific argumentation. *Journal of the Learning Sciences*, 20, 625-664.

Berland, L. K., & Reiser, B. J. (2011). Classroom communities' adaptations of the practice of scientific argumentation. Science Education, 95(2), 191-216.

Berland, L. K., Schwarz, C. V., Krist, C., Kenyon, L., Lo, A. S., & Reiser, B. J. (in press). Epistemologies in practice: Making scientific practices meaningful for students. *Journal of Research in Science Teaching*.

Brown, M. W. (2009). The teacher-tool relationship: theorizing the design and use of curriculum materials. In J. Remillard, B. Herbel-Sisenham, & G. Lloyd (Eds.), *Mathematics teachers at work: Connecting curriculum materials and classroom instruction* (pp. 17-37). New York: Routledge.

Buxton, C. A., Allexsaht-Snider, M., Kayumova, S., Aghasaleh, R., Choi, Y. J., & Cohen, A. (2015). Teacher agency and professional learning: Rethinking fidelity of implementation as multiplicities of enactment. *Journal of Research in Science Teaching*, 52(4), 489-502.

Century, J., Rudnick, M. & Freeman, C. (2010). A framework for measuring fidelity of implementation: A foundation for shared language and accumulation of knowledge. *American Journal of Evcaluation, 31*(2), 199-218.

Cho, J. (1998, April). Rethinking curriculum implementation: Paradigms, models, and teachers' work. Paper presented at the annual meeting of the American Educational Research Association, San Diego, CA.

Davis, E. A. & Krajcik, J. (2005). Designing educative curriculum materials to promote teacher learning. *Educational Researcher*, 34(3), 3-14.

Echevarria, J., Vogt, M. E., & Short, D. J. (2008). *Making content comprehensible for English learners: The SIOP Model* (3rd ed.). Boston, MA: Allyn & Bacon.

Ford, M. J. (2012). A dialogic account of sense-making in scientific argumentation and reasoning. *Cognition and Instruction*, 30(3), 207-245.

Hammer, D., Goldberg, F., & Fargason, S. (2012). Responsive teaching and the beginnings of *energy* in a third grade classroom. *Review of Science, Mathematics and ICT Education, 6*(1), 51-72.

Herrenkohl, L. R. & Cornelius, L. (2013). Investigating elementary students' scientific and historical argumentation. *Journal of the Learning Sciences, 22*, 413-461.

Jiménez-Aleixandre, M. P. & Erduran, S. (2008). Argumentation in science education: An Overview. In S. Erduran & M. P. Jimenez-Aleixandre (Eds.). *Argumentation in science education: Perspectives from classroom-based research.* (pp. 3-28), Dordrecht: Springer.

Knight-Bardsley, A. M. & McNeill, K. L. (in press). Teachers' pedagogical design capacity for scientific argumentation. *Science Education.*

Lee, O., Penfield, R., Maerten-Rivera, J. (2009). Effects of fidelity of implementation on science achievement gains among English language learners. *Journal of Research in Science Teaching*, 46(7), 836-859.

McNeill, K. L., González-Howard, M., Katsh-Singer, R. & Loper, S. (2016). Pedagogical content knowledge of argumentation: Using classroom contexts to assess high quality PCK rather than pseudoargumentation. *Journal of Research in Science Teaching, 53*(2). 261-290.

McNeill, K. L. & Knight, A. M. (2013). Teachers' pedagogical content knowledge of scientific argumentation: The impact of professional development on k-12 teachers. *Science Education*, 97(6), 936-972.

McNeill, K. L. (2009). Teachers' use of curriculum to support students in writing scientific arguments to explain phenomena. *Science Education*, 93(2), 233-268.

McNeill, K. L., Lizotte, D. J, Krajcik, J., & Marx, R. W. (2006). Supporting students' construction of scientific explanations by fading scaffolds in instructional materials. *Journal of the Learning Sciences*, 15(2), 153-191.

McNeill, K. L., & Pimentel, D. S. (2010). Scientific discourse in three urban classrooms: The role of the teacher in engaging high school students in argumentation. *Science Education*, 94(2), 203-229.

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

Mowbray, C. T., Holter, M. C., Teague, G. B., & Bybee, D. (2003). Fidelity criteria: Development, measurement and validation. *American Journal of Evaluation, 24*(3), 315-340.

National Research Council. (2005). *Advancing scientific research in education.* Committee on Research in Education. Lisa Towne. Lauress L. Wise and Tina M. Winters, Editors. Center for Education, Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press.

National Research Council (2015). Guide to Implementing the Next Generation Science Standards. Committee on Guidance on Implementing the Next Generation Science Standards. Board on Science Education, Division of Behavioral and Social Sciences and Education, Washington, DC: The National Academies Press.

NGSS Lead States. (2013). *Next generation science standards: For states, by states*. Washington, DC: The National Academies Press.

O'Donnell, C. (2008). Defining, conceptualizing, and measuring fidelity of implementation and its relationship to outcomes in K-12 curriculum intervention research. *Review of Educational Research*, 78(1), 33–84.

Osborne, J. (2010). Arguing to learn in science: The role of collaborative, critical discourse. *Science*, 328, 463-466.

Regents of the University of California (2013a). Metabolism. Filed trial version of Middle School science unit developed by the Learning Design Group. Lawrence Hall of Science.

Regents of the University of California (2013b). Microbiome. Filed trial version of Middle School science unit developed by the Learning Design Group. Lawrence Hall of Science

Remillard, J. T. (2005). Examining key concepts in research on teachers' use of mathematics curricula. *Review of Educational Research*, 75(2), 211-246.

Sampson, V., & Blanchard, M.R. (2012). Science teachers and scientific argumentation: Trends in view and practice. *Journal of Research in Science Teaching*, 49(9), 112-1148.

Shulman, L. (1990). Foreward. In M. Ben-Peretz, The teacher–curriculum encounter: Freeing teachers from the tyranny of texts (pp. vii-ix). Albany: State University of New York Press.

Stake, R. E. (2000). Case Studies. In N. K. Denzin & Y. S. Lincoln (Eds.). Handbook of Qualitative Research. Thousand Oaks, CA, Sage.