International Journal of Science Education

ISSN: 0950-0693 (Print) 1464-5289 (Online) Journal homepage: http://www.tandfonline.com/loi/tsed20

Teachers' enactments of curriculum: Fidelity to Procedure versus Fidelity to Goal for scientific argumentation

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

To cite this article: Katherine L. McNeill, Lisa M. Marco-Bujosa, María González-Howard & Suzanna Loper (2018): Teachers' enactments of curriculum: Fidelity to Procedure versus Fidelity to Goal for scientific argumentation, International Journal of Science Education, DOI: 10.1080/09500693.2018.1482508

To link to this article: https://doi.org/10.1080/09500693.2018.1482508

Published online: 23 Jun 2018.

INTERNATIONAL JOURNAL OF SCIENCE EDUCATION
https://doi.org/10.1080/09500693.2018.1482508

Teachers' enactments of curriculum: *Fidelity to Procedure* versus *Fidelity to Goal* for scientific argumentation

Katherine L. McNeill [a], Lisa M. Marco-Bujosa[a*], María González-Howard [b] and Suzanna Loper [c]

[a]Lynch School of Education, Boston College, Chestnut Hill, MA, USA; [b]Department of Curriculum and Instruction, University of Texas at Austin, Austin, TX, USA; [c]Lawrence Hall of Science, University of California, Berkeley, Berkeley, CA, USA

ABSTRACT
Fidelity of implementation (FOI) has received attention in calls for funding and research; however, there are numerous ways of conceptualising and measuring this construct. We argue that this conceptualisation is important for recent reform efforts focused on science practices. Consequently, we explored FOI in the context of the enactment of a middle-school curriculum focused on one particular science practice, argumentation. We coded videos of five teachers' enactments of argumentation lessons using two different fidelity coding schemes. First, *Fidelity to Procedure* targeted teachers' adherence to the order and types of procedures. Second, *Fidelity to Goal* examined teachers' adherence to the overarching argumentation goals. This analysis resulted in case studies that illustrate distinct patterns in the teachers' curriculum enactments. One case in particular, Ms Newbury, received a low score for *Fidelity to Procedure*, but a high score for *Fidelity to Goal*. She altered procedures to provide her students, all of whom were English Language Learners, with different linguistic supports, but maintained the overarching argumentation goals. Consequently, we argue that FOI for goals may better capture whether teachers' enactments are supporting students in the science practices. Furthermore, the results suggest the importance of educative curriculum including rationales for the curricular goals.

ARTICLE HISTORY
Received 1 September 2017
Accepted 26 May 2018

KEYWORDS
Argumentation; curriculum; science practices

In the current standards-based accountability era, there is a focus on efficacy or effectiveness studies that measure how closely implementation aligns with the original curricular intervention (Century, Rudnick, & Freeman, 2010; National Research Council, 2005). This trend, combined with an interest in the scale-up of interventions, has resulted in a number of education researchers focusing on fidelity of implementation (FOI) in education settings (Desimone & Hill, 2017; Harris et al., 2015; Lee & Chue, 2013). The goal underlying such research on implementation is to determine the impact of curricular interventions on student learning outcomes. Consequently, FOI has received attention in calls for funding and research. For example, in the United States, the Institute of Education Science (IES)

CONTACT Katherine L. McNeill kmcneillbc.edu; kmcneill@bc.edu Lynch School of Education, Boston College, 140 Commonwealth Avenue, Chestnut Hill, MA 02467, USA
*Present address: Department of Education & Counseling, Villanova University, Villanova, PA, USA

and National Science Foundation (NSF) developed common guidelines for education research and development that include FOI in the research design of impact studies (e.g. efficacy, effectiveness and scale-up) (IES & NSF, 2013). The FOI language has been included in all of the requests for application for education research grants from IES since then and in some NSF programme solicitations.

Furthermore, even in research on curriculum less driven by accountability, there is still a desire to understand key aspects of implementation that can include FOI as multiplicities of enactment, which considers the variety of ways teachers may choose to enact a given lesson given a range of contextual and personal factors (Buxton et al., 2015). Consequently, there are numerous ways of conceptualising and measuring this construct, ranging from a strict adherence to the procedural elements included in a curriculum (O'Donnell, 2008) to a consideration of the role of the teacher in making appropriate instructional decisions to support student learning (Brown, 2009).

We argue that the conceptualisation and measurement of FOI are particularly important to consider in relation to recent science education reform efforts that include a focus on science practices (NRC, 2012). Science practices require a significant shift in science instruction away from students memorising final form ideas to demonstrating knowledge in use as students construct explanations and develop models about the natural world (Berland et al., 2016). The focus on student-driven learning is a key aspect of the reform efforts, as it encourages students' active engagement in these science practices, which requires the teacher to support students in ways that differ from previous science instruction (NRC, 2015).

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 also result in new challenges for FOI. Consequently, we were interested in exploring different conceptualisations of FOI in relation to the enactment of a middle-school science curriculum focused on argumentation.

Conceptual framework

Fidelity of implementation

FOI is a prevalent construct in education research and implementation studies that has been defined in a variety of ways (Century et al., 2010). Drawing on research in public health, Lee and colleagues (Lee, Penfield, & Maerten-Rivera, 2009) defined FOI as 'the determination of how well an innovation is implemented according to its original program design or as intended' (p. 837). Lee and Chue (2013) also draw on the public health research describing FOI as 'the closeness between the formal/perceived and the operational/experiential curricula' (p. 2510). FOI could also 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 curricular intervention has been implemented as planned in order to appropriately attribute changes in student learning to an educational innovation (Lee et al., 2009). Yet, there are few studies to guide researchers on how FOI of 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 *fidelity to procedure* such as the number, order and alignment of methods prescribed in a curriculum (O'Donnell, 2008). In a literature review, O'Donnell (2008) emphasised a description of FOI as fidelity to structure (i.e. adherence, duration) and fidelity to process (i.e. quality of delivery, programme differentiation). Seraphin et al. (2017) used this perspective on FOI, particularly in terms of teachers' adherence to curricular guidelines, in their study with 28 teachers using aquatic science materials. They found that students had greater content gains when their teachers adhered more closely to the activities. Similarly, Mowbray, Holter, Teague, and Bybee (2003) emphasised fidelity criteria in terms of structure (i.e. the framework for service delivery) and process (i.e. the way in which services are delivered). A number of studies also define FOI in terms of instructional quality (O'Donnell, 2008). For example, Lee et al. (2009) operationalised FOI in terms of the quality of instructional delivery, dose or exposure and participant responsiveness. However, Lee et al.'s (2009) results indicate FOI, as measured by teachers' self-reports and classroom observations, had no significant effects on students' science achievement gains. The authors conjectured this result may have been due to measurement errors or their conceptualisation of FOI.

These measures of FOI for curriculum use do not take into consideration the adaptive and reactive aspects of teaching practice. As Shulman (1990) noted, 'While curriculum might be the backdrop for teaching, the two are not to be confused' (p. vii). Therefore, a new conceptualisation of FOI is needed in which the role of the teacher as curriculum user 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 reconceptualisation 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 that teachers have an important role within a unique instructional context, as they interpret, adapt and implement the curriculum. Teachers need to be responsive to the needs and ideas of their students (Hammer, Goldberg, & Fargason, 2012). Therefore, this perspective assumes the teacher is an active designer of curriculum rather than solely an implementer. Similarly, Buxton et al. (2015) extrapolated on the teacher–curriculum relationship by theorising 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 et al., 2015, p. 499). This suggests multiple different implementations could still align with high FOI, rather than only one 'correct' enactment.

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 conceptualisation and measurement focused on *goals* rather than procedural elements. *Fidelity to Goal* focuses on the alignment with the overarching goals in the curriculum rather than the prescribed methods in a lesson plan. This aligns with the work of Davis and Krajcik (2005) who argue that educative curriculum that support teacher learning should provide the rationale behind curricular decisions to better inform teachers' enactments. Furthermore, previous research focusing on teachers' adaptations to curricula suggests that they can oversimplify complex science practices, like argumentation, if they do not understand the underlying epistemic goals (McNeill, González-Howard, Katsh-Singer, & Loper, 2017).

To focus on goals, measuring FOI for curriculum use would involve specifying the critical components and processes of the curriculum's theory (Century et al., 2010). For example, in Debarger et al.'s (2017) work on purposeful science curriculum adaptation, they identified adaptation goals focused on classroom discourse norms and high leverage talk moves as being essential for the curriculum enactment. Furthermore, this view problematises the more traditional perspective on fidelity, as it considers that 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 if they align with the overarching goals. However, focusing on *Fidelity to Goal* may not capture other important aspects of teachers' curriculum enactment. Consequently, we investigated two potentially different conceptualisations of FOI from the literature – *Fidelity to Procedure* versus *Fidelity to Goal*. Specifically, we were interested in how these different conceptualisations relate to science practices in recent reform efforts.

Argumentation

Considering recent reform documents in science education, the conceptualisation of fidelity could have important implications for how to best support teachers and students in science practices, such as scientific argumentation. Argumentation is a potentially productive focus of this work because teachers' enactment of curriculum addressing argumentation varies greatly (Berland & Reiser, 2011; Herrenkohl & Cornelius, 2013). Furthermore, argumentation requires a different classroom culture and norms than traditional science classrooms (Osborne, Erduran, & Simon, 2004).

Argumentation in science has been described and assessed in a variety of ways (Sampson & Clark, 2008). Similar to others (Jiménez-Aleixandre & Erduran, 2008), we define argumentation in terms of both a *structural* and *dialogic* focus (McNeill, González-Howard, Katsh-Singer, & Loper, 2016). The structure of a scientific argument consists of a claim that is supported by both evidence and scientific reasoning (McNeill, Lizotte, Krajcik, & Marx, 2006). Evidence is scientific data, such as observations or measurements, about the natural world. Reasoning articulates why the evidence supports the claim using disciplinary core ideas (DCI). 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 emphasises a classroom goal of collaboratively making sense of phenomena and convincing peers of other potential claims (Berland & Reiser, 2011).

Teachers can have difficulty with both the structural and dialogic elements of this science practice (McNeill & Knight, 2013). This difficulty informed our identification of the key argumentation goals in this study. For the structural elements, teachers can oversimplify or struggle to understand what counts as evidence and reasoning. For example, Sampson and Blanchard (2012) found teachers did not provide solid evidence and reasoning in support of a claim, indicating teachers wrestled with these concepts. Teachers can also have difficulty knowing what counts as evidence in science and can struggle to use data from experiments (Crippen, 2012). In a study in which middle-school teachers enacted a science curriculum using the claim, evidence and reasoning structure, some teachers turned the structure into an algorithm, removing the sensemaking from the reasoning and instead only requiring a definition of a science term (McNeill, 2009). Consequently, in terms of structure, we view teachers' abilities to support students in both (1) using high-quality evidence (i.e. evidence) and (2) using scientific ideas to explain the link between the evidence and claim (i.e. reasoning) as essential curricular goals for enactment.

In addition to the structural elements, teachers can experience challenges with argumentation as a dialogic process. For example, in a study of curriculum supports for leading dialogic discussions in high school science, Alozie, Moje, and Krajcik (2010) found teachers altered lessons focused on student-to-student interactions into more traditional teacher-led discussions. Evagorou and Dillon (2011) discuss similar results around the enactment of a middle-school science curriculum in which one teacher adapted the argumentation lessons to follow a pedagogy of transmission in which the lesson was dominated by teacher talk. Consequently, one key curricular goal around the dialogic process is supporting students in building off of each other's ideas (i.e. interactions). When engaged in professional development focused on argumentation, teachers' instruction can incorporate more dialogic argumentation; however, some elements, such as supporting students in listening and talking, can be easier for teachers than other elements, such as engaging students in critique and counterarguments (Simon, Erduran, & Osborne, 2006). This highlights a final curricular goal for argumentation, around supporting students in critiquing and evaluating competing claims (i.e. competing claims).

Given these challenges with argumentation, the present study was designed to explicitly explore FOI in the context of a curriculum supporting this science practice. Specifically, we investigated the following research questions:

(1) What variation exists in teachers' enactments of a curriculum focused on argumentation considering *Fidelity to Procedure?*
(2) What variation exists in teachers' enactments of a curriculum focused on argumentation considering *Fidelity to Goal?*
(3) What similarities and differences exist in the two fidelity measures?

Methods

We used a multiple case study approach to examine differences in teachers' enactment of an argumentation curriculum (Yin, 2013). Specifically, we video recorded the teachers' enactment of key lessons and then used two different FOI coding schemes – *Procedure* and *Goal* – to examine the similarities and differences.

Curricular context

This study took place during the pilot of a life science curriculum for middle-school students. 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). Both curriculum units aligned with the three-dimensional science standards in the *Next Generation Science Standards* (NGSS Lead States, 2013), which include a focus on DCI, crosscutting concepts and science practices. The curriculum also incorporated a multimodal approach in which students engage in doing, talking, reading and writing about science (Pearson, Moje, & Greenleaf, 2010). In terms of key science concepts, the *Microbiome* unit focused on ideas about cells and population dynamics within the context of the human microbiome. The *Metabolism* unit focused on how body systems work together to bring needed molecules to the cells. Both units targeted key science practices, including engaging in argumentation from evidence. In addition to a web-based teacher's guide and other digital resources, teachers were provided with kits that included physical manipulatives for student investigations, as well as student notebooks.

The educative curriculum was designed to support teachers' abilities to incorporate scientific argumentation into their instruction. The curriculum was educative in that it was developed to support teacher learning (Davis & Krajcik, 2005) about both the structural and dialogic aspects of argumentation. The curriculum defined the structure of an argument to consist of a claim supported by evidence and scientific reasoning (McNeill et al., 2006). It also addressed the dialogic aspects of this science practice, emphasising argumentation as a social process in which students construct, evaluate and revise arguments through interaction with their classmates (Berland & Reiser, 2011). Educative supports were provided through both text and multimedia formats, such as videos, which offered teachers real examples of what the structural and dialogic aspects of argumentation looked like in practice as well as rationales for the argumentation activities and instructional strategies.

Participants

Across the United States, 20 teachers enacted the pilot curriculum. The participants in this study included 5 teachers selected based on their vicinity to the 2 research teams, which enabled the collection of videos of classroom enactment (Table 1). The 5 teachers had a range of teaching experience from a second-year teacher to over 20 years of teaching experience, and a range of degrees in education and science. Additionally, teachers were asked if they had ever participated in professional development around argumentation. The teachers reported attending between one and three argument trainings.

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 sixth- and seventh-grade students who were all native Spanish speakers with beginning English proficiency levels who had recently immigrated to the United States from Central or South America.

Table 1. Teachers' backgrounds.

Teacher	Science credential	Highest degree education	Highest degree science	Years Teaching Experience	# of argument trainings
Ms Majestic	None	MA	BA	20 or more	1
Ms Ransom	MS/HS	MA	BA	20 or more	1
Mr McDonald	MS/HS	MA	BA	6–10	2 or 3
Ms Newbury	MS/HS	MA	None	6–10	2 or 3
Mr Arlington	MS/HS	BA	BA	2	1

Note: MS/HS = middle school or high school science credential.

Data collection

To examine teachers' fidelity of argumentation, we selected six lessons focused on argumentation (Table 3). These six lessons varied with respect to the type of activities included as well as the argument goals addressed. For example, 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. 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, and *Metabolism* Lesson 2.10, in which students engaged in a class discussion, called a science seminar. Across the lessons, there was a range of activity structures (e.g. card sort, writing arguments, science seminar) targeting the argumentation goals.

Data analysis

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, & Saldaña, 2013).

Fidelity to Procedure

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 with one of three codes: aligned, modified or skipped (Table 4).

Table 2. School and class 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 P	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

Table 3. Summary of argumentation lessons.

Curriculum unit	Lesson focus	Lesson description
Microbiome	1.6: Identifying Claims and Evidence	• Introduce claims and evidence in a scientific argument about the effect of antibiotics. • Observe the effect of antibiotics on an agar plate, gather evidence. • Card sort to identify evidence supporting claims.
	1.9: Writing a Scientific Argument	• Review and discuss two arguments; how are they similar or different, and which is most persuasive? Discuss organisation and connections. • Highlight language of argumentation with sentence starters • Write scientific argument based on evidence from card sort.
	1.10: Presenting a Scientific Argument	• Discuss claims and evidence to develop a complete story about why the faecal transplant was successful. • Students work in groups to share their argument in favour of their assigned claim. • Plan and create video responses.
Metabolism	1.12: Writing an Argument	• Students prepare to write their own argument by reviewing the purpose and question. A template is provided. Justification is introduced. • Students write arguments.
	2.8: Using simulation to gather evidence	• 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.
	2.10: Science Seminar	• Prepare for the science seminar by writing their best ideas on the science seminar evidence sheet. • Review the purpose and structure of the science seminar. • Students participate in a Science Seminar.

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.

Fidelity to Goal

The second coding scheme focused on *Fidelity to Goal*. We coded each lesson for the quality of argumentation instruction centring on four argumentation goals. As described above, these four goals were identified based upon previously documented teacher challenges in supporting students with this complex science practice (McNeill et al., 2016). The first two argumentation goals focused on the structure of an argument: (1A) use of high-quality evidence (Evidence) and (1B) use of scientific ideas to explain the link between the evidence and claim (Reasoning). The second two goals emphasised 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, (2) Teachers' rationale, (3) Teacher models and prompts and (4) Student

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 followed 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 include the Do Now activity.

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 16 codes (4 goals each with 4 codes), we rated each lesson as 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 5 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 Goal*. 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.

Case studies

We used the two FOI coding schemes, *Procedure* and *Goal*, 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 et al., 2013) across both teachers and 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 around the two conceptualisations of FOI. Specifically, these case studies detailed the quality of each teacher's argumentation instruction in terms of the alignment with the four argumentation goals 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) about their enactment of the argumentation lessons. These case studies ranged from 9 to 11 pages single-spaced. Each case study was then read by two other members of the research team (the second and third authors), 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. After developing the five cases, we then read

Table 5. Coding scheme for *Fidelity to Goal* 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 or to decide which is the strongest among claims. (3) 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 and prompts for evidence. Examples could include: • Teacher models by providing an example of high-quality or low-quality evidence. • Teacher provides prompts (e.g. Remember to include evidence to support your claim.) • Teacher uses questions to prompt (e.g. What is your evidence?)	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?'
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.

across them to better understand the patterns that emerged from graphing the FOI results by teacher. Doing so offered insight into why the FOI scores differed for *Procedure* versus *Goal* across teachers.

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 of the case study teachers.

Fidelity to Procedure

In terms of procedure, there was variation across the five teachers' enactments of the argumentation lessons (Figure 1). Three teachers had high fidelity to procedure, with around 80% of the activities adhering 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.

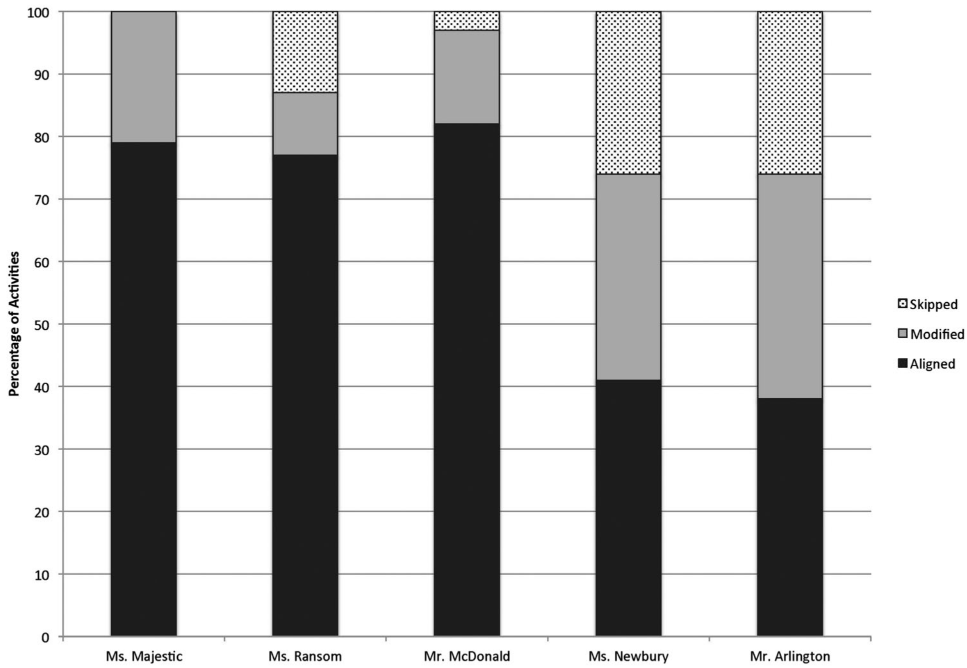

Figure 1. Fidelity to procedure.

Fidelity to Goal

For goals related to the structure and dialogic aspects 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.

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. Consequently, these three teachers received high scores for both *Fidelity to Procedure* and *Fidelity to Goal*. 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. Thus, his scores were low for both *Fidelity to Procedure* and *Fidelity to Goal*.

The one teacher whose instruction received very different scores for *Fidelity to Procedure* versus *Fidelity to Goal* was Ms Newbury. Ms Newbury's score for *Fidelity to Goal* was the highest among the five teachers despite receiving a low score for *Fidelity*

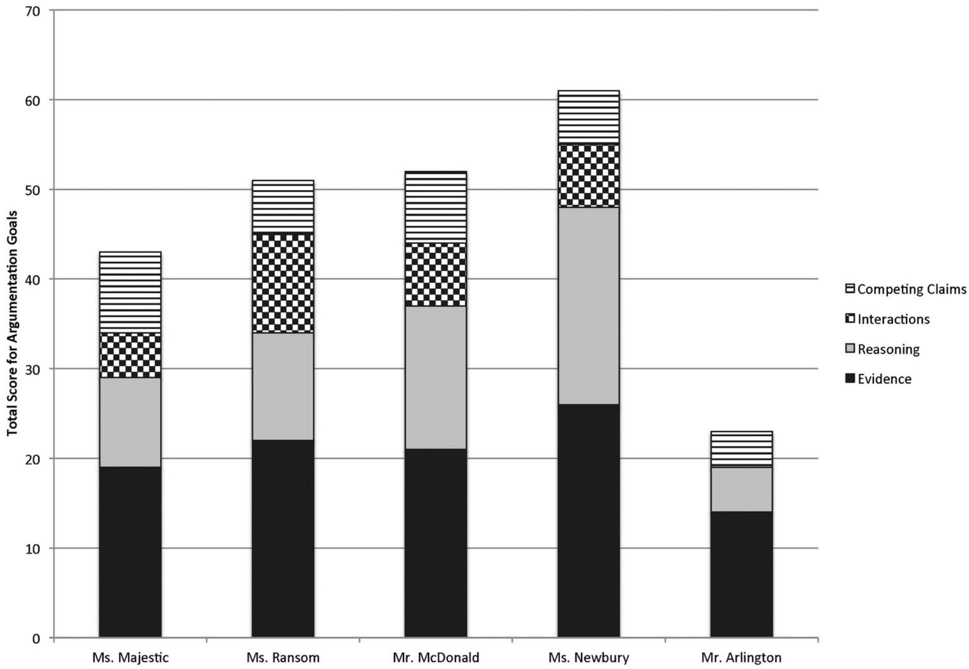

Figure 2. Fidelity to goal.

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 were high for both FOI scores, and Mr Arlington, who was low for both FOI scores.

Example Microbiome *lesson 1.9: writing a scientific argument*

The second argumentation lesson videotaped was *Microbiome* Lesson 1.9, which was the first time students were asked to write a scientific argument. Table 6 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. All five teachers' procedure 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.

Specifically, we will focus on the last of these activities, shaded in grey in Table 6. During this activity, the procedure in the curriculum suggested that the teacher present and highlight 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:

Table 6. 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 which of two arguments they think is more persuasive	Aligned	Aligned	Aligned	Aligned	Aligned
Shared listening: Comparing arguments	Student pairs: Student pairs engage in shared listening.	Modified	Aligned	Modified	Modified	Modified
Discussion: Comparing arguments	Full Class Discussion: Teacher leads a 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 and starters for the student writing.	Aligned	Aligned	Aligned	Modified	Modified
Student writing	Individual Students: Students write an argument for the question – Why did a faecal transplant cure the patient who was infected with c. difficile bacteria?	Aligned	Aligned	Aligned	Aligned	Aligned

> 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 Goal*.

Mr McDonald – High for Fidelity to Procedure and Goal

Mr McDonald was one of the three teachers who aligned closely to the curriculum in terms of 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 modelled and prompted using sentence starters (Level 2) to support students in including reasoning in their arguments. Consequently, his enactment received high codes for both procedure and goal.

Mr Arlington – low for Fidelity to Procedure and Goal

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 modelling 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 aloud 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 nor discuss the phrases in Argument B that made it stronger. He also received a lower score for *Fidelity to Goal* because he never provided a description, rationale or model of reasoning. The one element he did follow, relevant to both conceptualisations 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. Thus, the modifications that Mr Arlington made to the procedures also lowered the quality of the argumentation instruction because the argumentation goals were also not evident in his instruction.

Ms Newbury – low for Fidelity to Procedure*, but high for* Fidelity to Goal

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 goal of the lesson. 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 modelled 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.'

Although Ms Newbury was coded for high-quality support for reasoning for modelling 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 sixth- and seventh-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 6). 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 modelling 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 students to use. Although some of the student-generated sentence starters 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, particularly in relation to the structure. Consequently, her enactment of the curriculum received a low score for procedure, but it received a high score for the intended argumentation goals.

Discussion

The results from this study suggest that conceptualising and measuring FOI in various ways can present different evaluations of a teacher's curriculum enactment. Future research needs to explore the relationships between these FOI measurements and student outcomes. However, this work suggests that focusing on goals rather than more procedural elements of a curriculum may be more productive for curricula targeting science practices, such as argumentation. We discuss implications both in terms of measuring FOI, but also in terms of the design of future educative curriculum to support these essential goals.

Measuring FOI

A fundamental part of design work is often asking whether or not an intervention worked, which may include an evaluation of FOI (Century et al., 2010). However, similar to Buxton et al. (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 et al., 2012). This can appear to be at odds with a typical fidelity perspective that focuses on whether the implementation aligned with the designers' intentions (Century et al., 2010). However, if the designers' intentions are to support particular goals and not a set procedure, a goal-oriented perspective on FOI can be appropriate.

Teachers implement new ideas, such as argumentation instruction, differently in their classrooms (Simon et al., 2006). In our study, three of the teachers received high scores for both fidelity to procedure and fidelity to goal while one teacher received low scores for both conceptualisations of fidelity. 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 classroom. This suggests that different classroom contexts may offer important reasons to adapt procedural elements of curriculum to support students in specific goals.

Specifically, science practices offer both opportunities and challenges for culturally and linguistically diverse students because they are language intensive (Lee, Quinn, & Valdés, 2013). Teachers may need to use additional language supports to help English Language Learners in argumentation (González-Howard, McNeill, Marco-Bujosa, & Proctor, 2017). Consequently, strong curriculum should provide teachers with resources that enable them to respond to and adapt to their particular students. High FOI should be conceptualised as 'multiplicities of enactment' and not as only one appropriate procedure (Buxton et al., 2015, p. 499). Teachers' enactments should align with the target goals, but also meet the unique needs of their students. In this study, those target goals focused on argumentation; however, this focus will vary depending on the curriculum. For example, in Debarger et al.'s work (2017), they identified the adaptation goals of classroom discourse norms and high leverage teacher talk moves as essential for the enactment of a middle-school earth science unit. Identifying appropriate targets for *Fidelity to Goal* may depend on both the design principles of a curriculum as well as common challenges with classroom enactments. Determining appropriate goals may be more challenging than a procedural approach. However, we feel that FOI focused on curricular goals rather than procedures is one productive avenue for evaluating the impact of curriculum on teaching for the science practices. Future research should explore *Fidelity to Goal* with other curriculum materials, as well as investigate the relationship with student learning.

Designing educative curriculum

Furthermore, we argue that this work has implications for the future design of educative curriculum materials. Davis and Krajcik (2005) state that educative curriculum materials should include rationales for teachers to better understand the reasoning behind curricular recommendations. Educative curricula should include teacher supports for the *what* (i.e. What is the science practice?) the *how* (i.e. What is a strategy to engage students in this science practice?) and the *why* (Why does this strategy help support students in the science practice?) to integrate science practices in their classroom instruction (Bismack, Arias, Davis, & Palincsar, 2014). Our findings reiterate the rationale or 'why' as an essential aspect of educative curricula. More specifically, the rationale should not just be about why a science practice is important, but also why particular activities and strategies can help support students in the science practice. Teachers may choose to change a procedure, but if they understand the overarching goals those changes may look different. The rationale may be important to help them understand the goals behind that procedure and make appropriate modifications for their students, particularly around the demanding learning goals in the science practices.

Focusing on rationales and teachers' productive adaptations of curricula align with the argument made by Brown (2009) for the importance of developing teachers' pedagogical design capacity (PDC). Brown described PDC as the ability to mobilise instructional and teacher resources to better design instruction for the classroom. Considering teachers' PDC in designing curriculum materials suggests that curricula should support teachers in making productive adaptations, rather than providing one set procedure to follow. With text-based curriculum, this could be potentially challenging or overwhelming as it could require providing multiple variations of one lesson. However, technology offers new avenues for curriculum design that can not only provide images of classroom instruction, but also be more adaptive and provide teachers with different information depending on the backgrounds and needs of both the teacher and their students (Loper, McNeill, & González-Howard, 2017).

Implications

FOI can be a productive tool for evaluating the effectiveness of a curriculum or other interventions (Century et al., 2010). However, we argue that for recent reform efforts in science a focus on FOI for goals compared to procedures may better capture whether teachers' enactments are supporting students in the science practices. Because the science practices represent a shift in instruction in which students are constructing and critiquing knowledge (NRC, 2015) and because they are language intensive (Lee et al., 2013), students' backgrounds and experiences can impact the supports students need. As a community, we need to think more about how to measure FOI and design educative curricula that support teachers in responsive teaching in which they respond to the needs of the students, but also meet the overarching goals of the curriculum.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by the National Science Foundation from the Division of Research on Learning in Formal and Informal Settings under the Constructing and Critiquing Arguments in Middle School Science Classrooms: Supporting Teachers with Multimedia Educative Curriculum Materials grant [grant number DRL-1119584].

ORCID

Katherine L. McNeill http://orcid.org/0000-0003-3673-6637
María González-Howard http://orcid.org/0000-0003-3575-3937
Suzanna Loper http://orcid.org/0000-0002-7893-0801

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

Berland, L. K., Schwarz, C., Krist, C., Kenyon, L., Lo, A. S., & Reiser, B. J. (2016). Epistemologies in practice: Making scientific practices meaningful for students. *Journal of Research In Science Teaching*, *53*(7), 1082–1112.

Bismack, A. S., Arias, A. M., Davis, E. A., & Palincsar, A. S. (2014). Connecting curriulum materials and teachers: Elementary science teachers' enactment of a reform-based curricular unit. *Journal of Science Teacher Education*, *25*, 489–512.

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 Evaluation*, *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.

Crippen, K. J. (2012). Argument as professional development: Impacting teacher knowledge and beliefs about science. *Journal of Science Teacher Education*, *23*(8), 847–866.

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

Debarger, A. H., Penuel, W. R., Moorthy, S., Beauvineau, Y., Kennedy, C., & Boscaddin, C. K. (2017). Investigating purposeful science curriculum adaptation as a strategy to improve teaching and learning. *Science Education*, *101*, 66–98.

Desimone, L. M., & Hill, K. L. (2017). Inside the black box: Examining mediators and moderators of a middle school science invention. *Educational Evaluation and Policy Analysis*, *39*(3), 511–536.

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

Evagorou, M., & Dillon, J. (2011). Argumentation in the teaching of science. In D. Corrigan, J. Dillon, & R. Gunstone (Eds.), *The professional knowledge base of science teaching* (pp. 189–204). New York: Springer.

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

González-Howard, M., McNeill, K. L., Marco-Bujosa, L. M., & Proctor, C. P. (2017). Does it answer the question or is it French fries?: An exploration of language supports for scientific argumentation. *International Journal of Science Education*, *39*(5), 528–547.

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.

Harris, C. J., Penuel, W. R., D'Angelo, C. M., DeBarger, A. H., Gallagher, L. P., Kennedy, C. A., … Krajcik, J. S. (2015). Impact of project-based curriculum materials on student learning in science: Results of a randomized controlled trial. *Journal of Research in Science Teaching*, *52*(10), 1362–1385.

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

Institute of Education Sciences and National Science Foundation. (2013). *Common guidelines for education research and development*. https://ies.ed.gov/pdf/CommonGuidelines.pdf

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.

Lee, Y.-J., & Chue, S. (2013). The value of fidelity of implementation criteria to evaluate school-based science curriculum innovations. *International Journal of Science Education*, *35*(15), 2508–2537.

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.

Lee, O., Quinn, H., & Valdés, G. (2013). Science and language for English language learners in relation to next generation science standards and with implications for common core state standards for English language arts and mathematics. *Educational Researcher*, *42*, 223–233.
Loper, S., McNeill, K. L., & González-Howard, M. (2017). Multimedia educative curriculum materials (MECMs): Teachers' use of MECMs to support argumentation. *Journal of Science Teacher Education*, *28*(1), 36–56.
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., 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., González-Howard, M., Katsh-Singer, R., & Loper, S. (2017). Moving beyond pseudoargumentation: Teachers' enactments of an educative science curriculum focused on argumentation. *Science Education*, *101*(3), 426–457.
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., 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.
Miles, M., Huberman, A. M., & Saldaña, J. (2013). *Qualitative data analysis: A methods source book* (3rd edition). Thousand Oaks, CA: Sage.
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. (2012). *A framework for k-12 science education: Practices, crosscutting concepts and core ideas*. 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.
Osborne, J., Erduran, S., & Simon, S. (2004). Enhancing the quality of argumentation in school science. *Journal of Research in Science Teaching*, *41*(10), 994–1020.
Pearson, P. D., Moje, E., & Greenleaf, C. (2010). Literacy and science: Each in the service of the other. *Science*, *328*, 459–463.
Regents of the University of California. (2013a). *Metabolism: Filed trial version of middle school science unit developed by the learning design group*. Berkeley: 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*. Berkeley: 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), 1122–1148.

Sampson, V., & Clark, D. B. (2008). Assessment of the ways students generate arguments in science education: Current perspectives and recommendations for future directions. *Science Education*, *92*, 447–472.

Seraphin, K. D., Harrison, G. M., Philippoff, J., Brandon, P. R., Nguyen, T. T. T., Lawton, B. E., & Vallin, L. M. (2017). Teaching aquatic science as inquiry through professional development: Teacher characteristics and student outcomes. *Journal of Research in Science Teaching*, *54*(9), 1219–1245.

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

Simon, S., Erduran, S., & Osborne, J. (2006). Learning to teach argumentation: Research and development in the science classroom. *International Journal of Science Education*, *28*(2–3), 235–260.

Stake, R. E. (2000). Case studies. In N. K. Denzin, & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (pp. 435–453). Thousand Oaks, CA: Sage.

Yin, R. K. (2013). *Case study research: Design and methods*. Thousand Oaks, CA: Sage.