Mathematical and Scientific Argumentation in PreK-12

A Cross-Disciplinary Synthesis of Recent DRK-12 Projects

Eben Witherspoon, David Miller, Isabella Pinerua, and Dean Gerdeman

APRIL 2022

American Institutes for Research® | AIR.ORG

Advancing Evidence.
Improving Lives.

Mathematical and Scientific Argumentation in PreK-12

A Cross-Disciplinary Synthesis of Recent DRK-12 Projects

Eben Witherspoon, David Miller, Isabella Pinerua, and Dean Gerdeman

APRIL 2022

AIR® Headquarters
1400 Crystal Drive, 10th Floor
Arlington, VA 22202-3289
+1.202.403.5000 | **AIR.ORG**

17499_03/22

Acknowledgments

This project was supported by National Science Foundation Grant 1813777 to the American Institutes for Research (AIR).

AIR would like to thank Katherine McNeill and Jennifer Knudsen for serving as external advisors for this study. We also thank Sonica Dhillon, Danielle Ferguson, Jonathan Margolin, and Joe Taylor of AIR for their reviews of earlier drafts of this paper. Iszy Licht assisted with compiling figures and other supporting material.

Contents

Exhibits

Executive Summary

Context and Our Focus

Argumentation is a core disciplinary practice in mathematics and science that is important for both content understanding and everyday reasoning. In this report, we investigate how the National Science Foundation's (NSF's) recent research investments have advanced understanding and supported the development of interventions that improve the teaching and learning of argumentation in mathematics and science education. In the 5 years spanning 2011 to 2015, NSF's Discovery Research PreK–12 (DRK-12) program funded or cofunded 23 projects relating to argumentation, with more than $40 million awarded.

Findings

These 23 DRK-12 projects primarily focused on argumentation in high school and middle school and applied correlational/observational and longitudinal methods (rather than quasi-experimental or experimental methods), often reporting on the design and implementation of technological supports for the teaching and learning of argumentation. Our synthesis of empirical findings focused on how these projects studied both teacher- and student-facing interventions that improved the teaching and learning of argumentation, as well as naturalistic observations of argumentation in classroom settings that helped inform the design and development of future argumentation interventions:

- **Student Learning Trajectories in Argumentation.** A large number of empirical studies in this synthesis provided evidence that interventions that make the structure and components of an argument explicit for students, or that supported the facilitation of student-student discourse, both support students' learning of argumentation. A large proportion of interventions developed in projects from this period also featured technological tools that helped support these aspects of student learning.
- **Teacher Beliefs and Practices for Teaching Argumentation.** Compared with studies of student learning, relatively few empirical studies examined teacher learning through professional development. A mix of intervention studies and naturalistic observations suggested that certain teacher moves, such as using both closed- and open-ended questions and involving multiple students, may support students' learning of argumentation. A smaller number of teacher-focused studies reported on the role of teacher beliefs about how students learn argumentation.
- **Disciplinary Similarities and Differences in Mathematical and Scientific Argumentation.** Studies of both mathematical and scientific argumentation showed tools that help make the structure of an argument explicit by organizing student ideas and supporting student-

student discourse were effective for improving student argumentation. Differences between disciplines were in how argumentation was defined (e.g., evidence in science, justification in mathematics), differences in the study design (e.g., slightly more student interventions in science) and in the grade levels examined (e.g., fewer preK or early elementary studies in science).

- **Supporting Argumentation With Students From Marginalized Groups.** In the set of studies synthesized here, relatively few focused on supporting argumentation instruction with students from marginalized groups, beyond reporting data on participant characteristics. Those studies that did explore findings by demographic group in greater depth focused on understanding how contextual factors like English immersion courses and teacher beliefs about students abilities to learn argumentation presented barriers and supports to students' opportunities to learn.

Implications for the DRK-12 Portfolio

Our discussion of these projects' findings considers opportunities and priorities for future research, including potential gaps in the DRK-12 portfolio. In particular, new studies that investigate and compare the effectiveness of argumentation interventions in both math and science could help us understand what aspects of argumentation are specific to a particular discipline. The evidence from this synthesis suggests that supporting teachers in facilitating productive student discourse, although less common in mathematics, shows some evidence of improving argumentation in both subjects. Additional studies are needed to understand the nature of argumentation in preK and early elementary science classrooms to inform the design of interventions and pedagogical supports for teachers of younger students. Finally, more research is needed to better understand barriers in the teaching and learning of argumentation that may disproportionately impact students from historically marginalized groups. Drawing on the literature about culturally responsive practices in science, technology, engineering, and mathematics (STEM) could provide a fruitful avenue for projects in this area. Pursuing these relatively underexplored areas of argumentation could help support DRK-12's mission to significantly enhance preK–12 STEM teaching and learning.

Why This Topic?

The general practice of argumentation occurs in our everyday lives as the processes of understanding and evaluating others' claims, determining the validity of evidence, and reaching one's own reasoned conclusions. In this broad sense, engaging in argumentation is central to students' general education and preparing them for participation in the workforce and democratic society (Newton et al., 1999). In addition, it is considered an important part of the development of the communication and critical thinking skills applicable across many aspects of everyday life (Jiménez-Aleixandre & Erduran, 2007; Knudsen et al., 2017).

More specifically, scientific or mathematical argumentation is based on the beliefs and practices that are valued within the scientific or mathematical communities. Although each discipline defines argumentation somewhat differently, standards documents in both science and mathematics agree that fundamental components of argumentation, such as establishing the validity of claims based on reasoning and evidence, are core disciplinary practices and central to the work of scientists and mathematicians. For example, the Next Generation Science Standards (NGSS; NGSS Lead States, 2013) includes engaging in argument from evidence as one of the eight science and engineering practices. Similarly, the Common Core Standards for Mathematical Practice (Council of Chief State School Officers, 2010) include the ability to "construct viable arguments and critique the reasoning of others" as one of the eight standards for mathematical practice that educators should seek to develop in their students. However, in practice, the teaching and learning of scientific and mathematical argumentation remains a challenge for both teachers and students (Henderson et al., 2017; Weber, 2001.

In the 5 years spanning 2011 to 2015, the DRK-12 program funded or cofunded 23 projects relating to argumentation, totaling more than $40 million in value.

This report presents a cross-disciplinary synthesis of argumentation in mathematics and science education, as studied in recent projects funded by the National Science Foundation's (NSF's) Discovery Research PreK–12 (DRK-12) program. This synthesis includes studies of the outcomes of interventions aiming to improve both the teaching and learning of scientific and mathematical argumentation as well as studies aiming to describe and document the processes of argumentation that occur in science and math classes, with the goal of informing future interventions. Based on a structured review of recent DRK-12 award abstracts, we identified argumentation as a key area of NSF investment in preK–12 science, technology, engineering, and mathematics (STEM) education.[1] In the 5 years spanning 2011 to 2015, we found that the

[1] This report's synthesis on NSF-funded research on argumentation comes from a broader NSF project, *Advancing Methods and Synthesizing Research in STEM Education* (DRL-1813777), which aims to synthesize evidence of innovation and discovery in recent DRK-12 projects.

DRK-12 program funded or cofunded 23 projects relating to argumentation that totaled more than $40 million awarded, demonstrating the depth and importance of this area of research.

Defining Argumentation

Although a single definition of scientific or mathematical argumentation is unsettled in the literature, Toulmin's (1958) model of argumentation has been widely used in many humanities and social science fields, including science education research (Erduran et al., 2004) and mathematics education research (Inglis et al., 2007). In its simplest form, Toulmin's model posits that an argument consists of at least four components: (a) the claim (an assertion to be supported); (b) data (the facts to support a claim); (c) warrant (the "bridge" or reasoning that links the data and claim); and (d) backing (the support for the appropriateness of the warrant). Exhibit 1 summarizes these core components, including examples of disciplinary applications and other common names (e.g., Toulmin's "data" is sometimes called "grounds" or "evidence").

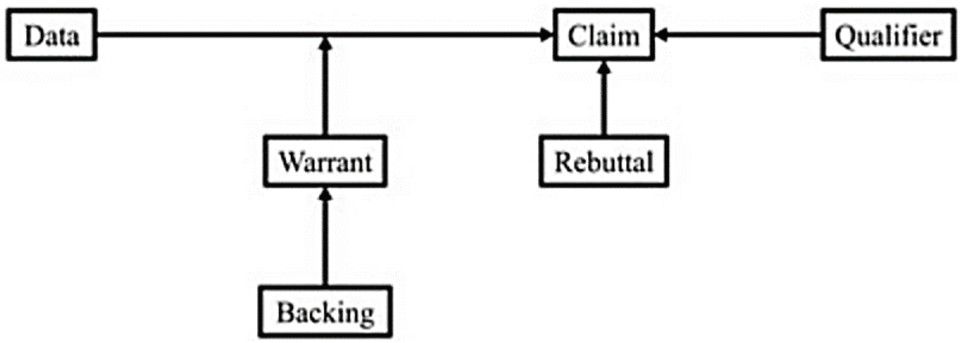
Exhibit 1. Three Key Components in Toulmin's Model of Argumentation

Component	Definition	Example	Other names
Claim	An assertion to be supported	*Math*: The sum of the angles in a triangle is always 180 degrees. *Science*: Light obeys the inverse square law.	Explanation
Data	The facts to support a claim	*Math:* Several triangles with angles of different sizes, tested by ripping off the angles and forming a line with them. *Science*: Collected data on the intensity of a light source at varying distances.	Grounds, evidence
Warrant	The bridge that links the data and claim	*Math*: Because the three angles always form a straight line, their sum is 180. *Science:* Statistical analysis fits the observed data to a power law relationship.	Reasoning[a]
Backing	Supports the appropriateness of the warrant	*Math*: When you bisect a triangle with a line parallel to one side of the triangle, the angles formed along the line are congruent to the angles in the triangle. *Science*: Light intensity is expected to follow a power law relationship.	Support

[a] Reasoning is sometimes used to describe both warrant and backing collectively.

Toulmin's model also expands to include other components found in more complex arguments: (a) *qualifiers* that limit the certainty of the claim and (b) *rebuttals* that specify conditions under which the claim might not be true. Exhibit 2 shows how these additional argument components support one another.

Exhibit 2. Graphical Depiction of Toulmin's Model of Argumentation



Note. Adapted from DeJarnette & González (2017).

Researchers use this framework as a methodological tool for coding and analyzing classroom discourse, but distinguishing argumentation components (e.g., data vs. warrant) is challenging in practice, with scholars viewing the model as useful but incomplete (Erduran et al., 2004; Kelly & Takao, 2002). Other scholars have suggested that Toulmin's model gives insufficient attention to argumentation as a social activity that involves back-and-forth dialogue and the persuasion of others (Jiménez-Aleixandre & Erduran, 2007; Manz, 2015; Walton, 1996). (See Sampson and Clark [2008] for a review of other common frameworks for argumentation in science education.) Therefore, science education and mathematics education researchers have built upon and extended the model. As one example from a recent DRK-12 project, McNeill and colleagues (2016) adapted Toulmin's framework to detail the structural aspects of argumentation while drawing on other research to frame the dialogic aspects of argumentation as a social activity.

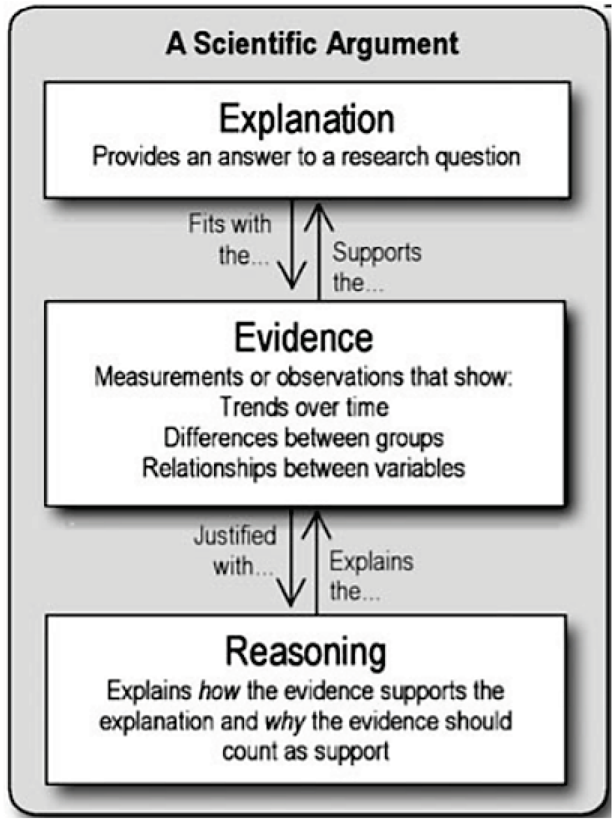
In addition, although Toulmin's model provides an overview of the fundamental components of argumentation that are common across disciplines, some disciplinary differences have emerged in the way argumentation has been defined within science and mathematics. In the following two sections, we use Toulmin's model as a framework to identify similarities and differences in how argumentation is discussed in these two disciplines.

Defining Argumentation in Science

The NGSS framework defines scientific argumentation ("engaging in argument from evidence") as a process for reaching agreements about scientific claims based on evidence and reasoning (see Exhibit 3). Using Toulmin's terminology, evidence is the "data," and reasoning is a combination of the "warrant" (i.e., how the evidence supports an explanation) and "backing" (i.e., why the warrant is appropriate).

The NGSS framework also distinguishes *argumentation* and *explanation*, noting that they are closely related but distinct practices (see also Berland & McNeill, 2012; Berland & Reiser, 2009; Osborne & Patterson, 2011). Explanations detail the causes for a phenomenon (e.g., why global temperatures have increased in recent decades), whereas arguments use evidence to identify the best explanation (e.g., using empirical data to support claims about specific causes of global climate change). Multiple explanations for a phenomenon may exist, often with varying degrees of supporting evidence (e.g., global climate change is driven by human activity; climate change is part of a normal temperature cycle). In Toulmin's terminology, a claim can be an explanation, and an argument is the entire package of an explanation (claim) supported by evidence (data) and reasoning (warrant/backing).

Exhibit 3. Model of Scientific Argumentation



Note. Adapted with permission from Sampson and Clark (2009).

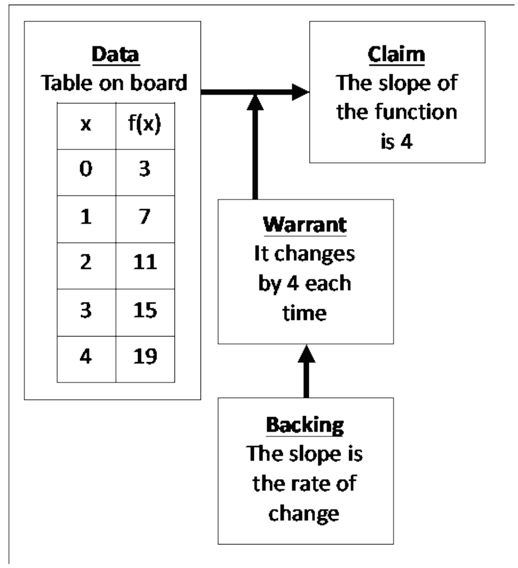
Defining Argumentation in Mathematics

The Common Core Standards for Mathematical Practice and many state mathematical standards document describe multiple components of students' proficiency in argumentation ("construct viable arguments and critique the reasoning of others"). These proficiencies can be categorized into three subcomponents: (a) *construction*, including exploring a mathematical situation, making claims about what might be true, and justifying those claims; (b) *reading*, including understanding, comparing, and evaluating others' arguments; and (c) *presentation*, including explaining an argument and convincing an audience of its validity (Mejía-Ramos & Inglis, 2009).

One important type of mathematical argument is proof, which usually has more formal epistemic requirements about "what counts" as a valid argument (Boero et al., 2010; Yackel & Cobb, 1996). Studies have shown evidence that children as young as elementary school can engage in discourse containing features of mathematical proof and lead to the generalization of mathematical ideas (Russell et al., 2011). In Toulmin's terminology, claims in proofs are constructed deductively from data with warrants established by mathematical theory (Durand-Guerrier et al., 2012; Pedemonte, 2007). However, proofs are only a subset of mathematical arguments, which can also include claims constructed inductively from warrants based on visual evidence, established patterns, or examples (Durand-Guerrier et al., 2012; Pedemonte, 2007). Exhibit 4 shows an example nonproof argument, described using Toulmin's model.

Despite subtle disciplinary differences, the past two decades of research in science and mathematics education have seen argumentation emerge as an important disciplinary practice for improving the teaching and learning of science and mathematics content as well as more general skills, such as critical thinking and communication. The following sections outline the extant research on these potential benefits of argumentation for teaching and learning.

Exhibit 4. Toulmin's Model Applied to a Hypothetical Mathematical Argument



Note. Adapted with permission from Singletary and Connor (2015).

Research on the Learning Benefits of Argumentation

Prior research on argumentation has identified at least three beneficial aspects of engaging in argumentation for students. First, because argumentation is a core disciplinary practice in science and mathematics, engaging in argumentation can help enculturate students into the social norms of the discipline (Manz, 2015; Yackel & Cobb, 1996). For example, elements of argumentation, such as dialogue and critique, are central features of the way scientific and mathematical knowledge is constructed (Ford & Wargo, 2011). Instructional activities, such as critiquing other students' ideas through writing tasks and classroom discussions, can help support students' epistemic views about acceptable forms of argumentation in science and mathematics (Ford, 2012; Yackel & Cobb, 1996). In addition, such activities can demonstrate that the nature of mathematical and scientific claims are tentative and involve convincing others in a community of their validity (González-Howard & McNeill, 2020). Engaging in argumentation allows students to engage in authentic practices of the discipline, which can improve their agency in applying these skills in novel situations (Brown et al., 1989). Indeed, Driver et al. (2000) argued that any education about science "must give the role of argument a high priority if it is to give a fair account of the social practice of science" (p. 287).

Second, engaging in argumentation could have cognitive benefits for students' understanding of mathematical and scientific content (Schwarz et al., 2009). Empirical studies have found that students' argumentation abilities can predict achievement in both mathematics (Cross, 2008) and science (Bathgate et al., 2015). Argumentation has been shown to support student learning through three primary mechanisms: (a) facilitating socio-cognitive conflict, (b) encouraging cognitive elaboration, and (c) improving interest and motivation (see Nussbaum, 2008, for a review). First, engaging in argumentation requires the explicit articulation and public exchange of ideas, which can facilitate the meta-cognitive reflection and self-explanation that leads to conceptual change (Chi et al., 1989). Second, through discourse, argumentation provides opportunities for students to recognize gaps or inconsistencies in their justifications, which compels them to revise and reorganize their knowledge to be more coherent (Greeno & van de Sande, 2007; Means & Voss, 1996). Third, students who view scientific and mathematical ideas as open to debate may be more interested, feel more agentic, and seek more opportunities to learn (Chinn, 2006).

Finally, engaging in argumentation is critical to the development of students' everyday communication and reasoning skills. Some scholars define argumentation as a process of discourse through which "something open to question is transformed into something mutually accepted, in principle, by participants in a debate" (Banegas, 2013, p. 49). In this broad sense, informal argumentation is common in both classroom discourse and everyday social interactions, and engaging students in argumentation can lead to improved communication and critical thinking skills necessary for engaged citizenship (Jiménez-Aleixandre & Erduran, 2007; Knudsen et al., 2017).

To summarize, researchers identified that engaging students in argumentation is (a) a critical component of developing and practicing disciplinary norms in science and mathematics, (b) beneficial for students' learning of scientific and mathematical content, and (c) a core component of the development of communication and reasoning skills that are applicable more broadly in everyday life.

Research in Mathematics Education

Prior educational research in mathematical argumentation has spent considerable effort in defining the structure and components of mathematical argumentation and proof (Staples et al., 2016), as well as understanding the challenges that students encounter and how to overcome them (Campbell et al., 2020). Students often struggle with the degree to which they can use specific examples as sufficient evidence (data, warrants) to support their mathematical claims (Cirillo & Hummer, 2019). Many studies suggest that providing generic examples (i.e., emphasizing how the example applies to the general) can support students in moving away from arguments based on specific examples and toward more formal, generalizable mathematical arguments (Aricha-Metzer & Zaslavsky, 2019; Ellis et al., 2017). Recent technological advances provide scaffolded opportunities for students to construct claims (independently or collaboratively) by generalizing from many examples (Campbell & Zelkowski, 2020). For example, dynamic geometry software allows students to discover warrants by testing claims using observational data in a virtual environment (e.g., GeoGebra; Alqahtani & Powell, 2015; Fukawa-Connelly & Silverman, 2015).

Research on teachers' instruction in mathematical argumentation remains an emerging area of study (Campbell et al., 2020; Walshaw & Anthony, 2008). Much of the research has focused on how teachers socialize students into mathematics by conveying what are acceptable disciplinary norms around argumentation (Walshaw & Anthony, 2008; Yackel & Cobb, 1996). Yackel and Cobb (1996) defined *sociomathematical norms* as social expectations specific to mathematical activity (e.g., what counts as acceptable mathematical explanation and justification), contrasting with more content-general social norms (e.g., expectations that students should generally explain their thinking). Such norms can support and encourage specific student activities for productive discourse (e.g., making bold conjectures that a student is not sure of but proposes anyway to stimulate discussion).

Teacher questioning is a key pedagogical practice in setting these norms and supporting student argumentation, but this practice can widely vary in terms of the question form (e.g., leading vs. probing; Kosko et al., 2014) and question content (e.g., attending to mathematical details, building on students' ideas; Singletary & Connor, 2015; Zambak & Magiera, 2018). Intervention research has built on these naturalistic studies to develop professional development (PD) programs. For example, Knudsen et al. (2014, 2017) developed a four-part model for supporting the pedagogy of mathematical argumentation in middle school classrooms, comprising *generating cases* (creating example as a basis for argument), *conjecturing* (coming up with claims with yet-undetermined mathematical validity), *justification* (using data and warrants to establish a claim's validity), and *concluding* (coming to agreement about a claim).

Research in Science Education

Despite the importance of scientific argumentation to the field, traditional science instruction rarely engages students in this practice (Driver et al., 2000; Newton et al., 1999). More commonly, students will learn a set of facts or explanations presumed to be scientific truths (e.g., how photosynthesis or plate tectonics work), without understanding the supporting empirical evidence (Berland et al., 2017). When asked to defend their claims, students often select inappropriate evidence that does not support the claim (McNeill & Krajcik, 2007) or fail to explain how the evidence supports the claim (i.e., do not provide the warrant; Bell & Linn, 2000; Sandoval & Millwood, 2005). Despite these challenges, students can exhibit nascent forms of scientific argumentation (Berland & Hammer, 2012; Berland & Reiser, 2009). Berland and McNeill (2010) introduced a learning progression for scientific argumentation that increasingly moves students from simple to more complex practices (e.g., "claims are defended” to “claims are defended with evidence” to “claims are defended with evidence and reasoning”; see also Osborne et al., 2016).

Teachers are critical in shaping students’ opportunities to practice scientific argumentation. They establish norms about classroom discourse and set expectations about whether students should defend their claims or not (Berland & McNeill, 2010; Osborne, 2010). Yet previous research suggests that science teachers can experience difficulty in integrating argumentation into their classrooms (Simon et al., 2006). Teachers even exhibit many of the same challenges that students have in defending and evaluating claims. In one study of 30 middle and high school science teachers, teachers struggled with the use of evidence when asked to evaluate alternative explanations and generate arguments; they rarely included evidence and reasoning to support their arguments, tending to rely on their content knowledge rather than use available data (Sampson & Blanchard, 2012). Recent studies, however, have begun to develop and test interventions to support teachers, especially through PD programs about facilitating classroom discourse (e.g., Berson et al., 2015; Fishman et al., 2017; Murphy et al., 2018).

Core Research Topics in Mathematical and Scientific Argumentation

Based on this review of the literature, our synthesis aimed to understand how recent DRK-12 projects, funded between 2011 and 2015, addressed critical gaps in research knowledge about argumentation in science and mathematics education. This topic is especially timely given the inclusion of argumentation in the NGSS and Common Core Standards for Mathematical Practice, which were both released in the early 2010s. Researchers designed and planned these recent DRK-12 projects at a critical time when educators needed further guidance and support to implement new expectations about engaging students in argumentation. The funding of DRK-12 projects on argumentation reflect this need; in just the 5-year span in which our synthesis focuses, the DRK-12 program funded 23 projects with a total value of $40.5 million. Specifically, our synthesis examines the body of DRK-12 funded work with regard to four core research topics identified in the recent argumentation literature: (a) understanding student learning trajectories in argumentation; (b) understanding teachers’ beliefs and practices about teaching argumentation; (c) a comparison of approaches to teaching and learning argumentation in science and mathematics; and (d) approaches to supporting student argumentation across varying cultural and linguistic backgrounds.

Research Topic 1: Student Learning Trajectories in Argumentation

Questions remain for the field about what tools and resources can most effectively aid in the developmentally appropriate support of students’ argumentation skills. Prior literature reviews have outlined potential student learning trajectories for moving students from nascent to more sophisticated disciplinary argumentation practices (e.g., Berland & McNeill, 2010; Osborne et al., 2016). But much remains to be known about how to most effectively advance students along these trajectories. New digital technologies (e.g., dynamic geometry software in mathematics, computational models in science) have provided useful scaffolds, helping support the collection of data, investigation of warrants, and construction of claims (Scheuer et al., 2010). Effectively using these tools requires additional testing, understanding for whom they work and for what type of outcomes (e.g., declarative knowledge about argumentation vs. quality of written arguments; Wecker & Fischer, 2014).

Research Topic 2: Teacher's Argumentation Beliefs and Instructional Practices

More research is needed to understand how teachers develop the knowledge and skills to enact argumentation instruction effectively. Research in mathematics and science education has identified key instructional skills and practices for the successful teaching and learning of argumentation, but less is known about how teachers develop these skills (Walshaw & Anthony, 2008). As noted earlier, this area of research is especially important given the new demands placed on preK–12 teachers by the NGSS and the Common Core Standards for Mathematical Practice. In one study, teachers felt especially unprepared in teaching about the practice of argumentation, compared with other NGSS science and engineering practices (e.g., ranking it seven out of eight in terms of teaching confidence; Kang et al., 2018). Using argumentation models like Toulmin's model in PD could help scaffold teachers' ability to recognize components of students' arguments in the "heat of the moment" of classroom discussions (McNeill & Knight, 2013).

Research Topic 3: Comparison of Argumentation in Science and Mathematics

Although some empirical studies have investigated argumentation across mathematics and science education (Kannan et al., 2018), most prior reviews on argumentation in education tend to be discipline specific (e.g., focusing on one field, such as science education; Jiménez-Aleixandre & Erduran, 2007) or discipline general (e.g., downplaying differences across fields; Rapanta et al., 2013). Therefore, empirical work remains to be done to understand the disciplinary aspects of argumentation (i.e., what is common and distinct across disciplinary fields).

Research Topic 4: Supporting Student Argumentation Across Varying Cultural and Linguistic Backgrounds

Finally, understanding how to support student argumentation across varying cultural and linguistic backgrounds is an important yet still relatively understudied area of research (Civil & Hunter, 2015). Given the focus on dialogue in argumentation instruction (i.e., speaking and writing), English learners (ELs) may require special supports to engage in these practices effectively. For instance, placing ELs into language-matched student pairs (e.g., both speak Spanish as their native language) can enable them to use both their native and second languages as linguistic resources in ways not possible in other classroom environments, such as whole-class discussions (González-Howard & McNeill, 2016).

What Was Studied?

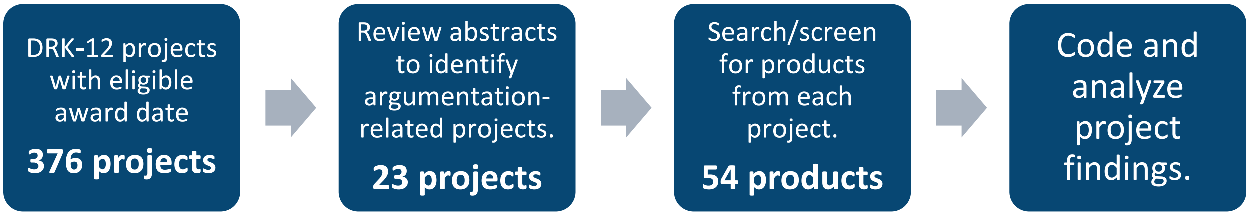
Our Synthesis Approach

This study examined published research and other products, which were developed with the support of DRK-12 grants, to synthesize existing knowledge and identify gaps in understanding to help guide future research. Appendix A details our review methodology, which we briefly summarize here. We examined DRK-12 projects with an original award date spanning January 2011 to December 2015, to focus on recently completed or close-to-completion projects. We downloaded the award abstracts for all DRK-12 awards in this date range using NSF's website. When screening projects for their relevance to argumentation, we defined argumentation as "a process for reaching agreement about explanations and design solutions" (see Appendix A for further elaboration and domain-specific definitions). Some award abstracts mentioned the term "argumentation" explicitly, whereas others referenced the concept in other ways (e.g., "using evidence to support claims"). This process yielded 23 eligible projects.

We identified products generated by these 23 projects by searching six sources: Web of Science, ERIC, PsycINFO, Google Scholar, Research.gov, and the CADRE website (cadrek12.org). The search strategy targeted (a) documents that referenced the numeric NSF award ID and (b) documents listed by project leaders on the Research.gov or CADRE websites, which together identified 54 research products that were then coded and reviewed (see Exhibit 5).

Appendix B contains a complete table with brief summaries of all *projects* (i.e., collections of materials produced from a single NSF award) and *products* (i.e., the specific studies, manuscripts, and other documentation of the work from each project) that are included in this synthesis. Throughout the synthesis, we will refer to *projects* using the seven-digit NSF award number for that project and refer to *products* using unique four- or five-digit numbers identifying specific products, which also can be used to look up additional details about a particular product from the table in Appendix B to reference.

Exhibit 5. Overview of Our Synthesis of 27 DRK-12 Projects Related to Argumentation



What Were the Major Areas of Research?

The focus of the reviewed projects on argumentation were categorized into four major areas of research across two different dimensions, as shown in Exhibit 6: (a) projects focusing on either **teachers** or **students** as the primary participants and (b) projects that either examined the impact of an **intervention** or that conducted **naturalistic** observations of teaching and learning in a "business-as-usual" classroom context.

Projects (i.e., awarded grants) were categorized based on the qualitative coding of the research *products* (i.e., disseminated reports of research) generated by each project. If a project contained at least one product that reported on a research study that fell within one of our four major areas of research, that project was placed within that category (e.g., a student-focused intervention project had at least one product that investigated student engagement or outcomes from an intervention). Therefore, as might be expected, projects often fell into more than one category (e.g., both student focused and teacher focused). For example, eight projects generated products that reported on research that was both teacher and student focused and were therefore counted toward both categories, whereas three projects generated products that contained both intervention and naturalistic studies.

Exhibit 6. The Number of Projects Within Each Major Area of Research

	Student focused	Teacher focused
Interventions	15 projects (30 products)	8 projects (13 products)
Naturalistic observations	7 projects (11 products)	4 projects (9 products)

Note. Projects could be counted more than once across different categories.

Projects were coded to the intervention category if they used any methodology (i.e., quantitative or qualitative) to examine the effect of a designed manipulation on students' or teachers' experience of the teaching and learning of argumentation. We also found value in findings from studies that did not directly evaluate an intervention but instead took a "fieldwork" approach. For example, many studies conducted naturalistic observations of students and teachers engaging in scientific and mathematical observation in "business-as-usual" classroom settings without any intervention or manipulation. We also include in this report a synthesis of results from projects that take this approach.

The largest group of projects were those that included student-focused intervention studies (15 projects), followed by teacher-focused intervention studies (eight projects), naturalistic studies of student argumentation (seven projects), and naturalistic studies of teachers' argumentation instruction (four projects). In the sections that follow, we use these categorizations as high-level organizers of our synthesis of the data and present descriptive findings and interpretive themes both overall and within each of these categories. In Exhibits 7–10, the darker shades and larger font size indicate relative size of the categories. As previously noted, our coding of these categories was not mutually exclusive; that is, it was possible for a study to be counted in multiple categories. Therefore, the rows and columns will not necessarily sum to the total of the overall projects in that category.

What Grade Levels and Disciplines Were Studied?

Overall, most of the projects were conducted in either middle school (16 projects) or high school (14 projects) settings, with relatively few projects taking place in elementary grades (five total; see Exhibit 7). None of the projects took place in preK. In terms of disciplines studied, the projects were relatively evenly split between science (13 projects) and mathematics (10 projects). However, within each of these larger disciplines, some subdisciplines were particularly common, such as earth science (seven projects) and biology (seven projects) within science, and geometry (seven projects) within mathematics.

Exhibit 7. Descriptive Frequencies of Grade and Discipline by Project

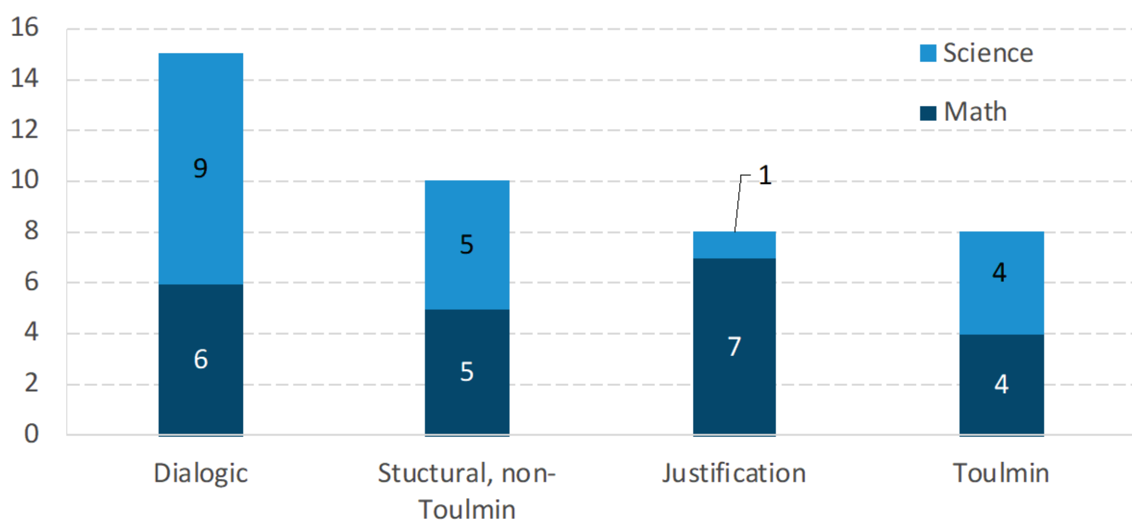
	All	Student intervention	Teacher intervention	Student naturalistic	Teacher naturalistic
Number of NSF projects overall	23	15	8	7	4
Grade level					
Middle school	16	9	4	4	3
High school	14	10	4	4	1
Late elementary	3	2	1	0	1
Early elementary	2	1	0	0	1
Pre-kindergarten	0	0	0	0	0
Major disciplinary field					
Science	13	10	3	2	1
Mathematics	10	5	3	5	3

There also were some differences when looking across the two dimensions of student/teacher and intervention/naturalistic projects by discipline. Specifically, there were a relatively high number of projects in science focusing on student interventions (10 projects) and fewer naturalistic projects (three projects), whereas mathematics had more projects that were at the intersection of intervention (eight projects) and naturalistic (eight projects). There also were a higher number of projects with intervention studies in science (13 projects) than in mathematics (eight projects).

How Was Argumentation Defined?

Overall, projects were split between those that defined argumentation as dialogic (i.e., argumentation as a social practice [15 projects]) and those that used a structural approach in their definition of argumentation (i.e., argumentation as a set of formal components [18 projects]; Exhibit 8). Eight projects also discussed argumentation as justification (i.e., providing explanations for why a claim is true); although seven of these projects were in mathematics, one project on scientific argumentation also used this definition. Within projects using structural definitions, researchers used Toulmin's model and other, non-Toulmin structural definitions at about the same rate (eight projects and 10 projects, respectively). By discipline, argumentation projects in science had a slightly higher number of dialogic studies, whereas science and mathematics had similar numbers of studies that were based on some structural definition of argumentation.

Exhibit 8. Stated Definition of Argumentation, by Project and Discipline



Note. "Use of scientific evidence" also was coded in 12 studies (science only), and "proof" was coded in seven studies (math only).

How Was Argumentation Measured?

For student-facing projects, the three most common measurement methods were short-answer responses (11 projects), multiple choice (10 projects), and observations of teacher-student interactions (10 projects), followed by student-student interactions (eight projects) and interviews (five projects; see Exhibit 9).

Within student-facing naturalistic studies, the most common measurement methods were teacher-student interactions and interviews (each with four projects). Somewhat surprising was that there were relatively few student argumentation projects that used Likert-scale surveys (three projects), possibly reflecting the difficulty of measuring argumentation using these types of scales. Examples of Likert-scale measures that were used were scales of confidence and motivation in science and scientific argumentation (project #1316799), the usefulness of automated feedback on students' arguments (project #1418019), and one measure of environmental science knowledge that asked students to rate the extent to which a scientist would agree with a particular statement (project #1316057). Also underrepresented were studies of students' argumentation measured using essays or written proof.

Exhibit 9. Descriptive Frequencies of Student Measurement Methods by Project

	All	Student intervention	Teacher intervention	Student naturalistic	Teacher naturalistic
Number of NSF projects overall	23	15	8	7	4
Student measurement method					
Short answer	11	9	1	3	0
Multiple choice	10	9	2	2	0
Teacher-student interactions	10	8	5	4	1
Student-student interactions	8	7	3	1	0
Interviews	5	3	1	4	0
Likert surveys	3	3	0	0	0
Essays and written proofs	2	2	1	0	0
Online discussions	1	1	0	0	0

For teachers, the most common form of measurement was observations of teacher-student interactions, which was most common in naturalistic studies (eight projects), followed by short-answer responses (five projects), which was most common in intervention studies (four projects). Less common for teachers were survey measures—both Likert-scale (three projects) or multiple choice (no projects; see Exhibit 10).

Exhibit 10. Descriptive Frequencies of Teacher Measurement Methods by Project

	All	Student intervention	Teacher intervention	Student naturalistic	Teacher naturalistic
Number of NSF projects overall	23	15	8	7	4
Teacher measurement method					
Teacher-student interactions	8	5	7	1	2
Short answer	5	1	3	0	3
Likert surveys	3	1	2	0	2
Interviews	3	0	2	0	2
Reflections	3	1	2	0	1
PD obsevations	2	0	2	0	0
Lesson plans	2	1	2	0	0
Log files	1	0	1	0	0

Note. The following codes did not appear in the data: multiple choice, online discussions, and clinical simulations.

Summary of What Was Studied

Overall, most of the argumentation projects reviewed focused on student interventions (15 of 23 projects, or 65%) and studied middle school (16 of 23 projects, or 70%) or high school (14 of 23 projects, or 60%) classrooms. The relatively small number of studies focused on argumentation in elementary classrooms (five of 23 projects, or 22%) and the absence of studies in preK suggests this may be an important gap in the research on argumentation.

Descriptive analyses of what was studied also revealed some disciplinary differences in the focus on argumentation and the way is has been defined in science and mathematics. Science projects had slightly more emphasis on dialogic definitions of argumentation than mathematics, whereas there was a more even split of structural definitions across the two disciplines. Disciplinary differences also appeared in study design and approach, with relatively more science projects producing intervention studies, whereas there was a more even representation of naturalistic studies in science and mathematics.

Finally, there were few survey measures of argumentation for students and none for teachers, possibly signifying the difficulty in measuring argumentation outcomes using these methods, as well as the relatively small number of empirical, teacher-focused intervention studies that examine practices and beliefs, outcomes often measured using these kinds of items. The most common measurement for student interventions was either short answer or multiple choice, whereas the most common form of measurement for teachers was teacher-student interactions and short answer.

What Was Learned About the Outcomes of Argumentation Interventions?

Projects (i.e., collections of work under a single NSF award) focusing on the outcomes of argumentation interventions were grouped first based on the population that was the focus of the intervention (e.g., students or teachers), and then according to themes derived from the coding of each *product* (i.e., studies, manuscripts, and other research output from a project; see Exhibit 11).[2] Findings below are reported by theme, with the projects that fall under that theme reported in blue boxes and tagged with the NSF project number and project title. Within each theme, findings from individual products that were coded to that theme are reported and tagged with a study-specific product number for reference (see Appendix B for a full list of projects and associated products). For each product, the study sample size, relevant discipline, and grade level also are reported.

These products employed a range of research methodologies and designs. It is important to note that some designs (e.g., randomized controlled trials [RCTs], some quasi-experimental designs) provide stronger causal evidence that the resulting effect is indeed a result of the intervention, whereas other designs (e.g., pre-post tests) leave open the possibility that these changes could result from other factors (e.g., retesting effects, history effects/confounds, natural growth over time [Marsden & Torgerson, 2012]). Therefore, to exercise caution and provide clarity about the relative certainty of the evidence presented, in the following sections we note the study design when mentioning evidence from difference research products and report statistical information on the size of effects only for intervention studies that employ experimental or quasi-experimental designs comparing at least two groups (i.e., a treatment and control group). Effects are reported as Cohen's *d* unless otherwise stated.

[2] Themes were identified by coding products, creating narrative summaries, and then regrouping them based on a secondary round of thematic coding (Flick, 2009).

Exhibit 11. Summary of Themes of What Was Learned About the Outcomes of Argumentation Interventions

	Student focused	Teacher focused
Interventions	**Interventions that make structure explicit** (5 projects[a]) **Interventions that use multiple representations** (5 projects[a]) **Interventions that provide feedback** (2 projects) **Interventions that support student discourse** (5 projects)	**Interventions that prepare teachers to support student discourse** (7 projects[b]) **Interventions focused on change in teacher knowledge and beliefs** (2 projects[b])

[a] Two projects were coded to both of these themes.
[b] One project was coded to both of these themes.

Student-Focused Interventions

The largest group of projects focused on student interventions (15 projects). These projects fell under four main themes, based on the approach taken to improve a variety of outcomes: students' learning of argumentation (14 projects), student's content learning (seven projects) or students' engagement with the intervention (eight projects).

Across each of these themes, a large proportion of these student intervention projects (six of 15 projects, or 40%) also developed and tested specific educational technology tools as a means to support student argumentation; in the project lists below, the specific tools associated with each project are reported in italics next to the project name.

Interventions That Make Structure Explicit

Five projects emphasized the importance of helping students notice, attend to, or name the underlying components of an argument (e.g., warrants, claims).

- #1119670 CLASS: Continuous Learning and Automated Scoring in Science (*Idea Manager*)
- #1417757 Learning Labs: Using Videos, Exemplary STEM Instruction, and Online Teacher Collaboration to Enhance K–2 Mathematics and Science Practice and Classroom Discourse
- #1317034 Learning Algebra and Methods for Proving
- #1418136 Building High School Students' Understanding of Evolution
- #1316057 Developing Critical Evaluation as a Scientific Habit of Mind: Instructional Scaffolds for Secondary Earth and Space Sciences

Findings. First, five research products on student intervention focused on **tools that support students' argumentation by organizing student ideas to make the structural components of their arguments explicit**. For example, one correlational study in high school chemistry (#19533; *N* = 164) found that using an online tool, the Idea Manager, supported students in recording their ideas and organizing their explanations by separating the explanation process into a concrete set of steps and was associated with argument coherence. One pre-post study (#5637; *N* = 944) and one randomized control trial (#5639, *N* = 2,269), both in high school biology, scaffolded the construction of evidence-based arguments using the claims, evidence, and reasoning (CER) framework. Evidence from the experimental study showed that students using the curriculum had significantly larger gains (*d* = 0.58) in conceptual content understanding than control students. Not all these interventions required advanced technological tools; one observational study in a single second-grade general science classroom (#75682) showed how using simple agree/disagree T-charts in an elementary classroom can help make students' ideas visible and connected to a specific claim, which supported them in generating and refuting claims as well as integrating new evidence.

Second, two additional products reported on **instructional approaches that use specific language or strategies to provide structure to students' arguments**. For example, one experimental study in middle school algebra (#91096, *N* = 210) reported on a PD program that encouraged teachers to ask students to develop explicit claims using mathematical language (e.g., "for all" and "there exists") and to support claims with viable referents and warrants. Findings demonstrate that treatment students outperformed comparison students on an argument and reasoning assessment and state standardized assessments (*d* = 0.68). Another pre-post study of a teaching experiment with a single middle school algebra student (#52559) showed how encouraging that student to approach argumentation through a process of "eliminating counterexamples" improved the student's ability to construct, critique, and validate contrapositive arguments.

Summary. There is evidence from two experimental studies that providing tools and supports that make the structure of arguments explicit can be an effective intervention for improving both content understanding (#5639) and general reasoning assessments (#91096). Interestingly, this approach seems to also be one that has common evidence across science and mathematics, suggesting an approach to learning argumentation that may not be discipline-specific. Other nonexperimental studies reviewed here describe related interventions (e.g., providing specific vocabulary or including specific virtual and paper-based tools) that could be ripe for future empirical evaluation.

Interventions That Use Multiple Representations

Five projects helped highlight important differences in argument characteristics by varying the way that an argument was presented or represented.

- #1118773 Collaborative Research: Computer-Supported Math Discourse Among Teachers and Students (*VMTwG*)
- #1453493 CAREER: Proof in Secondary Classrooms: Decomposing a Central Mathematical Practice
- #1324977 DIP: Community Knowledge Construction in the Instrumented Classroom (*WallCology*)
- #1316057 Developing Critical Evaluation as a Scientific Habit of Mind: Instructional Scaffolds for Secondary Earth and Space Sciences
- #1119670 CLASS: Continuous Learning and Automated Scoring in Science (*Idea Manager*)

Findings. A set of five research products across five projects examined how **the use of multiple forms of representations helped students learn argumentation**. To help understand why this approach may be successful, a pre-post study in high school geometry (#92500, $N = 389$) suggested that students tend to attend to superficial features of arguments that are presented in class (e.g., as two-column proofs) and not see underlying structural similarities of arguments when these features are changed (e.g., when proof is presented as a diagramming task). Multiple representations may reduce the importance of superficial features of arguments and help students focus on the underlying common elements of the argument (Ainsworth, 2006). For example, one quasi-experimental study in high school earth science (#52093, $N = 64$) showed significant increases in argument plausibility and knowledge scores for students when two alternative explanations were prompted, and no change with only one alternative. Two observational studies using dynamic geometry software in middle school (#43994, $N < 10$) and high school (#70936, $N < 10$) showed how that tool provided scaffolds for evidence (through multiple constructions) and warrants (through program rules), which allowed students' focus to shift from empirical explorations and task completion toward deductive justifications. Another pre-post study on using graphs to support argumentation in middle school biology (#90385, $N = 117$) showed that students who critiqued multiple graphical representations of evidence in their arguments were significantly more likely to identify important aspects of the graph and describe the narrative process represented in the graph. Constructing their own graphs during argumentation improved students' science understanding and increased the use of scientific content in their explanations from pre- to posttest. Other evidence from observational studies in biology with middle and high school students suggested that textual evidence may provide greater support than visual evidence in making connections to arguments (#73216, $N = 45$) and that using a variety of representations for persuasive purposes, rather than as artifacts, improves students' appropriate use of CER in their scientific explanations (#57402, $N = 47$).

Summary. There is some empirical evidence about the effect of the use of multiple representations on students' learning of argumentation, with a single quasi-experimental study that reported positive effects. Several observational and pre-post studies in both mathematics (e.g., multiple representations of proof, dynamic geometry software) and science (e.g., use of graphs, relative benefits of visual vs. textual evidence) offer design considerations that provide a useful framework for future research and development.

Interventions That Provide Feedback

Two projects focused on improving students' argumentation through opportunities for feedback and revision of their and others' arguments.

- #1220756 High Adventure Science: Earth Systems and Sustainability (*HASbot*)
- #1418019 Investigating How to Enhance Scientific Argumentation Through Automated Feedback in the Context of Two High School Earth Science Curriculum Units

Findings. Three products across two projects emphasized the **importance of providing students with real-time feedback on their arguments**. One hypothesis about the mechanism for these interventions is that feedback helps students to identify sources of uncertainty in their argumentation and engages them in improving their arguments by providing opportunities to generate, collect, and analyze their own data. One example of an educational technology tool designed for this purpose is HASbot, an automated text scoring and real-time feedback system designed to support student revision of scientific arguments. Results from a pre-post study (#9600, *N* = 343) show that students' use of HASbot was related to significant gains in scientific argumentation, above and beyond demographic characteristics. Another correlational study in high school earth science (#32872, *N* = 1,180) developed a computer-based formative assessment to support students' construction and revision of scientific arguments. A pilot study showed students' scientific argumentation skills improved during their revision process on 11 of 16 items. One pre-post study, also in high school earth science (#93829, *N* = 183), showed that the majority of students made revisions after receiving feedback and that students with higher initial scores were more likely to revise their responses. In addition, students who revised had significantly higher final scores than those who did not. Analysis on item difficulty shifts showed that written scientific argumentation became easier after students used the automated feedback.

Summary. The three products within this theme provide preliminary evidence for effects based on pre-post and observational studies. These products highlight several interesting technological tools that support revision and feedback, which could be a focus of more rigorous evaluation.

Interventions That Support Discourse

Five projects examined interventions that facilitate and improve the quality of argumentation discourse between students. These projects focused on argumentation as a discursive process—that is, arguments as the process of discussion between multiple actors to reach consensus. See Project Spotlight 1 below for more details on a project from this group.

- #1119518 Mathematical Argumentation in Middle School: Bridging From Professional Development to Classroom Practice (*Geometers' Sketchpad*)
- #1223021 Collaborative Research: Researching the Efficacy of the Science and Literacy Academy Model
- #1316347 Integrating Quality Talk Professional Development to Enhance Professional Vision and Leadership for STEM Teachers in High-Need Schools
- #1316799 Enhancing Argumentation With Social Media: Supporting Teacher Professional Development (*Reason Racer*)
- #1417895 Preparing Urban Middle Grades Mathematics Teachers to Teach Argumentation Throughout the School Year

Findings. One set of four research products examined **how teachers learned to support students to engage in productive argumentation discourse**. For example, a report of a pre-post study (#5631[3]) examining the effect of a Quality Talk (QT) PD program in high school physics and chemistry found an increase in students' scores on a measure of using scientific reasoning to build arguments across three time points. Importantly, students were better able to enact these discourse practices on their own, without teacher support, after QT Science discourse lessons. A quasi-experimental study in science (#24924, *N* = 301) reported that QT students produced significantly stronger written scientific arguments than comparison students. Another quasi-experimental study in elementary biology (#78764, *N* = 44) found that although there were no differences in teacher outcomes, the students of teachers who attended a practice-based PD program with an additional practicum component that included coaching and reflection made statistically significant improvements in their science discourse practices compared with students of teachers who attended the PD program only. A technical report of another PD program in mathematics (#90058[4]) presented design principles that included (a) establishing classroom norms to structure interpersonal argumentation and (b) using dynamic geometry software collaboratively in student group discussions. A pre-post study (#24152, *N* = 97) of the same program in middle school geometry classrooms showed that in all the target classrooms, students demonstrated significant and substantial learning gains.

[3] The sample size for this study was not reported.
[4] The sample size for this study was not reported.

Another set of three products examined the outcomes of **tools that promote student discourse**. For example, one observational study (#24600, *N* > 500) examined the argumentation discourse of middle school science students engaged in an online multiplayer game (Reason Racer). An experimental study (#56233, *N* = 402) showed that a group of high school biology students using Reason Racer scored significantly higher overall than the comparison group on their knowledge of scientific argumentation and a significant overall increase in confidence regarding scientific argumentation (effect sizes were not reported for this study). Results from another quasi-experimental study focusing on attitudinal outcomes showed that middle school students who played the game at least 10 times also reported a significant increase in confidence and motivation to engage in science compared with students who did not play the game (#53147, *N* = 249). Results also showed that students who discussed a topic related to social issues were significantly more likely to use science vocabulary and have social interactions than students who discussed a purely scientific topic but were less likely to have substantive discussions.

Summary. There is empirical evidence from experimental and quasi-experimental studies that technological tools and PD aiming to support productive classroom discourse could be a powerful approach to improving students' argumentation and content knowledge. Specifically, pedagogical approaches and technological tools, such as Reason Racer or dynamic geometry software, that provide a structure to student-student discussions appear to be particularly powerful mechanisms for improved student outcomes.

Project Spotlight 1: Using Technology to Design Experimental Studies of Argumentation

Enhancing Argumentation With Social Media: Supporting Teacher Professional Development (NSF award #1316799; total funded amount = $429,784)

Why Spotlight This Project?

The project used observational methods to develop a technology platform (Reason Racer) with an innovative student-student discourse feature and then used this tool to facilitate an experimental study of its impact on students' scientific argumentation. Reason Racer is an online, multiplayer game in which students race cars between "PitStops," and they are asked to evaluate the claims, evidence, and reasoning of a scientific argument.

What Was Studied?

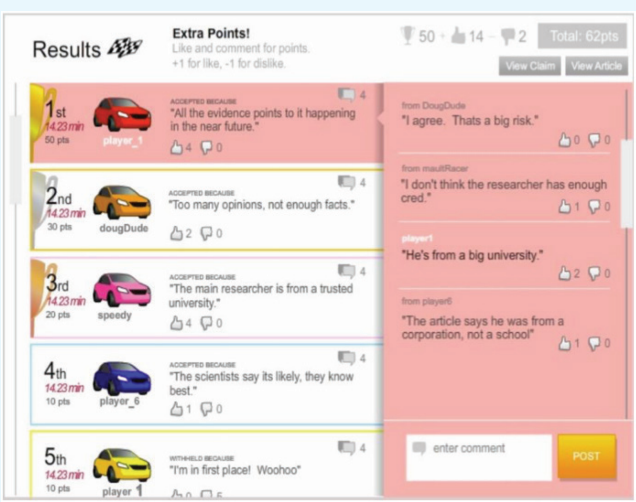
The project used a mix of observational, quasi-experimental, and experimental methods to test the efficacy of the intervention with middle and high school biology students. Both a structural (Toulmin's model) and dialogic definition of argumentation were used.

What Was Found?

Observational study. Craig-Hare et al. (2017) collected online discussions from more than 500 middle school students playing Reason Racer with both scientific and socio-scientific prompts. Results indicated that students were just as likely to use scientific argumentation vocabulary in both topics but be more social and less substantive with socio-scientific topics.

Quasi-experimental study. Ault et al. (2015) showed that additional gameplay was associated with improvement in scientific argumentation, with students who played Reason Racer at least 10 times showing significant improvement in their performance and confidence toward scientific argumentation.

Experimental study. Rowland et al. (2017) conducted an experimental study showing that students who participated in a unit using Reason Racer and social media scored significantly higher on measures of both performance and confidence in scientific argumentation than a comparison group.



The Reason Racer student discourse environment (used with permission from Craig-Hare et al., 2017)

Teacher-Focused Interventions

A smaller set of eight projects comprised interventions that were focused on teacher outcomes. All of the projects targeted improving teachers' argumentation instruction, through a combination of workshops (eight projects) and curricular interventions (seven projects). The projects generally focused on interventions that aimed to improve teachers' argumentation instruction either through preparing teachers to better support students (e.g., through facilitating classroom discussions), or by directly influencing teachers' knowledge and beliefs about argumentation.

Interventions That Prepare Teachers to Support Student Discourse

Seven projects focused on developing teacher moves and practices that help scaffold student learning and encourage productive classroom discussions around argumentation.

- #1119518 Mathematical Argumentation in Middle School: Bridging From Professional Development to Classroom Practice
- #1223021 The Role and Use of Examples in Learning to Prove
- #1316347 Integrating Quality Talk Professional Development to Enhance Professional Vision and Leadership for STEM Teachers in High-Need Schools
- #1417895 Preparing Urban Middle Grades Mathematics Teachers to Teach Argumentation Throughout the School Year
- #1118773 Collaborative Research: Computer-Supported Math Discourse Among Teachers and Students
- #1418136 Building High School Students' Understanding of Evolution
- #1503511 Developing Teachers' Capacity to Promote Argumentation in Secondary Science

Findings. Four products emphasized **how PD helped teachers develop specific instructional moves** that supported students' learning. One report provided design principles for PD programs for teaching argumentation that included improvisational games and curriculum supports as methods of improving instruction. An experimental evaluation of the effectiveness of a PD program based on these principles (#24152, N = 20) showed that middle school geometry teachers who experienced the PD used more argumentative discourse than control teachers (d = 0.3–0.5). Those teachers were more likely to (a) use both closed- and open-ended questions and (b) facilitate and encourage the participation by multiple students, which are both practices that support the learning of argumentation skills. A case study (#70936) of a single teacher's use of a virtual geometry technology that supports joint problem solving showed that the teacher used collaboration, mathematical reasoning, and technology to move students from making empirical explorations to deductive justifications. For example, the teacher established group collaboration norms to encourage students to develop and communicate convincing arguments that satisfy their peers and added more explicit questions to curriculum prompts, which motivated learners to use the virtual technology tool to notice geometric relationships while manipulating objects. Another quasi-experimental (#10006, *N* = 44) and observational (#38071, *N* < 10) study, both in upper elementary science, noted how video-based reflection can provide opportunities for teacher practice and reflection that may be valuable features of PD programs. Specifically, authors identified three components potentially related to teacher instructional change: talk format (e.g., instruction, whole/small group, pair, individual), activities (e.g., charting ideas, "four corners," cartoons), and teacher moves (e.g., ask, support, press, link). After PD, the teacher used more talk formats that encouraged interaction, more activities that provide opportunities for discussion, and teacher moves that supported students' engagement in scientific discourse.

A second set of four products focused on **how PD prepared teachers to facilitate student discussions in their classrooms**. A practitioner-focused report in mathematics (#51831, *N* < 10) suggested that teachers' classroom discussions followed a certain pattern. First, teams or individuals generate conjectures, and the whole class justifies them. Then, the whole class together generates and justifies conjectures. Next, teams or individuals come up with conjectures and justifications, which are shared with the class. Finally, there is a conclusion in which the class comes to a consensus that a conjecture is either true or false, based on the argument's merits. One quasi-experimental study (#5631, *N* < 10) showed that participation in the Quality Talk PD helped high school chemistry and physics teachers to more effectively teach and promote discourse focused on substantive questions about core scientific phenomena and elaborate and appropriately critique responses to those questions, whereas another study (#78764, *N* = 37) showed that all elementary biology teachers who attended these PD sessions and their students made statistically significant improvements in their science discourse practices. A second observational study of middle school (*N* = 12) and high school (*N* = 14) science teachers (#35668) suggested that providing a storyline model that anchored the discussion to a specific phenomenon helped increase the number of open-ended questions teachers used in the classroom as well as the amount of student-student dialogue.

Summary. This group of projects provides evidence of effective approaches for supporting teachers' facilitation of argumentation in the classroom. Evidence from experimental and quasi-experimental studies suggests that teachers who receive PD on facilitating discussions are significantly more likely to use instructional strategies that encourage participation from multiple students and keep classroom discussions focused on core disciplinary concepts.

Interventions Focused on Change in Teacher Knowledge and Beliefs

Two projects focused on teachers' own growth in their understanding of argumentation and their beliefs about the teaching and learning of argumentation.

- #1119584 Constructing and Critiquing Arguments in Middle School Science Classrooms: Supporting Teachers With Multimedia Educative Curriculum Materials
- #1316347 Integrating Quality Talk Professional Development to Enhance Professional Vision and Leadership for STEM Teachers in High-Need Schools

Findings. Two research products focused on the outcomes of **PD interventions on teachers' own knowledge and beliefs** about the teaching and learning of argumentation. For example, one quasi-experimental study in high school chemistry and physics (#24924, $N < 10$) revealed that treatment teachers' discourse practices better reflected critical-analytic thinking and argumentation at posttest relative to comparison classrooms after receiving the Quality Talk PD. An observational (#68330, $N = 10$) study with middle school earth science teachers suggested that there were three main influences on teachers' curricular decision making in classes with higher quality argumentation: (a) teachers' understanding of argumentation as an epistemic practice (rather than surface-level features), (b) teachers as critically reflective curriculum users, and (c) teachers problematizing their prior teaching experiences.

Summary. There is limited evidence (few studies overall and no experimental or quasi-experimental research) about teachers' knowledge and beliefs. Considering the strong connection between teachers' beliefs and practices (Fang, 1996) and research suggesting that teachers feel particularly underprepared to teach argumentation compared with other NGSS science and engineering practices (Kang et al., 2018), understanding teacher beliefs could be an important element in ensuring that argumentation instruction takes place in the classroom.

What Was Learned That Could Inform the Development of Future Interventions?

Projects that conducted naturalistic studies examined the teaching and learning of argumentation in "business as usual" settings were grouped first based on the population that was the focus of the intervention (e.g., students or teachers), and then according to themes derived from the coding of each product (see Exhibit 12). These projects provide insight into students' and teachers' learning trajectories, beliefs, and approaches to argumentation, which could offer useful guidance for the development of future interventions.

As in the section on intervention studies, findings here are reported by theme, with the projects that fall under that theme reported in blue boxes and tagged with the NSF project number and project title. Within each theme, findings from individual products that were coded to that theme are reported, and also tagged with a study specific product number for reference (see Appendix B for a full list of projects and associated products). For each product, the study sample size, relevant discipline, and grade level are also reported.

Exhibit 12. Summary of Themes of What Was Learned From Naturalistic Studies That Could Inform the Development of Future Interventions

	Student focused	Teacher focused
Naturalistic observation	**Using theoretical frameworks to map student argumentation** (5 projects) **Uncovering student misconceptions and learning trajectories** (2 projects)	**Teacher beliefs about argumentation and student ability** (2 projects[a]) **Teacher pedagogical learning trajectories** (4 projects)

[a] These two projects contained products coded to both themes.

Student Naturalistic Studies

A set of naturalistic studies were focused on observations of student argumentation (seven projects). Overall, these projects sought to identify patterns in the way that students constructed their arguments and common pathways to understanding that students took when learning mathematical or scientific argumentation.

Using Theoretical Frameworks to Map Student Argumentation

Five projects examined student argumentation in naturalistic settings by using different theoretical frameworks and models of argumentation to identify and map out components of arguments as they took place in the classroom.

- #1149436 CAREER: Learning to Support Productive Collective Argumentation in Secondary Mathematics Classes
- #1253081 CAREER: Noticing and Using Students‘ Prior Knowledge in Problem-Based Instruction
- #1220623 The Role and Use of Examples in Learning to Prove
- #1119584 Constructing and Critiquing Arguments in Middle School Science Classrooms: Supporting Teachers With Multimedia Educative Curriculum Materials
- #1418423 GRIDS: Graphing Research on Inquiry With Data in Science

Findings. Two products used **Toulmin's model** (described in detail earlier): (a) one to document the different contributions (data, warrants, or claims) of students and teachers during a mathematics classroom discussion in high school geometry (#37794), and (b) one to identify how students used their prior knowledge about visual perspective to make and justify claims about geometric diagrams (#32744, *N* = 23).

Four products used the **Criteria, Affordances, Purposes, and Strategies (CAPS) framework** as a tool for examining students' and experts' criteria for selection and use of examples during conjecturing and proving activities in middle school (*N* = 12) and high school (*N* = 4) math (#64561). Findings suggest that "successful provers" (i.e., those who produced either a complete or incomplete but viable justification for why a conjecture was true in general) were more likely to (a) note the *affordances* of an example for generalization and gaining insight into the conjecture; (b) have *strategies* (e.g., pattern search, attempting to disprove) for using examples rather than just for choosing them; and (c) identify the *purpose* of the example as conveying a general argument (#5609). A second study (#54759) concluded that using a generic example (i.e., specific instances of a more general case) is more likely to be successful and that it is important for teachers to draw students' attention to the generalizable elements of the examples they provide. A final study (#71161) showed that although both students and experts used a *direct approach* (constructed examples to satisfy the hypothesis) and a *contradiction approach* (constructed examples that satisfy the hypothesis and the negation of the conclusion; i.e., counterexamples), only experts use a *contrapositive approach* (i.e., construct examples that satisfy the counterexample).

Frameworks used to study argumentation:

- Toulmin's model
- Criteria, Affordances, Purposes, and Strategies (CAPS)
- Knowledge integration (KI)
- Community of practice (CoP)
- Claims, evidence, and reasoning (CER)

Another product used the **knowledge integration (KI) framework** to develop a measure of graph comprehension, critique, and construction in middle school earth science and physics classrooms (#35709, *N* = 460). Using graphs is a core scientific competency that is integral to scientific argumentation, as graphs are often drawn upon as a source of evidence. The measure also revealed several challenges that students face when interpreting graphs embedded in a science context, including (a) interpretation of complex graphs, (b) critiquing graphs, (c) constructing graphs from narrative accounts of scientific phenomena, and (d) using scientific knowledge to interpret graphs.

Finally, one product used a **community of practice (CoP) framework** to investigate the relationship between students' argumentation in a single middle school, sheltered English immersion (SEI) biology classroom (#45935, *N* < 10). Findings suggest that student movement in and out of the immersion class influenced their opportunities for legitimate peripheral participation, specifically in that new students did not have a chance to interact with more advanced students who had transitioned out. Findings suggest that heterogenous grouping by experience level to allow for peer modeling and allowing students to use their first language as they acquire fluency in science practices can both support students in these classrooms.

Another framework, **claims, evidence, and reasoning (CER)**, also was used to guide a number of intervention studies in scientific argumentation (#5637, #5639, #57402). These products are reported on in more depth above, in the intervention section of this synthesis.

Summary. These products provide insight into core theoretical frameworks used by DRK-12 projects in mathematics and science to identify and map components of students' arguments as they occur in naturalistic settings. Although Toulmin's model is still often cited across studies on argumentation as a guiding and foundational framework, other frameworks, such as the CAPS framework, are becoming more widely used, particularly in math.

Uncovering Student Misconceptions and Learning Trajectories

Two projects focused on understanding student learning trajectories and common misconceptions that students encounter as they develop more sophisticated arguments.

- #1317034 Learning Algebra and Methods for Proving
- #1453493 CAREER: Proof in Secondary Classrooms: Decomposing a Central Mathematical Practice

Findings. Three research products identified **common misconceptions that students encounter when learning argumentation**. In one study (#56552, $N = 15$), researchers interviewed high-achieving high school students in geometry to better understand their approach to proving when confronted with four alternative proving tasks: a two-column proof, a diagramming task, a task in which students were asked to draw conclusions, and a task in which they were asked to determine a theorem from a completed proof. Students struggled to complete tasks that differed from what they had normally encountered in class. The authors conclude that students had turned proving into a rote task, in which they expected to mark a diagram and complete a written two-column proof in which some information was given to them, but struggled to recognize similar activities that did not look like what they thought of as "doing proof."

Two products identified **specific content and instructional strategies that target common misconceptions** and can help mitigate the difficulty they cause for learning argumentation. For example, explicit instruction prior to teaching proof about assumptions and drawing valid conclusions from diagrams can help prevent later confusion in high school geometry (#92500[5]). Another study drawing on data from a broader Learning Algebra and Methods of Proof (LAMP)

[5] The sample size for this study was not reported.

intervention (#84212) used a single case study to examine the relationship between middle school students' understanding of repeating decimals and rational numbers. The author used this case to suggest that the representation of rational numbers as repeated decimals (e.g., if .333… equals 1/3 or if .999… equals 1) can act as key content for engaging students in argumentation related to complex mathematical ideas.

Summary. These five research products identify common misconceptions that students encounter while learning argumentation. Interestingly, all these products were from mathematics projects, suggesting that DRK-12 research on scientific argumentation has placed less of an emphasis on student misconceptions. Also, all the products in this theme are related to argumentation as proof, suggesting that a focus on misconceptions may be particularly common under this definition of argumentation.

Teacher Naturalistic Studies

A similar number of projects conducted naturalistic observation of teachers' argumentation instruction in "business-as-usual" classroom settings. Overall, these projects aimed to understand how teachers learn to teach argumentation and how their beliefs about argumentation may influence their instructional practices.

Teacher Beliefs About Argumentation and Student Ability

Two projects focused on understanding how teachers' beliefs about the importance, benefits, and constraints to teaching argumentation influenced their instruction.

- #1119584 Constructing and Critiquing Arguments in Middle School Science Classrooms: Supporting Teachers With Multimedia Educative Curriculum Materials
- #1149436 CAREER: Learning to Support Productive Collective Argumentation in Secondary Mathematics Classes

Findings. A single research product focused on **how teachers' beliefs about argumentation influenced their instruction** in middle school earth science classrooms. Teachers in this study stated that their own learning goals had the most effect on their argumentation instruction, whereas context, policy, and assessment were all less important (#37635, N = 42). This was in part because teachers saw current school policies and assessments as misaligned with the learning goals of argumentation. Authors suggest that helping teachers identify what counts as argumentation and providing support to try argumentation in their classes may increase teachers' confidence and encourage integration of argumentation in instruction.

A second product (#43829, $N = 15$) of preservice algebra teachers differentiated between three different perspectives on student thinking with different sources: (a) a teacher perspective (drawing on professional training), (b) a student perspective (drawing on their own experience as a student), and (c) a combined perspective. Researchers suggest that the *student perspective* may be skewed based on teachers' subsequent educational experiences (i.e., an "expert blind spot") and could lead teachers to not present proofs at a developmentally appropriate level.

A third product showed that middle school biology teachers' beliefs about the benefit of argumentation, students' capabilities for argumentation, and their role and capacity to support argumentation instruction varied by the socioeconomic status (SES) of their students in science (#16454, $N = 34$). For example, some responses showed a possible deficit view of their students' abilities to engage in argumentation, especially teachers of students with low SES, ELs, students in special education, with teachers often describing their role as providing scaffolds to reduce the difficulty of argumentation for these students. Interestingly, teachers who believed that all students could engage in argumentation used scaffolding as an example of why all students *are* capable.

Summary. Products in this group identify common teacher beliefs about argumentation that influence their practice, including potential beliefs that may act as barriers to effective instruction. These beliefs are both internal (i.e., beliefs about students' abilities, teachers' "expert blind spot" that prevents them from seeing potential student misconceptions) and external (i.e., school support, alignment of curriculum with argumentation instruction).

Teacher Pedagogical Learning Trajectories

Four projects examined how both preservice and in-service teachers learn to teach argumentation.

- #1149436 CAREER: Learning to Support Productive Collective Argumentation in Secondary Mathematics Classes
- #1119584 Constructing and Critiquing Arguments in Middle School Science Classrooms: Supporting Teachers With Multimedia Educative Curriculum Materials
- #1316241 Developing Rich Media-Based Materials for Practice-Based Teacher Education
- #1350802 CAREER: L-MAP: Pre-service Middle School Teachers' Knowledge of Mathematical Argumentation and Proving

Findings. Four products focused on **how preservice teachers learn to teach argumentation** in teacher preparation programs. One study (#5650, *N* = 52) showed that preservice teachers in late elementary mathematics discussed mathematical argumentation from the perspective of the individual, with less focus on argumentation as a social activity. A second study (#5651, *N* = 37) suggested that teachers who were stronger at explaining their own solutions also were stronger at critiquing students' explanations. A study of preservice high school geometry teachers (#37794, *N* < 10) used Toulmin's model to map student contributions and teacher supports during a mathematical discussion, and identified specific question prompts teachers used to support students' argumentation. Importantly, asking "why" questions or asking students to explain their reasoning helped make warrants explicit and encouraged them to contribute warrants. A fourth study (#8663, *N* = 34) focused on the presence of three teaching competencies: (a) professional noticing of student mathematical reasoning and strategies, (b) the ability to assess the validity of student reasoning and strategies, and (c) the ability to select student strategies for class discussion. Results suggest that supporting preservice teachers' ability to notice the mathematically significant aspects of student reasoning and strategies can help them to better assess the validity of student reasoning and strategies and that selecting strategies with the purpose of engaging students in justifications can advance their conceptual understanding.

Teacher competencies for argumentation instruction:

- Noticing student reas1oning and strategies
- Assessing the validity of student strategies
- Selecting student strategies for class discussion
- Including language supports that:
  - facilitate dialogue
  - focus on argument structure
  - provide a rationale for argumentation practices
- Active (rather than passive) facilitation that solicits student explanations and does not take over their thinking 1

Another set of two products focused on **how in-service teachers learn to improve their argumentation instruction**. One study (#76223) focused on teaching argumentation with ELs in a single middle school biology classroom, and identified three components of teachers' practice that acted as supports for ELs: (a) including additional language supports focused on argument structure, (b) facilitating dialogic interactions with productive language supports, and (c) providing language supports that offer a rationale for argumentation. They further conclude that making rationales explicit by explaining the *why* behind science practices could be beneficial for all students, not just ELs. A second product (#47795, *N* = 20) reported on the use of an online application that allows middle school algebra teachers to create storyboard-like depictions of classroom scenarios. Results from coded depictions of teacher actions show two themes: (a) *passive facilitation of argumentation*, in which teachers used *statements* to take over the thinking, or the use of *silence* and *generating discussions*, which failed to elicit the key mathematical ideas from students; and (b) active facilitation of argumentation, in which teachers used *probing* and *orienting and focusing* to solicit student explanations. Active facilitation was positively correlated with teacher experience, suggesting that teachers with more classroom experience may be more likely to engage in active facilitation, whereas teachers who are less experienced may engage in more passive facilitation.

Summary. This group of products focuses on the teacher as a learner, providing insight into the trajectories teachers take as they learn to teach argumentation, and suggesting core competencies that teacher develop. The products include studies of both preservice teachers (four studies) and in-service teachers (two studies). A number of studies produce lists of potentially useful practices that teachers could implement in their argumentation instruction.

What Was Learned About Argumentation With Learners From Marginalized Groups in STEM?

We identified projects across the categories above with a specific focus on the teaching and learning of argumentation with students who are historically underrepresented in STEM. These included students from lower SES backgrounds, students from rural locations, students who are ELs, women, and students from historically marginalized racial groups (e.g., Black, Hispanic, and Native American students). These projects were identified by coding for studies that explored findings by demographic groups beyond simply identifying these groups in the sample population. For example, studies that only reported the racial demographics of a sample would not be coded as addressing the needs of diverse groups of learners, but studies in which the researcher explicitly discussed a pedagogical intervention that focused on engaging ELs in argumentation would be included. Under this definition, of the 23 projects, only three projects (#1119584, #1418019, and #1418423) explicitly studied approaches to teaching and learning argumentation that aimed to support students from marginalized groups. See the Project Spotlight 2 for more detail on one of these projects.

- #1119584 Constructing and Critiquing Arguments in Middle School Science Classrooms: Supporting Teachers With Multimedia Educative Curriculum Materials
- #1418019 Investigating How to Enhance Scientific Argumentation Through Automated Feedback in the Context of Two High School Earth Science Curriculum Units
- #1418423 GRIDS: Graphing Research on Inquiry With Data in Science

Findings. Three products discussed **potential barriers and supports for learning argumentation for ELs**. For example, one study (#45935, $N < 10$) that took place in a sheltered English immersion middle school biology classroom suggested that the transition of students into and out of the classroom disrupted their opportunities to learn argumentation from more experienced students. This study also suggested that it is important to allow students to use their first language as they acquire fluency in science practices and that the common practice of front-loading language-heavy content and vocabulary may be counterproductive in providing ELs with opportunities for scientific sensemaking. A pre-post study in high school earth science (#93829, *N* = 183) and an observational study in a middle school biology classroom (#76223) also provided evidence that interventions such as giving feedback on argumentation or making rationales explicit by explaining the "why" behind science practices were shown to benefit ELs and are likely to be beneficial for all students.

Two products emphasized **how students' SES influenced their opportunities to learn argumentation.** One study in middle school earth science and physics (#35709, *N* = 460) hypothesized that the lower scores observed on a measure of argumentation were related to reduced opportunities to learn for students from lower SES backgrounds. Another study (#16454, *N* = 34) found that middle school biology teachers of students from low-SES backgrounds were more likely to (a) feel that an emphasis on state testing influenced their ability to focus on argumentation instruction and (b) have a deficit view of their students' abilities to engage in argumentation. As mentioned earlier, some teachers discussed scaffolding argumentation as a way to reduce difficulty for their students, whereas others used scaffolding as an example of why all students are capable of argumentation.

Summary. Only three projects focus on understanding argumentation with learners from marginalized groups. This finding demonstrates a significant gap in the research on argumentation. More research is needed to understanding approaches to teaching and learning of argumentation that can best support students from groups that are historically marginalized in mathematics and science.

Project Spotlight 2: Supports and Barriers to Opportunities to Learn Argumentation for Students From Historically Marginalized Groups

Constructing and Critiquing Arguments in Middle School Science Classrooms: Supporting Teachers With Multimedia Educative Curriculum Materials (NSF award #1119584; total funded amount = $3.15 million)

Why Spotlight This Project?

This larger project investigated the effects of multimedia educative curricular materials on teachers' learning and beliefs about scientific argumentation. However, the project also explicitly focused on supports and potential barriers for the teaching and learning of argumentation with students from groups that have been historically marginalized in mathematics and science (e.g., English learners [ELs]; students from lower socioeconomic status [SES] backgrounds).

What Was Studied?

The project consisted of naturalistic observations focused on middle school teachers and their students in biology and earth science classrooms, using both a sociocultural and structural (non-Toulmin) definition of argumentation.

What Was Found?

- **Supports for ELs.** González-Howard et al. (2017) conducted a case study of a science teacher's instructional strategies for argumentation in an EL classroom. The teacher provided language supports in three ways: (a) supporting language around the structure of the argument (e.g., pointing out specific elements of an argument); (b) facilitating dialogic interactions (e.g., providing sentence frames and having students work in peer groups); and (c) offering a rationale for argumentation practices (e.g., using analogy to demonstrate the meaning and importance of "relevance" in argumentation).
- **Barriers for ELs.** Gonzalez-Howard et al. (2016) reported how the structure of an SEI classroom limited students' opportunities to engage in productive argumentation because more experienced students were moved out of the classroom as they developed English proficiency.
- **Barriers for lower SES students.** Katsh-Singer et al. (2016) also investigated teachers' beliefs about students' abilities to engage in argumentation. Teachers of students from lower SES backgrounds were more likely to hold a deficit view of students' abilities and either not teach or heavily scaffold argumentation to reduce the perceived difficulty of the content. They also suggest that teachers in schools that serve students from lower SES backgrounds may experience more pressure to reach state testing requirements and therefore not include lessons on argumentation if argumentation is not seen as a skill that supports test taking.

Summary of What Was Learned

In the following sections, we provide a high-level overview of the findings across both intervention and naturalistic studies as they related to the four lines of research identified from the literature review and outlined in the introduction of this report: (a) student learning trajectories in argumentation, (b) teacher beliefs and practices for teaching argumentation, (c) disciplinary differences in mathematical and scientific argumentation, and (d) barriers and support to the teaching and learning of argumentation for individuals from historically marginalized groups in STEM. This review provided a synthesis of 23 DRK-12 research projects on argumentation in science and mathematics, which together produced 54 research products identified for review. The largest number of these products reported findings from student-focused intervention studies, with a number also reporting on teacher-focused interventions, and naturalistic studies of teaching and learning of argumentation. Several of these products reported on the use of educational technologies that supported both argumentation instruction and student learning, whereas a relatively small number focused on considerations for the teaching and learning of argumentation for groups historically marginalized in STEM.

Research Topic 1: Student Learning Trajectories in Argumentation

Student-focused projects made up the bulk of the studies reviewed in this synthesis. Both intervention projects and naturalistic projects of student argumentation suggested the importance of providing structure for students' nascent forms of argumentation. For example, eight naturalistic studies of student argumentation suggested four different frameworks that could be used to map students' arguments: (a) Toulmin's model, (b) the CAPS framework, (c) the KI framework, and (d) argumentation as a CoP. A fifth framework, CER, was often used in intervention studies of scientific argumentation. Evidence from two experimental studies shows that digital technologies that support making the structure of students' arguments explicit could be a potentially effective intervention for supporting students' learning trajectories in argumentation.

Experimental evidence also showed positive effects of tools that supported argumentative discourse in the classroom. Many of these studies included teacher-focused interventions, suggesting that learning to facilitate classroom discussions for productive argumentation is important. Interestingly, an experimental evaluation of student-student interactions in an online forum provided useful evidence of the benefits of teachers' facilitation of student discourse. Such platforms can perhaps help to scaffold some of the complexities of in-person facilitation of argumentation discourse for teachers.

Research Topic 2: Teacher Practices and Beliefs About Teaching Argumentation

A set of intervention and naturalistic studies reported findings about preservice teachers and in-service teachers as learners. Studies offered a wide array of potentially "high-leverage" practices that teachers employed, including different talk formats (e.g., instruction, whole classroom/small group, pair, individual), activities (e.g., charting ideas, "four corners," cartoons, T-charts), teacher discourse moves (e.g., ask, support, press, link), question prompts (both closed- and open-ended; requesting a factual answer, method, idea, elaboration, or evaluation), and approaches to encouraging participation from multiple students (e.g., providing language supports, providing a rationale behind scientific practices, considering both one's own educational experiences and what students may or may not know). These reported findings demonstrate the abundance of practices encountered by teachers looking for advice about how to teach argumentation. More studies are needed to provide experimental evidence of the effectiveness of these practices and to understand how teachers learn to adaptively apply these practices in context.

Although both naturalistic and intervention studies suggested the importance of teachers' beliefs on their argumentation instruction, no studies provided experimental evidence of the impact of teachers' beliefs on their practice or student learning. Most studies reported naturalistic observations that examined the relationship between teachers' beliefs and practice and the implications for students' equitable opportunities to learn. Relatively few studies focused on changing teachers' beliefs about argumentation through their engagement in PD interventions.

Research Topic 3: Disciplinary Similarities and Differences in Mathematics and Science

Some findings demonstrated components of interventions specific to mathematics or science, whereas others showed some evidence of effects across both. Naturalistic studies suggested that certain definitions of argumentation (e.g., as "proof" in mathematics or as "evidence-based reasoning" in science) were discipline specific, as well as certain frameworks for identifying components of arguments (e.g., the CAPS framework in math and the CER framework in science). The content of some subdisciplines unique to mathematics and science also appeared to be more amenable to the study of argumentation, though for different reasons. For example, a number of studies on scientific argumentation took place in earth science and biology courses, which address topics of contemporary scientific debates (e.g., climate change, evolution). However, many studies in mathematics were conducted in geometry courses, due to the historical relationship of argumentation in the development of that discipline (e.g., Euclidian construction, geometric proofs).

Across both math and science, experimental evidence from two student intervention studies (one in math and one in science) suggest that tools that provide opportunities and scaffolds for students to engage in productive argumentative discourse could be effective in supporting students' learning of argumentation in both disciplines. However, for student-focused interventions, only studies in mathematics focused on identifying common student misconceptions, suggesting that there may be less emphasis on this particular line of research in scientific argumentation. This may reflect a growing emphasis in science education research to consider students' naïve ideas about science as opportunities for sensemaking rather than as barriers or deficits to understanding (Campbell et al., 2016).

Research Topic 4: Supporting Argumentation With Students From Marginalized Groups

Few projects in our review emphasized argumentation with students from marginalized groups beyond reporting sample demographic information. Of the three projects that did focus on this topic, the groups that were most commonly included were ELs and students from low SES backgrounds. Studies focused largely on the effect of the learning context on students' ability and opportunities to learn argumentation, both in terms of the classroom structure (i.e., English immersion courses) and the influence of teacher beliefs about their abilities to learn argumentation. These studies also provided insight into specific barriers and supports for groups with different experiences of marginalization in STEM. For example, EL students experienced challenges to learning argumentation as a result of changing enrollments in English immersion classrooms, while students from low-SES backgrounds experienced barriers in access to argumentation instruction because of the emphasis placed in their schools on preparing for state accountability measures that do not directly test for argumentation.

Opportunities for Future Research

In this section, we will highlight some of the important gaps in the DRK-12 portfolio of projects on argumentation and provide suggestions for future prioritization based on our syntheses of recent projects.[6] Our recommendations focus both on the existence of relatively strong experimental evidence to support the continued implementation and development of certain lines of research, and areas of common interest in which there are naturalistic studies that could inform the development of future intervention studies.

Opportunity 1: Rigorous Study of the Use of Technological Tools for Argumentation

One of the key findings from this report was that a number of projects reported on the design and implementation of technological tools that helped support the teaching and learning of

[6] Two external content experts also contributed significantly to this synthesis including these reflections: Jennifer Knudsen at TERC (expert on argumentation in mathematics) and Katherine McNeill at Boston College (expert on argumentation in science). However, this section does not necessarily reflect their endorsement of our reflections.

argumentation. More than one fourth of the projects (six of 23) used some form of educational technology as a means to support teacher instruction or student understanding in argumentation. Four main functions of these tools emerged as a core set of possible design features that can provide guidance for future development and study of technologies that support argumentation. Experimental studies indicated positive effect of two types of tools: (1) tools that help make the structure of students' argument explicit, and (2) tools that support opportunities for students to engage in argumentative discourse. Pre-post, correlational, or observational studies reported positive findings for two other types of tools: (1) tools that provide students with feedback and opportunities to revise their arguments, and (2) tools that provide multiple representations of arguments. Experimental studies are needed to determine if these other forms of tools also show significant impacts on teaching or student learning of argumentation. Importantly, rigorous studies designed to test specific functions of these tools can contribute to a theoretical framework for argumentation by identifying the underlying learning mechanisms that these tools help support and provide guidance for the development of new tools to support argumentation.

Opportunity 2: Pairing Naturalistic Studies With Quantitative Methodologies

An important finding from this synthesis is the depth, value, and contribution of naturalistic studies of argumentation to understanding the forms that teaching and learning of argumentation take in the absence of any interventions. These studies identify important questions and potentially fruitful areas for additional research and provide the basis for the design of future interventions and additional studies of argumentation. Although the methods of these descriptive studies reflect an intentional emphasis on the design and development goals of the DRK-12 program, they do not provide evidence of the effectiveness of argumentation interventions. Therefore, in addition to the continued contribution of naturalistic studies to the DRK-12 portfolio, we also note that it is important to complement or extend these studies with quantitative methodologies and experimental designs that can provide evidence for the effect of interventions in argumentation.

These efforts may be supported by measurement studies that work on developing quantitative measures of specific argumentation skills and competencies. For example, project #1316057 used Likert-scale survey measures to assess students' content knowledge and understanding of discipline-specific argumentation practices. Knowledge items asked students to rate how closely a scientist would agree with a particular statement (e.g., "Earthquakes are caused by slips in Earth's crustal plates"), thereby reflecting perceptions of scientific consensus rather than their own beliefs of knowledge on the topic. A second set of items had students rate the plausibility of two scientific models (i.e., from "greatly implausible or even impossible" to "highly plausible") based on the available evidence. Such measures may be a fruitful avenue for future studies seeking to quantitatively measure argumentation as a form of disciplinary practice.

Opportunity 3: Studying Convergences in Mathematical and Scientific Argumentation

This synthesis provided some empirical evidence of argumentation interventions that could be applied across both mathematics and science. For example, findings suggest that intervention that make the structure of students' arguments explicit may be productive in both mathematics and science. Studies that take a cross-disciplinary approach to argumentation or aim to directly compare such interventions in mathematics and science could help in the understanding of what approaches to teaching and learning argumentation can be applied across contexts and which are grounded in a particular discipline.

Our synthesis also found evidence that supporting teachers' facilitation of student discourse could improve argumentation outcomes but was more common in science. In part, developing lessons that involve students in authentic scientific debates on topics like climate change or evolution may be more obvious in scientific subdisciplines like earth science and biology. However, while historically mathematical disciplines like geometry have taught argumentation through other forms like the two-column proof, evidence suggests that students are often unable to transfer this knowledge to other contexts, and that engaging in discourse could be a productive learning opportunity for students. Studies that support teachers in developing opportunities for students to use authentic mathematical practices to engage in discourse around relevant topics (e.g., the Monty Hall problem; does .999… = 1?) could provide can opportunity to improve the evidence base on this approach to teaching argumentation in mathematics.

Opportunity 4: Studies of Argumentation in Early Elementary and PreK

A notable gap in studies of argumentation across both mathematics and science is research on argumentation with early elementary and preK students, although earlier work in mathematics not included in the time frame for this synthesis has explored this area (e.g., Yackel & Cobb, 1996). Although this scarcity may arise from the assumption that preschoolers are too young to engage in the kind of inferential thinking required for argumentation, recent research suggests that scientific argumentation can be observed in participants as young as 3 years old (Convertini, 2021). Additional naturalistic and intervention studies are required in this setting to understand the extent to which these younger students in both mathematics and science exhibit nascent forms of argumentation and how educators can best support their learning.

Opportunity 5: Understanding Barriers and Supports for Teaching and Learning Argumentation for Students Experiencing Marginalization in STEM

Finally, an important finding from the synthesis of these studies is that despite the importance and specific call for proposals that focus on broadening participation in STEM, this area has been relatively understudied in the recent DRK-12 portfolio of projects related to argumentation. In particular, very few studies were found to reach the threshold of our definition of attending to diverse populations that went beyond the simple reporting of sample demographics. This area is particularly important because it is cross-cutting and underlies all the other categories; understanding the barriers and supports that are particularly important for marginalized communities is integral to the success of other intervention efforts. Despite a wealth of literature on the beneficial effect of culturally responsive practices for improving the attitudes, achievement, and persistence of students from racial groups that have been historically marginalized in STEM (Brown et al., 2019), no studies explicitly investigated how these practices may be useful in the teaching and learning of argumentation.

Identifying ways to provide equitable access to opportunities to participate in and learn argumentation as a core practice in these two disciplines seems essential for the goal of broadening participation in STEM. In addition, research that develops and studies culturally relevant and responsive approaches to teaching argumentation is needed to provide further insight into aspects of teaching and learning argumentation that will benefit all students and teachers of mathematical and scientific argumentation.

Conclusions

Although the importance of argumentation is well documented in the research literature and core standards documents for both mathematics and science, many challenges to the teaching and learning of argumentation in preK–12 classrooms still exist. This review synthesized insights from 23 DRK-12 projects on argumentation in mathematics and science which totaled more than $40.5 million in awards, and identified several key areas of research in scientific and mathematical argumentation that could provide fruitful avenues for new or continued lines of DRK-12 research. Specifically, new digital technologies could offer effective ways to support students in organizing their ideas and making the structure of their arguments explicit. In addition, striking a balance between naturalistic studies that inform new areas for research and rigorous experimental designs that provide strong evidence of effectiveness will be important for strengthening the DRK-12 portfolio in the future. This synthesis also provided a first step toward understanding aspects of teaching and learning argumentation in preK–12 that may be discipline-specific and which frameworks, interventions, and naturalistic forms of argumentation may be common across science and mathematics. Further research in this area could provide key insights that could advance argumentation research in both fields and provide a shared language for understanding the teaching and learning of argumentation. Finally, continued emphasis on barriers and supports for teaching and learning of argumentation for students from groups that have been historically marginalized would support the mission of the DRK-12 portfolio to significantly enhance preK–12 STEM teaching and learning and the larger mission of NSF to broaden participation in STEM fields.

References

Ainsworth, S. (2006). DeFT: A conceptual framework for considering learning with multiple representations. *Learning and Instruction*, *16*(3), 183–198.

Alqahtani, M. M., & Powell, A. B. (2015). Co-action and dynamic geometry knowledge. In K. Beswick, T. Muir, & J. Wells (Eds.), *Proceedings of 39th psychology of mathematics education conference* (Vol. 2, pp. 17–24). PME.

Aricha-Metzer, I., & Zaslavsky, O. (2019). The nature of students' productive and non-productive example-use for proving. *The Journal of Mathematical Behavior*, *53*, 304–322. https://doi.org/10.1016/j.jmathb.2017.09.002

Ault, M., Craig-Hare, J., Frey, B., Ellis, J. D., & Bulgren, J. (2015) The effectiveness of Reason Racer, a game designed to engage middle school students in scientific argumentation. *Journal of Research on Technology in Education, 47*(1), 21–40. https://doi.org/10.1080/15391523.2015.967542

Banegas, J. A. (2013). Argumentation in mathematics. In A. Aberdein & I. J. Dove (Eds.), *The argument of mathematics* (pp. 47–60). Springer. https://doi.org/10.1007/978-94-007-6534-4

Bathgate, M., Crowell, A., Schunn, C., Cannady, M., & Dorph, R. (2015). The learning benefits of being willing and able to engage in scientific argumentation. *International Journal of Science Education*, *37*(10), 1590–1612. https://doi.org/10.1080/09500693.2015.1045958

Bell, P., & Linn, M. C. (2000). Scientific arguments as learning artifacts: Designing for learning from the web with KIE. *International Journal of Science Education*, *22*(8), 797–817. https://doi.org/10.1080/095006900412284

Berland, L. K., & Hammer, D. (2012). Framing for scientific argumentation. *Journal of Research in Science Teaching*, *49*(1), 68–94. https://doi.org/10.1002/tea.20446

Berland, L. K., & McNeill, K. L. (2010). A learning progression for scientific argumentation: Understanding student work and designing supportive instructional contexts. *Science Education*, *94*(5), 765–793. https://doi.org/10.1002/sce.20402

Berland, L. K., & McNeill, K. L. (2012). For whom is argument and explanation a necessary distinction? A response to Osborne and Patterson. *Science Education*, *96*(5), 808–813. https://doi.org/10.1002/sce.21000

Berland, L. K., McNeill, K. L., Pelletier, P., & Krajcik, J. (2017). Engaging in argument from evidence. In C. V. Schwarz, C. Passmore, & B. J. Reiser (Eds.), *Helping students make sense of the world using next generation science and engineering practices* (pp. 229–257). National Science Teachers Association.

Berland, L. K., & Reiser, B. J. (2009). Making sense of argumentation and explanation. *Science Education*, *93*(1), 26–55. https://doi.org/10.1002/sce.20286

Berson, E., Borko, H., Million, S., Khachatryan, E., & Glennon, K. (2015). Practice what you teach: A video-based practicum model of professional development for elementary science teachers. *Orbis*, *2*, 35–53. https://doi.org/10.14712/23363177.2015.79

Boero, P., Douek, N., Morselli, F., & Pedemonte, B. (2010). Argumentation and proof: A contribution to theoretical perspectives and their classroom implementation. In M. M. F. Pinto & T. F. Kawasaki (Eds.), *Proceedings of the 34th conference of the International Group for the Psychology of Mathematics Education* (vol. 1, pp. 179–204). PME.

Brown, B. A., Boda, P., Lemmi, C., & Monroe, X. (2019). Moving culturally relevant pedagogy from theory to practice: Exploring teachers' application of culturally relevant education in science and mathematics. *Urban Education, 54*(6), 775–803. https://doi.org/10.1177/0042085918794802

Brown, J. S., Collins, A., & Duguid, P. (1989). Situated cognition and the culture of learning. *Educational Researcher*, *18*(1), 32–42. https://doi.org/10.3102/0013189x018001032

Campbell, T. G., Boyle, J. D., & King, S. (2020). Proof and argumentation in K-12 mathematics: A review of conceptions, content, and support. *International Journal of Mathematical Education in Science and Technology*, *51*, 754–774. https://doi.org/10.1080/0020739X.2019.1626503

Campbell, T., Schwarz, C. V., & Windschitl, M. (2016). What we call misconceptions may be necessary stepping-stones toward making sense of the world. *Science & Children, 83*(3), 69–74. https://doi.org/10.2505/4/sc16_053_07_28

Campbell, T. G., & Zelkowski, J. (2020). Technology as a support for proof and argumentation: A systematic literature review. *International Journal for Technology in Mathematics Education*, *27*(2), 113–124.

Chi, M. T. H., Bassok, M., Lewis, M. W., Reimann, P., & Glaser, R. (1989). Self-explanations: How students study and use examples in learning to solve problems. *Cognitive Science*, *13*(2), 145–182. https://doi.org/10.1207/s15516709cog1302_1

Chinn, C. A. (2006). Learning to argue. In A. M. O'Donnell, C. E. Hmelo-Silver, & G. Erkens (Eds.), *Collaborative learning, reasoning, and technology* (pp. 355–383). Erlbaum.

Cirillo, M., & Hummer, J. (2019). Addressing misconceptions in secondary geometry proof. *The Mathematics Teacher*, *112*(6), 410–417. https://doi.org/10.5951/mathteacher.112.6.0410

Civil, M., & Hunter, R. (2015). Participation of non-dominant students in argumentation in the mathematics classroom. *Intercultural Education*, *26*(4), 296–312. https://doi.org/10.1080/14675986.2015.1071755

Community for Advancing Discovery Research in Education (CADRE). (2014). *A targeted study of gaming and simulation projects in DR K-12.* https://cadrek12.org/sites/default/files/A%20Targeted%20Study%20of%20Gaming%20and%20Simulation%20Projects%20in%20DR%20K-12.pdf

Convertini, J. (2021). An interdisciplinary approach to investigate preschool children's implicit inferential reasoning in scientific activities. *Research in Science Education, 51,* 171–186. https://doi.org/10.1007/s11165-020-09957-3

Craig-Hare, J., Ault, M., & Rowland, A. (2017). The effect of socioscientific topics on discourse within an online game designed to engage middle school students in scientific argumentation. *Journal of Education in Science, Environment and Health (JESEH), 3*(2), 110–125. https://doi.org/10.21891/jeseh.325783

Cross, D. I. (2008). Creating optimal mathematics learning environments: Combining argumentation and writing to enhance achievement. *International Journal of Science and Mathematics Education*, *7*(5), 905–930. https://doi.org/10.1007/s10763-008-9144-9

Driver, R., Newton, P., & Osborne, J. (2000). Establishing the norms of scientific argumentation in classrooms. *Science Education*, *84*(3), 287–312. https://doi.org/10.1002/(sici)1098-237x(200005)84:3<287::aid-sce1>3.0.co;2-a

Durand-Guerrier, V., Boero, P., Douek, N., Epp, S. S., & Tanguay, D. (2012). Argumentation and proof in the mathematics classroom. In G. Hanna & M. de Villiers (Eds.), *Proof and proving in mathematics education* (pp. 349–367). Springer.

Ellis, A. B., Ozgur, Z., Vinsonhaler, R., Dogan, M. F., Carolan, T., Lockwood, E., Lynch, A., Sabouri, P., Knuth, E., & Zaslavsky, O. (2019). Student thinking with examples: The criteria-affordances-purposes-strategies framework. *The Journal of Mathematical Behavior*, *53*, 263–283. https://doi.org/10.1016/j.jmathb.2017.06.003

Erduran, S., Simon, S., & Osborne, J. (2004). TAPping into argumentation: Developments in the application of Toulmin's Argument Pattern for studying science discourse. *Science Education*, *88*(6), 915–933. https://doi.org/10.1002/sce.20012

Fang, Z. (1996). A review of research on teacher beliefs and practices. *Educational Research, 38*(1), 47–65. https://doi.org/10.1080/0013188960380104

Fishman, E. J., Borko, H., Osborne, J., Gomez, F., Rafanelli, S., Reigh, E., Tseng, A., Million, S., & Berson, E. (2017). A practice-based professional development program to support scientific argumentation from evidence in the elementary classroom. *Journal of Science Teacher Education*, *28*(3), 222–249. https://doi.org/10.1080/1046560x.2017.1302727

Flick, U. (2009). *An introduction to qualitative research*. Sage Publications.

Ford, M. J. (2012). A dialogic account of sense-making in scientific argumentation and reasoning. *Cognition and Instruction*, *30*(3), 207–245. https://doi.org/10.1080/07370008.2012.689383

Ford, M. J., & Wargo, B. M. (2011). Dialogic framing of scientific content for conceptual and epistemic understanding. *Science Education*, *96*(3), 369–391. https://doi.org/10.1002/sce.20482

Fukawa-Connelly, T., & Silverman, J. (2015). The development of mathematical argumentation in an unmoderated, asynchronous multi-user dynamic geometry environment. *Contemporary Issues in Technology and Teacher Education*, *15*(4), 445–488. https://doi.org/10.34944/dspace/6278

Garrett, R., Citkowicz, M., & Williams, R. (2019). How responsive is a teacher's classroom practice to intervention? A meta-analysis of randomized field studies. *Review of Research in Education*, *43*(1), 106–137. https://doi.org/10.3102/0091732X19830634

González-Howard, M., & McNeill, K. L. (2016). Learning in a community of practice: Factors impacting English-learning students' engagement in scientific argumentation. *Journal of Research in Science Teaching*, *53*(4), 527–553. https://doi.org/10.1002/tea.21310

González-Howard, M., & McNeill, K. L. (2020). Acting with epistemic agency: Characterizing student critique during argumentation discussions. *Science Education*, *104*(6), 953–982. https://doi.org/10.1002/sce.21592

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. https://doi.org/10.1080/09500693.2017.1294785

Greeno, J. G., & van de Sande, C. (2007). Perspectival understanding of conceptions and conceptual growth in interaction. *Educational Psychologist*, *42*(1), 9–23. https://doi.org/10.1080/00461520709336915

Haddaway, N. R., Collins, A. M., Coughlin, D., & Kirk, S. (2015). The role of Google Scholar in evidence reviews and its applicability to grey literature searching. *PLoS One*, *10*(9):e0138237. https://doi.org/10.1371/journal.pone.0138237

Henderson, J. B., McNeill, K. L., González-Howard, M., Close, K., & Evans, M. (2018). Key challenges and future directions for educational research on scientific argumentation. *Journal of Research in Science Teaching, 55*(1), 5–18.

Inglis, M., Mejia-Ramos, J. P., & Simpson, A. (2007). Modelling mathematical argumentation: The importance of qualification. *Educational Studies in Mathematics*, *66*(1), 3–21. https://doi.org/10.1007/s10649-006-9059-8

Jiménez-Aleixandre, M. P., & Erduran, S. (2007). 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). Springer.

Kang, E. J. S., Donovan, C., & McCarthy, M. J. (2018). Exploring elementary teachers' pedagogical content knowledge and confidence in implementing the NGSS science and engineering practices. *Journal of Science Teacher Education*, *29*(1), 9–29. https://doi.org/10.1080/1046560x.2017.1415616

Kannan, P., Zapata-Rivera, D., Mikeska, J., Bryant, A., Long, R., & Howell, H. (2018). Providing formative feedback to pre-service teachers as they practice facilitation of high-quality discussions in simulated mathematics and science methods classrooms. In E. Langran & J. Borup (Eds.), *Proceedings of Society for Information Technology and Teacher Education International Conference* (pp. 1570–1575). Association for the Advancement of Computing in Education (AACE). https://www.learntechlib.org/primary/p/182737/

Katsh-Singer, R., McNeill, K. L., & Loper, S. (2016). Scientific argumentation for all? Comparing teacher beliefs about argumentation in high, mid, and low socioeconomic status schools. *Science Education, 100*(3), 410–436. https://doi.org/10.1002/sce.21214

Kelly, G. J., & Takao, A. (2002). Epistemic levels in argument: An analysis of university oceanography students' use of evidence in writing. *Science Education*, *86*(3), 314–342. https://doi.org/10.1002/sce.10024

Knudsen, J., Lara-Meloy, T., Stevens, H. S., & Rutstein, D. W. (2014). Advice for mathematical argumentation. *Mathematics Teaching in the Middle School*, *19*(8), 494–500. https://doi.org/10.5951/mathteacmiddscho.19.8.0494

Knudsen, J., Stevens, H. S., Lara-Meloy, T., Kim, H. J., & Schechtman, N. (2017). *Mathematical argumentation in middle school—the what, why, and how: A step-by-step guide with activities, games, and lesson planning tools*. Corwin Mathematics.

Kosko, K. W., Rougee, A., & Herbst, P. (2014). What actions do teachers envision when asked to facilitate mathematical argumentation in the classroom? *Mathematics Education Research Journal*, *26*(3), 459–476. https://doi.org/10.1007/s13394-013-0116-1

Lynch, K., Hill, H. C., Gonzalez, K. E., & Pollard, C. (2019). Strengthening the research base that informs STEM instructional improvement efforts: A meta-analysis. *Educational Evaluation and Policy Analysis*, *41*(3), 260–293. https://doi.org/10.3102/0162373719849044

Manz, E. (2015). Representing student argumentation as functionally emergent from scientific activity. *Review of Educational Research*, *85*(4), 553–590. https://doi.org/10.3102/0034654314558490

Marsden, E., & Torgerson, C. J. (2012). Single group, pre- and post-test research designs: Some methodological concerns, *Oxford Review of Education, 38*(5), 583–616. https://doi.org/10.1080/03054985.2012.731208

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. https://doi.org/10.1002/tea.21252

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. https://doi.org/10.1002/sce.21274

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. https://doi.org/10.1002/sce.21081

McNeill, K. L., & Krajcik, J. (2007). Middle school students' use of appropriate and inappropriate evidence in writing scientific explanations. In M. C. Lovett & P. Shah (Eds.), *Thinking with data (Carnegie Mellon Symposium on Cognition Series)* (pp. 233–265). Erlbaum.

Means, M. L., & Voss, J. F. (1996). Who reasons well? Two studies of informal reasoning among children of different grade, ability, and knowledge levels. *Cognition and Instruction*, *14*(2), 139–178. https://doi.org/10.1207/s1532690xci1402_1

Mejía-Ramos, J. P., & Inglis, M. (2009). What are the argumentative activities associated with proof? *Research in Mathematics Education*, *11*(1), 77–78. https://doi.org/10.1080/14794800902732258

Murphy, P. K., Greene, J. A., Allen, E., Baszczewski, S., Swearingen, A., Wei, L., & Butler, A. M. (2018). Fostering high school students' conceptual understanding and argumentation performance in science through *Quality Talk* discussions. *Science Education*, *102*(6), 1239–1264. https://doi.org/10.1002/sce.21471

National Center for Science and Engineering Statistics. (2021). *Women, minorities, and persons with disabilities in science and engineering: 2021* (Special Report NSF 21-321). National Science Foundation. Available at https://ncses.nsf.gov/wmpd

National Governors Association Center for Best Practices & Council of Chief State School Officers. (2010). *Common Core State Standards for mathematics*. http://www.corestandards.org/wp-content/uploads/Math_Standards1.pdf

Newton, P., Driver, R., & Osborne, J. (1999). The place of argumentation in the pedagogy of school science. *International Journal of Science Education*, *21*(5), 553–576. https://doi.org/10.1080/095006999290570

NGSS Lead States. (2013). *New Generation Science Standards: For states, by states* (appendixes). National Academies Press.

Nussbaum, E. M. (2008). Collaborative discourse, argumentation, and learning: Preface and literature review. *Contemporary Educational Psychology*, *33*(3), 345–359. https://doi.org/10.1016/j.cedpsych.2008.06.001

Osborne, J. F. (2010). Arguing to learn in science: The role of collaborative, critical discourse. *Science*, *328*(5977), 463–466. https://doi.org/10.1126/science.1183944

Osborne, J. F., Henderson, J. B., MacPherson, A., Szu, E., Wild, A., & Yao, S.-Y. (2016). The development and validation of a learning progression for argumentation in science. *Journal of Research in Science Teaching*, *53*(6), 821–846. https://doi.org/10.1002/tea.21316

Osborne, J. F., & Patterson, A. (2011). Scientific argument and explanation: A necessary distinction? *Science Education*, *95*(4), 627–638. https://doi.org/10.1002/sce.20438

Pedemonte, B. (2007). How can the relationship between argumentation and proof be analysed? *Educational Studies in Mathematics*, *66*(1), 23–41. https://doi.org/10.1007/s10649-006-9057-x

Rapanta, C., Garcia-Mila, M., & Gilabert, S. (2013). What is meant by argumentative competence? An integrative review of methods of analysis and assessment in education. *Review of Educational Research*, 83(4), 483–520. https://doi.org/10.3102/0034654313487606

Rowland, A., Craig-Hare, J., Ault, M., Ellis, J., & Bulgren, J. (2017) Social media: How the next generation can practice argumentation, *Educational Media International, 54*(2), 99–111, https://doi.org/10.1080/09523987.2017.1362818

Russell, S. J., Schifter, D., & Bastable, V. (2011). Developing algebraic thinking in the context of arithmetic. In J. Cai & E. Knuth (Eds.), *Early algebraization: A global dialogue from multiple perspectives* (pp. 43–69). Springer-Verlag.

Sampson, V., & Blanchard, M. R. (2012). Science teachers and scientific argumentation: Trends in views and practice. *Journal of Research in Science Teaching*, *49*(9), 1122–1148. https://doi.org/10.1002/tea.21037

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*(3), 447–472. https://doi.org/10.1002/sce.20276

Sampson, V., & Clark, D. B. (2009). The impact of collaboration on the outcomes of scientific argumentation. *Science Education*, *93*(3), 448–484. https://doi.org/10.1002/sce.20306

Sandoval, W. A., & Millwood, K. A. (2005). The quality of students' use of evidence in written scientific explanations. *Cognition and Instruction*, *23*(1), 23–55. https://doi.org/10.1207/s1532690xci2301_2

Scheuer, O., Loll, F., Pinkwart, N., & McLaren, B. M. (2010). Computer-supported argumentation: A review of the state of the art. *International Journal of Computer-Supported Collaborative Learning*, *5*(1), 43–102. https://doi.org/10.1007/s11412-009-9080-x

Schwarz, B. B., Prusak, N., & Hershkowitz, R. (2009). Argumentation and mathematics. In K. Littleton & C. Howe (Eds.), *Educational dialogues: Understanding and promoting productive interaction* (pp. 103–127). Routledge.

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. https://doi.org/10.1080/09500690500336957

Singletary, L. M., & Conner, A.-M. (2015). Focusing on mathematical arguments. *The Mathematics Teacher*, *109*(2), 143–147. https://doi.org/10.5951/mathteacher.109.2.0143

Staples, M., Newton, J., Kosko, K. W., Conner, A.-M., Cirillo, M., & Bieda, K. (2016). Conceptions and consequences of what we call argumentation, justification, and proof. In M. B. Wood, E. E. Turner, M. Civil, & J. A. Eli (Eds.), *Proceedings of the 38th annual meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education* (pp. 1704–1712). The University of Arizona. http://www.pmena.org/pmenaproceedings/PMENA%2038%202016%20Proceedings.pdf

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

Walshaw, M., & Anthony, G. (2008). The teacher's role in classroom discourse: A review of recent research into mathematics classrooms. *Review of Educational Research*, *78*(3), 516–551. https://doi.org/10.3102/0034654308320292

Walton, D. N. (1996). *Argumentation schemes for presumptive reasoning*. Lawrence Erlbaum.

Weber, K. (2001). Student difficulty in constructing proofs: The need for strategic knowledge. *Educational Studies in Mathematics, 48*(1), 101–119.

Wecker, C., & Fischer, F. (2014). Where is the evidence? A meta-analysis on the role of argumentation for the acquisition of domain-specific knowledge in computer-supported collaborative learning. *Computers & Education*, *75*, 218–228. https://doi.org/10.1016/j.compedu.2014.02.016

Westgate, M. J. (2019). revtools: An R package to support article screening for evidence synthesis. *Research Synthesis Methods*, *10*(4), 606–614. https://doi.org/10.1002/jrsm.1374

What Works Clearinghouse. (2020). *What Works Clearinghouse standards handbook* (Version 4.1). U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance. https://ies.ed.gov/ncee/wwc/handbooks

Wickham, H. (2020). *rvest: Easily harvest (scrape) web pages* (R package Version 0.3.6) [Computer software]. https://CRAN.R-project.org/package=rvest

Wilson-Lopez, A., Strong, A. R., Hartman, C. M., Garlick, J., Washburn, K. H., Minichiello, A., Weingart, S., & Acosta-Feliz, J. (2020). A systematic review of argumentation related to the engineering-designed world. *Journal of Engineering Education*, *109*(2), 281–306. https://doi.org/10.1002/jee.20318

Witherspoon, E. B., Ferrer, N. B., Correnti, R. R., Stein, M. K., & Schunn, C. D. (2021). Coaching that supports teachers' learning to enact ambitious instruction. *Instructional Science*, *49*(6), 877–898. https://doi.org/10.1007/s11251-021-09536-7

Yackel, E., & Cobb, P. (1996). Sociomathematical norms, argumentation, and autonomy in mathematics. *Journal for Research in Mathematics Education*, *27*(4), 458–477. https://doi.org/10.2307/749877

Zambak, V. S., & Magiera, M. T. (2018). Pre-service K-8 teachers' professional noticing and strategy evaluation skills: An exploratory study. *Eurasia Journal of Mathematics, Science and Technology Education*, *14*(11), Article 1572. https://doi.org/10.29333/ejmste/92021

Appendix A. Review Methodology

This supplemental appendix provides additional details about our review methodology, including the procedures to select the NSF projects, as well as search, select, code, and synthesize their products.

Project Selection

From NSF's website, we searched for awards meeting the following criteria: (a) had an original award date between January 1, 2011, to December 31, 2015; (b) were tagged with DRK-12 program element code 7645; and (c) were active or completed. This search yielded 428 awards. However, some awards linked to the same project. For instance, a "collaborative research" project will have multiple NSF awards given to separate institutions, though the awards are part of the same project. After removing duplicate awards based on matching titles and abstracts, we identified 376 unique projects. We selected the award date range to focus on projects that were recently completed or are close to completion.

We searched for argumentation-related keywords in the award abstracts (*argu**, *critique*, *proof*, *proving*, *discourse*, *evidence*, *expla**, *justif**), narrowing the list of potential projects to 191 projects. The second author (Miller) reviewed award abstracts for these 191 projects, eliminating ones that were clearly irrelevant and designating borderline cases for group discussion. In this screening process, we operationally defined argumentation as "a process for reaching agreements about explanations and design solutions," along with providing domain-specific definitions for science ("argument is a process based on evidence and reasoning that leads to explanations acceptable by the scientific community") and mathematics ("mathematical argumentation is the construction and critique of mathematical conjectures and justifications"). Some abstracts mentioned the term "argumentation" explicitly, whereas others described the concept in other ways (e.g., "using evidence to support claims").

The award abstracts have at least two major limitations: (a) they are brief synopses that likely do not capture the full extent of each project's goals, and (b) the project goals may change between the time of award and the time of research. Given these limitations, we therefore erred on the side of inclusion when in doubt, so that we could use the associated products to inform our eligibility decisions. We removed projects if we could not find at least one produced document that was relevant to argumentation (see next section). This process yielded 23 argumentation-related projects.

Product Search

We used six sources to identify the publications and resources that the selected projects produced: Web of Science, ERIC, PsycINFO, Google Scholar, Research.gov, and the CADRE website (cadrek12.org). This search strategy targeted (a) documents that referenced the numeric award ID and (b) documents that project leaders listed on the Research.gov or CADRE websites.

Using the three literature databases (Web of Science, ERIC, and PsycINFO), we searched for the numeric award ID in the funding information search fields (e.g., the grant number field for Web of Science). Using Google Scholar, we searched for documents whose full text contained the numeric award ID and the terms "NSF" or "National Science Foundation." Google Scholar can complement searches of scientific databases by finding relevant gray literature sources (Haddaway et al., 2015). We conducted these searches using the full list of award IDs connected to the 23 argumentation-related projects. For instance, a collaborative research project will have multiple award IDs, and we searched for documents containing any of those award IDs. To complement these award ID-based search methods, we developed web scrapers in the *rvest* package in R (Wickham, 2019) to automatically extract citations and other resources (e.g., links to project websites and videos) from the project-specific pages on the Research.gov and CADRE websites. For Research.gov, this search included the public project outcome reports.

We merged the search results from these six different sources using the *revtools* package in R (Westgate, 2018), yielding 466 unique citations after removing duplicates. These citations indexed a diverse set of records and abstracts, including journal articles, conference presentations, book chapters, project websites, project outcome reports, videos hosted on the CADRE website, and other miscellaneous records. After identifying these citations, we also sent emails to each project's PI listing the citations we found, asking the PIs to provide any other products associated with the project.

Product Selection

For each NSF project, we identified between one to three products that were most closely related to argumentation. Product screening occurred in two main phases: (1) identify the products related to argumentation, and (2) select the one to three products per project that were the most complete and relevant to argumentation. We limited the maximum number to three products per project for reasons of practicality (i.e., create a manageable number of products to review) while ensuring representation across projects.

For Phase 1 screening, we identified documents that reported empirical research (quantitative or qualitative) focusing on argumentation as an educational topic in a science or mathematics domain, including (a) interventions or curriculum to teach students argumentation skills or processes, (b) assessments of students' or teachers' argumentation skills, and (c) teachers' knowledge, beliefs, or attitudes about argumentation (e.g., including but not limited to PCK about argumentation). We excluded studies that focused solely on teachers' argumentation performance (e.g., how teachers support scientific claims) without also studying their teaching of argumentation (either directly through classroom observations or indirectly through interviews and surveys). Phase 1 screening reduced the number of citations from 466 to 141.

For Phase 2 screening, we further restricted the corpus by limiting the maximum of products to three per project, for reasons of practicality, as noted earlier. We prioritized products that were

- peer-reviewed (e.g., journal article as opposed to conference poster),
- the most relevant to argumentation (e.g., had argumentation as its central as opposed to tangential focus), and
- provided the most complete reporting of argumentation-related results (e.g., when similar sets of results were reported across multiple products, such as a journal article and a conference paper).

This report's second author (Miller) trained junior staff on the screening Phase 1. Training steps included (a) providing example study screening decisions during the initial training phase, (b) listing common reasons for exclusion, and (c) conducting periodic dual screening checks on the junior screeners' decisions. Phase 2 screening was, admittedly, more subjective, so the second author conducted the Phase 2 screening rather than training junior staff. Phase 2 screening reduced the number of citations from 141 to 54.

Product Structured Coding

We quantitatively coded the products for the presence of key features, such as the component of argumentation studied or the research methods used. The "What Was Studied?" section provides results from this coding. As noted in that section, we coded at the product level and then summarized frequencies at the project level (e.g., 23 NSF projects contained 54 unique products). We summarized at the project level as a meaningful unit of analysis that gave equal weight across projects (rather than weighting toward projects that produced many documents).

We created a sheet using Google Forms with the structured codes and text descriptions for each coding category (see Appendix C for a complete list of codes). The lead and second author trained junior staff on an example set of three study articles, met with them on a weekly basis to address questions about the coding categories, and reviewed their codes to help ensure consistency across coders.

Synthesis of Empirical Findings

We also summarized the studies' empirical results in three steps. First, we coded lines of text from the results sections using NVivo, categorizing each relevant text section on results about argumentation into one of four categories of research: student intervention studies, teacher intervention studies, student naturalistic studies, and teacher naturalistic studies. Second, we created interpretative bullet point summaries in Microsoft Word for each document, separately by the study focus (i.e., intervention or naturalistic studies, and teacher- or student-focused). Third, we further condensed these summaries into Tables B1–B4 in Appendix B. We chose a qualitative synthesis approach, rather than a quantitative meta-analysis approach, because the studies varied widely in research methods and were often in qualitative in nature (Thomas & Harden, 2008). For instance, many studies were in-depth qualitative case studies based on interview data, for which extraction of quantitative effect sizes and formal meta-analysis would be inappropriate.

Limitations

Our synthesis focuses only on argumentation research funded by NSF's DRK-12 program, meaning it does not cover the entire field of recent STEM education research on argumentation. Also, because of limitations in the award abstracts, our synthesis may not cover all recent DRK-12 projects that studied argumentation . In addition, the methods rely on publicly available publications and products (or those provided to us by PIs), restricting the observable data about projects to what is reported in these documents. Last, the goals, interventions, methods, and outcomes of these projects varied considerably, presenting challenges in coherently synthesizing contributions across projects.

In defense of our synthesis, however, several points are worth noting: (a) the DRK-12 program is a major funder of U.S.-based preK–12 STEM education research; (b) the selected projects are likely representative of the recent DRK-12 portfolio on argumentation research, even if some projects might have been missed invertedly; (c) we took extensive effort to find relevant products, including contacting PIs by email; and (d) dividing the results into different lines of empirical research on argumentation helped us identify meaningful themes across projects, even if the methods varied.

Appendix B. Supplemental Tables

Exhibit B1. Findings From Intervention Studies of Student Argumentation (15 Projects, 30 Products)

NSF award ID/ title/amount	Domain	Sample	Method and measure	Key findings related to argumentation
#1118773: Collaborative Research: Computer-Supported Math Discourse Among Teachers and Students ($1,514,213.00)	Math: geometry	3 middle school students	Observational: student-student interactions	This project tests the impact of Virtual Math Teams with GeoGebra (VMTwG), geometry software that allows students to collaboratively solve geometry problems. Researchers used Toulmin's model to analyze students' interactions while using VMTwG (#43994). Results showed how students' approach to argumentation shifted from an initial focus on task completion to the following: a. using the tool as a way of gathering evidence b. using the programs rules as warrants c. introducing the notion of certainty d. debating warrants e. making providing reasonings normative f. using past constructions from the tools to support new hypotheses
		2 teams of 3–4 high school students	Observational: student-student interactions, student-teacher interactions	A case study of teachers' use of VMTwG with students (#70936[a]) showed how the teachers supported students' trajectories from empirical explorations to deductive justifications. Analyses show that a team of three students (Team 6) improved their collaboration, explorations, and mathematical reasoning.
#1119518: Mathematical Argumentation in Middle School: Bridging From Professional Development to Classroom Practice ($499,933.00)	Math	Sample size and grade level not reported	Observational: teacher-student interactions	This project examines how a professional development program that uses Geometer's Sketchpad influences teachers' argumentation instruction and students' conceptual understanding of mathematics and draws on ideas and processes from improvisational theater to support teachers in facilitating argumentation in their classrooms as a form of "disciplined improvisation." The results (#90058[a]) suggest that two important and related categories of teacher moves are moves that support math content learning and moves that support argumentation.

NSF award ID/ title/amount	Domain	Sample	Method and measure	Key findings related to argumentation
#1119670: CLASS: Continuous Learning and Automated Scoring in Science ($3,147,729.00)	Science: chemistry	164 high school students	Correlational: short-answer responses	This project produced a set of studies that use a technology-enhanced online science inquiry curriculum, the "Web-based Inquiry Science Environment" (WISE). One study (#19533) investigates the Idea Manager, an inquiry scaffolding tool that aims to help students record their ideas and construct coherent scientific explanations. Researchers scored students' explanations and used regression models to predict features of these explanations that predicted their coherence. Results suggest that explanations that were validly organized (e.g., organized in a way to support a valid explanation) were a significant predictor of argument coherence ($b = .41$, $p < .05$), above and beyond the number of ideas generated.
	Science: earth science	2 middle school students	Observational: student-student interactions Pre-post: short answer, multiple choice	A qualitative case study (#51492) of two students' discussions while using WISE identified some unique challenges of using computer-supported visual evidence in argumentation, including the following: a. Perceptions are strongly influenced by prior expectations. b. Subjective interpretations are difficult to refute. c. Criteria for evaluating visual evidence are not apparent.
	Science: biology	30 middle school students (Study 1); 117 middle school students (Study 2)	Experimental: short answer, multiple choice	A third set of studies (#90385) examined the impact of incorporating graph construction (Study 1) and graph critiquing (Study 2) into science inquiry lessons on students' conceptual understanding of scientific concepts. Results from Study 1 showed large pre-post gains ($d = 2.17$) in students' science understanding. Study 2 showed that students who critiqued graphs were significantly more likely to identify important aspects of the graph ($d = 2.49$) and describe the narrative process represented in the graph ($d = 1.58$), whereas students who constructed graphs were more likely to use science content ideas to explain their graphs ($d = 0.52$).

NSF award ID/ title/amount	Domain	Sample	Method and measure	Key findings related to argumentation
#1220756: High Adventure Science: Earth Systems and Sustainability ($2,328,593.00)	Science: biology	343 middle and high school students	Pre-post: short answer, student-student interactions	This project reports on the use of HASbot, an automated text scoring and real-time feedback system designed to support student revision of scientific arguments. Results (#9600) show students' use of HASbot was related to significant gains from pretest to posttest in scientific argumentation (*ES* = 1.52 *SD, p* < 0.001), above and beyond demographic characteristics. Authors hypothesize that HASbot helped students identify sources of uncertainty in their argumentation and that they were more actively engaged in improving their arguments when they were given the opportunity to generate, collect, and analyze their own data than observing someone else's data.
	Science: earth science, physics	302 high school students	Observational: short answer	Another study (#18510) found that (a) although the majority of students did not express uncertainty in their scientific explanations, students who expressed uncertainty did so scientifically without being prompted; (b) students' uncertainty ratings revealed a mix of their personal confidence and uncertainty related to science; and (c) if a task presented noisy data, students were less likely to express uncertainty in their explanations. The authors also develop a 2 x 3 coding framework for uncertainty: problematic vs. unproblematic thinking; and personal, empirical, signal, and conceptual uncertainty.
	Science: biology	512 middle and high school students	Experimental: short answer, multiple choice	Experimental results (#75991) indicate that (a) a majority of students (75.5%) incorporated models as evidence to support their claims, (b) most students used graphical model output to confirm their claim rather than to explain processes, (c) students' dependence on model results and uncertainty diminished as models became more complex, (d) some students' misconceptions interfered with observing and interpreting model results or simulated processes, and (e) students' uncertainty sources reflected more frequently on their assessment of personal knowledge or abilities related to the tasks than on their critical examination of scientific evidence.

NSF award ID/ title/amount	Domain	Sample	Method and measure	Key findings related to argumentation
#1223021: Collaborative Research: Researching the Efficacy of the Science and Literacy Academy Model ($1,547,637.00)	Science: biology	Late elementary students of 37 teachers	Quasi-experimental: student-student interactions, teacher-student interactions	This project (#78764[a]) examined the impact of a practice-based professional development (PD) program on the scientific discourse practices of teachers and their students. All students, regardless of the teacher cohort, made statistically significant improvements in their science discourse practices after attending the PD ($p < .001$, $\eta p2 = .63$). Interestingly, attending a practicum did not improve teacher outcomes ($p = .883$) but did improve their students' outcomes ($p = .076$, $\eta p2 = .09$).
#1316057: Developing Critical Evaluation as a Scientific Habit of Mind: Instructional Scaffolds for Secondary Earth and Space Sciences ($449,567.00)	Science: earth science	64 high school students	Quasi-experimental: short answer, multiple choice	This project (#52093) developed and tested instructional scaffolds to help students evaluate lines of evidence with respect to alternative scientific explanations. Result showed significant increases in plausibility and knowledge scores for students when two alternative explanations were prompted and no change with only one alternative. A structural equation model suggests that students' evaluation of evidence may influence plausibility and knowledge.
		299 high school students	Quasi-experimental, pre-post, correlational: short answer, survey	Another study (#58682) investigated changes to plausibility judgments and knowledge as a result of instructional scaffolds called model-evidence link (MEL) activities. Results showed that participants' plausibility judgments shifted toward scientifically accepted explanations and increased their content knowledge. Structural equation modeling revealed that 10% of the variance in knowledge scores was related to participants' evaluations, above and beyond background knowledge, which accounted for 26% of the variance.
		Middle school students (sample size not reported)	Descriptive: short answer	An additional study (#88577) developed a rubric for assessing the quality of student evaluations when engaging in the MEL activity, specifically in the written explanations about the connections between evidence and explanations. This rubric features four distinct categories of evaluation: (a) erroneous, (b) descriptive, (c) relational, and (d) critical.

NSF award ID/ title/amount	Domain	Sample	Method and measure	Key findings related to argumentation
#1316347: Integrating Quality Talk Professional Development to Enhance Professional Vision and Leadership for STEM Teachers in High-Need Schools ($2,106,207.00)	Science: chemistry, physics			This project (#5631[a]) tested the efficacy of the Quality Talk (QT) professional development support, materials, and coaching. QT led to a significant increase in students' use of scientific reasoning to build arguments (0.65 to 2.02 out of 3 points from Time 1 to Time 4). Importantly, students were better able to enact these discourse practices on their own, without teacher support, after QT Science discourse lessons. In turn, students who experienced QT Science produced stronger written arguments, a key indicator of scientific literacy.
		301 high school students	Observational, quasi-experimental, pre-post: teacher-student interaction, multiple choice, short answer	A quasi-experimental study (#24924[a]) revealed that QT discourse practices reflected better critical-analytic thinking and argumentation relative to comparison classrooms. Similarly, students at QT produced stronger written scientific arguments than comparison students ($p < .001$); $\eta p2 = .059$.
#1316799: Enhancing Argumentation With Social Media: Supporting Teacher Professional Development ($429,784.00)	Science: general	More than 500 middle school students	Observational: online student discussions	This project (#24600) examined the argumentation discourse of students engaged in an online multiplayer game (Reason Racer). Results indicated that socio-scientific topics produced collaborative discourse episodes that were positive, supportive, and civil within an argumentation framework. Scientific argumentation vocabulary was used 43.9% of the time for socio-scientific topics ($n = 108$) and 39.1% of the time for scientific topic scenarios ($n = 270$). Students in the scientific topic group had more substantive discussions ($t(436) = 2.03$, $p = .04$) and fewer social interactions ($t(400) = -2.72$, $p = .01$).
	Science: biology	402 high school students	Experimental: multiple choice, survey	An experimental study (#56233) showed that the group using Reason Racer ($M = 27.52$) scored significantly higher ($p = .00$) overall than the comparison group ($M = 24.41$) on their knowledge of scientific argumentation and showed a significant overall increase ($p = .00$) in confidence regarding scientific argumentation.

NSF award ID/ title/amount	Domain	Sample	Method and measure	Key findings related to argumentation
	Science: general	249 middle school students	Quasi-experimental, pre-post: short answer, survey, student-student interactions	A quasi-experimental study (#53147) showed that students who played the game at least 10 times also reported an increase in confidence (p = .001, ηp2 = .015) and motivation (p = .001, ηp2 = .012) to engage in science compared with students who did not play the game.
#1317034: Learning Algebra and Methods for Proving (LAMP) ($370,746.00)	Math: algebra	210 middle school students	Experimental: short answer, multiple choice	This study (#91096) reports on the Learning Algebra and Methods for Proving (LAMP). LAMP teachers are encouraged to include viable argumentation in every lesson they teach and on assessments. This includes asking students to develop explicit claims using the language of mathematics (e.g., “for all” and “there exists”) and to support claims with viable referents and warrants. Findings demonstrate that treatment students outperformed comparison students on an argument and reasoning assessment and state standardized assessments.
		1 middle school student	Observational, pre-post: short answer, interviews	This study (#52559) used a teaching experiment and retrospective analysis to develop a learning trajectory for improving a Grade 8 student’s ability to construct, critique, and validate contrapositive arguments. The student’s learning trajectory demonstrates how a conception of contrapositive proving as eliminating counterexamples can be useful in developing, critiquing, and validating contrapositive arguments. Analysis revealed that the student’s contrapositive reasoning could be improved through a teaching experiment that encouraged the eliminating counterexamples approach.
#1324977: DIP: Community Knowledge Construction in the Instrumented Classroom ($1,236,076.00)	Science: biology	47 5th- and 6th-grade students	Observational: teacher-student interactions	This project (#57402) uses the claims, evidence, and reasoning (CER) framework to evaluate the use of WallCology, a digital ecosystem. Results show that student groups were proficient in their scientific explanations and effectively used a variety of representation types across all three components of CER in their explanations. Authors hypothesize that the emphasis on persuasion as a goal and on turning representations that were “of something” (e.g., simulation screens that were visible in the classroom) into representations that were “for something” helped improve students’ explanations.

NSF award ID/ title/amount	Domain	Sample	Method and measure	Key findings related to argumentation
		45 middle school and 51 high school students	Observational: short answer	This study (#73216) reports on two learning environments in which students used visual evidence (digital photographs) or textual evidence for the scientific practices of planning and argumentation. Textual evidence was found to provide greater support (M = 3.0, SD = 1.4 vs. M = 2.63, SD = 1.34) than high-level visual evidence in making connections to investigation plans and arguments.
#1417757: Learning Labs: Using Videos, Exemplary STEM Instruction, and Online Teacher Collaboration to Enhance K–2 Mathematics and Science Practice and Classroom Discourse ($2,998,840.00)	Science: general	Students from a single 2nd-grade teacher's classroom	Observational: short answer	This project consisted of a single practitioner-focused report (#75682) that examines the use of agree/disagree T-charts for supporting students in the following: • generating and refining claims about a phenomenon • rallying evidence from different unit activities • understanding how arguments and models co-develop over time, as new evidence is integrated • making their reasoning public and developing a community that values learning together
#1417895: Preparing Urban Middle Grades Mathematics Teachers to Teach Argumentation Throughout the School Year ($2,999,737.00)	Math: geometry	97 middle school students	Experimental, observational, pre-post: interviews, teacher-student interactions	This technical report (#24152[a]) presents design principles and an evaluation of a Bridging PD program. In all the target classrooms, students demonstrated substantial learning gains, with an effect size of 1.45 (t(96) = 9.94, p <.0001), on average gaining 10.31 points out of 36 from pretest to posttest.

NSF award ID/ title/amount	Domain	Sample	Method and measure	Key findings related to argumentation
#1418019: Investigating How to Enhance Scientific Argumentation Through Automated Feedback in the Context of Two High School Earth Science Curriculum Units ($2,495,604.00)	Science: earth science	1,180 (Phase 1) and 42 (Phase 2) high school students	Longitudinal, correlational: short answer	This project (#32872) developed a computer-based formative assessment to support students' construction and revision of scientific arguments. In addition to a psychometric validation of the tool (Phase 1), a pilot study (Phase 2) showed that students' scientific argumentation skills improved during their revision process on 11 of 16 items ($d \geq .20$, $p < .05$).
		183 high school students	Correlational, pre-post: short answer, survey	Results from another study (#93829) showed that the majority of students (77%) made revisions after receiving the feedback, and students with higher initial scores were more likely to revise their responses. Students who revised had significantly higher final scores than those who did not ($d = 0.96$), and each revision was associated with an average increase of 0.55 on the final scores. Analysis on item difficulty shifts showed that written scientific argumentation became easier after students used the automated feedback.
#1418136: Building High School Students' Understanding of Evolution ($3,000,000.00)	Science: biology	2,269 high school students	Experimental, correlational, pre-post: essays, multiple choice	This project (#5639) reports on the development and testing of a biology unit that aims to build students' abilities in analyzing and interpreting data about phenomena from published scientific research and engage students in the construction of evidence-based arguments. Treatment students had significantly higher gains than control students ($d = .58$).
		944 high school students	Pre-post: multiple choice	Results from another study (#5637[a]) showed statistically significant gains with large effect sizes from pretest to posttest in students' conceptual understanding of evolution and genetics ($d = .96$). Students also gained skill in identifying claims, evidence, and reasoning in scientific arguments.
#1453493: CAREER: Proof in Secondary Classrooms: Decomposing a Central Mathematical Practice ($874,017.00)	Math: geometry	389 high school students	Pre-post: multiple choice	This project (#92500[b]) was part of a larger body of work aimed at decomposing proofs in geometry. A single study emphasized the importance of teaching certain proof competencies before teaching proof.

[a] Study included findings for both student and teacher participants. [b] Study contained both intervention and naturalistic components.

Exhibit B2. Findings From Intervention Studies of Teacher Argumentation (Eight Projects, 13 Products)

NSF award ID/title	Domain	Sample	Method and measure	Key findings related to argumentation
#1118773: Collaborative Research: Computer-Supported Math Discourse Among Teachers and Students ($1,514,213.00)	Math: geometry	Single teacher	Observational: lesson plans	This case study (#70936[a]) reports on one teacher's use of Virtual Math Teams with GeoGebra (VMTwG), which provides technology that supports joint problem solving. Results showed how the teacher supported students' trajectories from empirical explorations to deductive justifications. Specifically, supports fell into three main categories: a. collaboration b. mathematical reasoning c. the use of technology
#1119518: Mathematical Argumentation in Middle School: Bridging from Professional Development to Classroom Practice ($499,933.00)	Math: geometry	4 middle school teachers	Observational: teacher-student interactions	This practitioner-focused project (#51831) reports on successful episodes of argumentation conducted by teachers and students participating in the Bridging Professional Development (BPD) project, as well as several teacher moves that could help engage students' argumentation. Results suggest that in BPD classrooms, conjectures are often based on an examination of patterns and the suggestion to be creative. Justifying takes on four forms: a. Teams or individuals generate conjectures, and the whole class justifies them. b. The whole class together generates and justifies conjectures. c. Teams or individuals come up with conjectures and justifications, which are shared with the class and brought to conclusion. d. Conclusions are, in essence, a time for class agreement that a conjecture is true or false, assuming that an argument has been persuasive and correct.

NSF award ID/title	Domain	Sample	Method and measure	Key findings related to argumentation
#1119584: Constructing and Critiquing Arguments in Middle School Science Classrooms: Supporting Teachers with Multimedia Educative Curriculum Materials ($3,147,015.00)	Science: earth science	46 middle school teachers	Correlational, pre-post: log files, PD observation, survey, short answer	This project (#51974) describes the development and study of teachers' use of multimedia educative curriculum materials (MECMs) to support argumentation instruction. Teacher background characteristics didn't predict MECM use, but teachers were more likely to access MECMs that were embedded in lesson plans or reflective self-assessment prompts than when they were in a separate library. In addition, teachers were more likely to watch videos earlier in the curriculum and the first time a new activity structure was introduced.
		10 middle school teachers	Observational: interviews, teacher-student interactions	One study (#68330) analyzed videos and interviews of 10 teachers' curricular decision making as they taught a lesson on scientific argumentation. Results suggest three main influences on teachers' curricular decision making in classes with higher quality argumentation: (a) teachers' understanding of argumentation as an epistemic practice (rather than surface-level features), (b) teachers as critically reflective curriculum users, and (c) teachers problematizing their prior teaching experiences.
#1223021: The Role and Use of Examples in Learning to Prove ($1,547,637.00)	Science: biology	44 late elementary teachers	Quasi-experimental: teacher-student interactions	This project (#78764[a]) examines the impact of a practice-based PD on the scientific discourse practices of teachers and their students. Features of the PD include content focus, active learning, coherence, sufficient duration, and collective participation. One cohort attended the full program (institute, practicum, and follow-up sessions), while another comparison group but did not take part in the practicum. All teachers and their students made statistically significant improvements in their science discourse practices ($p < .001$, ηp2 = .64). However, teachers who attended the full PD (with practicum) did not outperform the teachers who did not attend the practicum ($p = .883$).
	Science: general	44 late elementary teachers	Quasi-experimental: PD observations, teacher reflections, teacher-student interactions	One study (#10006) showed that video-based reflection can provide opportunities for practice and reflection that are potentially valuable features of professional development programs.

NSF award ID/title	Domain	Sample	Method and measure	Key findings related to argumentation
		Single late elementary teacher	Observational, pre-post: PD observations, interviews, short answer, teacher-student interactions	Another study (#38071) reports on a teacher's enactment of instructional strategies learned through reflection on video observations of teaching. The analysis focused on three components of practice to document teacher instructional change: (a) talk format (e.g., instruction, whole/small group, pair, individual); activities (charting ideas, "four corners," cartoons); and teacher moves (ask, support, press, link). After PD, the teacher used talk formats that encouraged interaction, activities that provided opportunities for discussion, and teacher moves that supported students' engagement in scientific discourse.
#1316347: Integrating Quality Talk Professional Development to Enhance Professional Vision and Leadership for STEM Teachers in High-Need Schools ($2,106,207.00)	Science: chemistry, physics	7 high school teachers	Observational, quasi-experimental, pre-post: teacher-student interaction, multiple choice, short answer	This project (#5631[a]) tested the efficacy of the Quality Talk Science professional development support, materials, and coaching. QT Science led to changes in both teachers' and students' discourse practices. Teachers could more effectively (a) teach and promote discourse focused on substantive questions about core scientific phenomena and (b) elaborate and appropriately critique responses to those questions.
			Quasi-experimental, observational, pre-post: teacher-student interactions	A quasi-experimental study (#24924[a]) revealed that treatment teachers' discourse practices better reflected critical-analytic thinking and argumentation at posttest relative to comparison classrooms.
#1417895: Preparing Urban Middle Grades Mathematics Teachers to Teach Argumentation Throughout the School Year ($2,999,737.00)	Math: geometry	24 middle school teachers	Experimental, observational, pre-post: teacher-student interactions	This project (#24152[a]) presents design principles and an evaluation of a Bridging PD program. Design principles include the following: (a) Mathematical knowledge for teaching is foundational in learning to teach for argumentation; (b) curriculum aids in supporting classroom argumentation; (c) teaching is improvisational, and teaching moves can be taught through improvisation games; (d) planning is the complement to improvisation; and (e) argumentation for students is improvisational and can be learned through improvisational games. Treatment teachers showed more use of argumentative discourse in both years 1 ($ES = 1.4$) and 2 ($ES = .7$); teaching moves that support argumentation: closed-ended questions [$t = 2.2$, $p < .05$], open-ended questions [$t = 2.8$, $p < .05$], and moves to facilitate and encourage participation by multiple students [$t = 2.4$, $p < .05$].

NSF award ID/title	Domain	Sample	Method and measure	Key findings related to argumentation
#1418136: Building High School Students' Understanding of Evolution ($3,000,000.00)	Science: biology	16 high school teachers	Correlational, observational, pre-post: teacher reflection, short answer, survey, teacher-student interactions	This project (#5637[a]) reports the development and use of a biology curriculum unit that provides opportunities for students to ask scientific questions, use models, analyze data from published scientific studies, and construct evidence-based arguments. Twelve teachers (66.7% of respondents) reported that the unit was better than prior curriculum materials (e.g., the use of real-world data, the CER scaffold and opportunities to build the practice of argumentation, unit design that allows students to take ownership over their learning, and the scientific research that went into designing the activities). Three teachers (16.67%) noted that it was as good as their current materials (preferred other materials for lower reading levels for their special education and low-achieving students). The remaining three teachers (16.7%) indicated that some parts of the unit were better than materials they had used in the past and that some parts were not as good.
#1503511: Developing Teachers' Capacity to Promote Argumentation in Secondary Science ($2,770,500.00)	Science: general	12 middle school teachers 14 high school teachers	Observational: lesson plans, teacher-student interactions	This project reported an observational study (#35668) of teachers' dialogic moves both with and without PD support. Without support, teachers tended to dominate classroom conversations and limit student-student interactions. However, PD that included a storyline model that encouraged teachers to anchor the discussion to a particular phenomenon showed some increase in teachers' use of open-ended questions and more student-student dialogue.

[a] Study included findings for both student and teacher participants.

Exhibit B3. Findings From Naturalistic Studies of Student Argumentation (Seven Projects, 11 Products)

NSF award ID/title	Domain	Sample size	Outcome and measure	Key findings related to argumentation
#1119584: Constructing and Critiquing Arguments in Middle School Science Classrooms: Supporting Teachers with Multimedia Educative Curriculum Materials ($3,147,015.00)	Science: biology	5 middle school students	Student challenges and competencies: interviews, teacher-student interactions	Part of a larger project investigating the effects of multimedia educative curricular materials (MECMs) on teachers' learning and beliefs about scientific argumentation, this project (#45935) used a community of practice (CoP) framework to investigate the relationship between students' argumentation in a single sheltered English immersion (SEI) classroom. Findings suggest that student movement in and out of the immersion class impacted their opportunities for legitimate peripheral participation, specifically in that new students could not interact with more advanced students who had transitioned out. Heterogenous grouping by experience level to allow for peer modeling was suggested as a way to support students' participation. Findings also suggest the importance of allowing students to use their first language as they acquire fluency in science practices as a way to encourage participation and that front-loading language-heavy content and vocabulary, as is often suggested by SEI approaches, may be counterproductive in providing students with opportunities for scientific sensemaking.
#1149436: CAREER: Learning to Support Productive Collective Argumentation in Secondary Mathematics Classes ($535,007.00)	Math: geometry	Students in 2 high school classrooms	Student challenges and competencies: teacher-student interactions	This project (#37794[a]) uses Toulmin's model to map components of an argument during a mathematical discussion, including student contributions and teacher supports. The researchers focus on collective argumentation and documenting the different contributions (data, warrants, claims) of students and teachers during a classroom mathematical discussion.

NSF award ID/title	Domain	Sample size	Outcome and measure	Key findings related to argumentation
#1220623: The Role and Use of Examples in Learning to Prove ($995,955.00)	Math: general		Student challenges and competencies: researcher-student interviews	This project uses the Criteria, Affordances, Purposes, and Strategies (CAPS) framework (#64561) as a tool for examining students' criteria for choosing examples during conjecturing and proving activities to examine the use of examples in learning to develop mathematical proofs, specifically in developing a viable justification for why a conjecture must be true.
		12 middle school and 16 high school students		Three studies used the same data set to examine different aspects of example use. Findings of one study (#5609) suggest that "successful provers" (produced a complete or incomplete but viable justification for why a conjecture was true in general) were more likely to a. note the *affordances* of an example for generalization and gaining insight into the conjecture; b. have *strategies* for using examples (e.g., pattern search, attempting to disprove), whereas unsuccessful provers were largely restricted to strategies about choosing examples; and c. identify the *purpose* of the example as conveying a general argument.
				A second study (#54759) distinguished between empirical (using specific examples to test conjectures) versus generic (viewing these examples as specific instances of a more general case) use of examples. The authors conclude that generic example use is more likely to be productive than empirical example use, which over rely on specific examples and can lead to errors in inference. Also, teacher-selected examples may be more productive than student-generated ones; however, it is important for teachers to draw students' attention to the generalizable elements of the examples they provide.
				A final study (#71161) focused on differences in the coordination of purposes and strategies of proving activities between the reasoning of students and experts as they developed conjectures. Although both used a *direct approach* (constructed examples to satisfy the hypothesis) and a *contradiction approach* (constructed examples that satisfy the hypothesis and the negation of the conclusion, i.e., counterexamples), only experts use a *contrapositive approach* (i.e., construct an example that satisfies the counterexample).

NSF award ID/title	Domain	Sample size	Outcome and measure	Key findings related to argumentation
#1253081: CAREER: Noticing and Using Students' Prior Knowledge in Problem-Based Instruction ($644,320.00)	Math: geometry	23 high school students	Student challenges and competencies: short answer	This project (#32744) used Toulmin's model to examine how the visual support of a 1-point perspective drawing influenced students' argumentation about figures. The authors analyzed transcripts of students as they discussed the relative dimensions of objects in the drawing. Findings suggest that students used both artistic and everyday knowledge about perspective to help them develop warrants for their claims. Students used given features of the drawings, added features, or measured components in order to justify their claims about the figures.
#1317034: Learning Algebra and Methods for Proving (LAMP) ($449,567.00)	Math: rational numbers	Single middle school student	Student challenges and competencies: short answer, interviews, student-student and teacher-student interactions	Drawing on data from a broader Learning Algebra and Methods of Proof (LAMP) intervention, this project (#84212) uses a single case study to examine the relationship between students' understanding of repeating decimals and rational numbers. The author uses this case to suggest that argumentation about the representation of rational numbers as repeated decimals (e.g., if .333… equals 1/3 or if .999… equals 1) can act as core mathematical content for teaching and learning.
#1418423: GRIDS: Graphing Research on Inquiry with Data in Science ($2,999,748.00)	Science: earth science, physics	460 middle school students	Student challenges: short answer, multiple choice	This project (#35709) reports on the development of a measure of graph comprehension, critique, and construction based on the knowledge integration framework and evidence of student difficulties in critiquing or constructing graphs for scientific modeling, reasoning, and communication. This is important because graphs can often be used as a source of evidence during scientific argumentation. Authors quantitatively and qualitatively analyze student responses to graphing comprehension, critique, and construction items. Results suggest multiple challenges that students face when interpreting graphs embedded in a science context, including the following: a. interpretation of complex graphs b. critiquing graphs c. constructing graphs from narrative accounts of scientific phenomena d. using scientific knowledge to interpret graphs

NSF award ID/title	Domain	Sample size	Outcome and measure	Key findings related to argumentation
				The authors suggest that experiences that require students to construct graphs, explain their reasoning, and challenge the reasoning of others may help improve students' understanding and use of graphs for scientific understanding. They also note that interpretation and critique items were particularly challenging for English learners and suggest that it may be particularly important to adapt materials that involved interpreting and critiquing graphs for those students.
#1453493: CAREER: Proof in Secondary Classrooms: Decomposing a Central Mathematical Practice ($874,017.00)	Math: geometry	15 high school students	Student challenges: interviews	This project addresses the challenge of teaching proof in secondary geometry classrooms and examines common misconceptions and challenges students face in learning proof. In one study (#56552), researchers interviewed high-achieving students in geometry to better understand their approach to proving when confronted with four alternative proving tasks: a. a two-column proof b. a diagramming task c. a task in which students were asked to draw conclusions d. a task in which students were asked to determine a theorem from a completed proof
		High school students (sample not reported)	Student challenges: interviews, multiple choice, teacher-student interactions	Students struggled to complete the tasks that differed from what they normally encountered in class. The authors conclude that students turned proving into a rote task, in which they expected to mark a diagram and complete a written two-column proof in which some information was given to them, but struggled to recognize similar activities that did not look like what they thought of as "doing proof."
				A second study (#92500[b]) identifies particular conceptual obstacles students face when proving and provides some instructional suggestions for ways to mitigate these obstacles. Specifically, they identify five common conceptual obstacles and offer five corresponding instructional strategies: 1. *"You can draw conclusions from diagrams."* Teach students to draw valid conclusions before teaching proof.

NSF award ID/title	Domain	Sample size	Outcome and measure	Key findings related to argumentation
				2. *"You cannot make assumptions about diagrams."* Teach students what they can and can't assume about diagrams. 3. *"A definition can include all the properties that one knows about the geometric object."* Have students practice defining, and continually emphasize the importance of knowing definitions. 4. *"Bisectors divide triangles in half or act as lines of symmetry."* Focus repeatedly on the three types of bisectors, and formatively assess students' progress. 5. *"When attempting to prove a conjecture as a theorem, one assumes the conclusion of the statement."* Teach students to rewrite conjectures as conditional statements and identify the hypothesis as the "given" and the conclusion and the "prove" statement.

[a] Study included findings for both student and teacher participants. [b] Study contained both intervention and naturalistic components.

Exhibit B4. Findings From Naturalistic Studies of Teacher Argumentation (Four Projects, Nine Products)

NSF award ID/title	Domain	Sample size	Outcome and measure	Key findings related to argumentation
#1119584: Constructing and Critiquing Arguments in Middle School Science Classrooms: Supporting Teachers with Multimedia Educative Curriculum Materials ($3,147,015.00)	Science: earth science	42 middle school teachers	Teacher beliefs: short answer, surveys, interviews	This project is part of a larger effort investigating the effects of multimedia educative curricular materials (MECMs) on teachers' learning and beliefs about scientific argumentation. One study (#37635) used surveys and interviews to identify the factors that teachers believe impact their argumentation instruction. Teachers cited their own learning goals as the most important, whereas context, policy, and assessment were all less important, in part because teachers saw policy and assessment as misaligned with the learning goals of argumentation. The authors suggest that because teachers see argumentation as important, helping them identify what counts as argumentation and providing support to try argumentation in their classes may increase teachers' confidence and encourage integration of argumentation in their instruction.
	Science: biology	Single middle school teacher	Teacher practice: teacher-student interactions	A second study (#76223) focused on specific challenges and strategies for teaching argumentation with English learners (ELs). The authors used a case-study methodology to examine one teacher's approach to teaching argumentation and identified three components of their practice that acted as supports for ELs: a. more language supports that focused on argument structure b. dialogic interactions that were most often facilitated by productive language supports c. some language supports that offered a rationale for argumentation
		34 middle school teachers	Teacher beliefs: short answer, interviews, survey	The authors conclude that making rationales explicit by explaining the *why* behind science practices could be beneficial for all students, not just ELs. A third study (#16454) examined beliefs about argumentation of teachers at schools serving students from primarily low, medium, and high SES backgrounds. Results show three emergent themes of teachers' beliefs about argumentation:

NSF award ID/title	Domain	Sample size	Outcome and measure	Key findings related to argumentation
				a. *Teachers saw different benefits based on the SES of their students.* Teachers of high and mid SES students said argumentation was a good way for them to learn how to listen and build on each other's ideas. Teachers of low SES students said argumentation provided a good outlet for students to talk, which they want to do but have few opportunities to do during the school day. b. *Teachers had varied beliefs about students' capabilities for argumentation and their roles in supporting them.* Some responses showed a possible deficit view of their students' abilities to engage in argumentation, especially teachers of students with low SES, ELs, and students in special education, despite also providing contradicting survey responses suggesting that all students can do argumentation. Relatedly, these teachers often described their role as providing scaffolds to reduce the difficulty of argumentation. However, teachers who believed that all students could engage in argumentation used scaffolding as an example of why all students *are* capable. c. *Teachers in low SES schools felt that pressure from standards and state testing impacted their ability to teach argumentation, more so than teachers in higher SES schools.* However, teachers of low SES students differed in that some used accountability pressures to explain why they did not teach argumentation (as it was not explicitly on the test), whereas others suggest argumentation is a skill that can support test taking.
#1149436: CAREER: Learning to Support Productive Collective Argumentation in Secondary Mathematics Classes ($535,007.00)	Math: geometry	2 high school teachers	Teacher practice: teacher-student interactions	This project focuses on how teachers learn to support students' argumentation, following preservice teachers through their preparation courses and their first 2 years of teaching. One study (#37794[a]) uses Toulmin's model to map components of an argument during a mathematical discussion, including student contributions and teacher supports. The researchers identified five types of question prompts that teachers used to support students' argumentation, including the following:

NSF award ID/title	Domain	Sample size	Outcome and measure	Key findings related to argumentation
				a. Request a factual answer. b. Request a method. c. Request an idea. d. Request an elaboration. e. Request an evaluation.
	Math: algebra	15 middle school teachers	Teacher beliefs: interviews	Importantly, questions posed by the teacher helped make warrants explicit. In particular, asking "why" questions or asking students to explain their reasoning encouraged them to contribute warrants. A second study (#43829) asked preservice secondary teachers to evaluate the hypothetical arguments of middle school students. Researchers differentiate between three different perspectives on student thinking, based in different rationales: a. a *teacher perspective* that used professional training in mathematical knowledge for teaching proof (MKT-P) to make hypotheses about students' prior knowledge (6 teachers) b. a *student perspective*, which relied on their own experiences as students to guide their instructional decisions (9 teachers) c. a *combined perspective*, based on both thinking about one's own experiences and what they knew about what students might know or how they might think about mathematics (2 teachers) Researchers suggest that the *student perspective* may be skewed based on teachers' other experiences and could lead teachers to not present proofs at a developmentally appropriate level.

NSF award ID/title	Domain	Sample size	Outcome and measure	Key findings related to argumentation
#1316241: Developing Rich Media-based Materials for Practice-based Teacher Education ($2,634,873.00)	Math: algebra	20 middle school teachers	Teacher practice: survey, short answer	This study (#47795) is part of a larger project using an online application that allows teachers to create storyboard-like depictions of classroom scenarios. Results from coded depictions of the teacher actions (e.g., probing, generating discussions) show two themes which were correlated with survey measures of teacher experience: a. *Passive facilitation of argumentation*. Teachers used statements to take over the thinking or used *silence* (negatively correlated with teaching experience, $\rho = -.55$, $p < .10$) and generating discussions (negatively correlated with time spent on whole-class discussions, $\rho = -.49$, $p < .10$), which failed to elicit the key mathematical ideas from students. b. *Active facilitation of argumentation.* Teachers used probing (negatively correlated with several math courses, $\rho = -.45$, $p < .05$) and orienting and focusing (positively correlated with class time spent on whole-class discussions, $\rho = .53$, $p < .05$) to solicit student explanations. These results suggest that teachers with more classroom experience and who spend more time on whole-class discussion may be more likely to engage in active facilitation, whereas teachers who are less experienced, spend less time on whole-class discussion, and take more math classes may engage in more passive facilitation (perhaps suggesting an expert blind spot).
#1350802: CAREER: L-MAP: Pre-service Middle School Teachers' Knowledge of Mathematical Argumentation and Proving	Math: general	52 late elementary preservice teachers	Teacher beliefs, teacher practice: teacher reflections	This project focused on how preservice teachers' knowledge of mathematical argumentation and proving develops in teacher preparation programs. One study (#5650) analyzed preservice teachers' written journals to gain insight into their conceptualizations of the meaning of mathematical argumentation. Findings show that the majority of preservice teachers discussed mathematical argumentation from the perspective of the individual, with less focus on argumentation as a social activity.
($791,854.00)	Math: basic arithmetic, rational numbers	34 late elementary preservice teachers	Teacher practice: short answer	A second study (#8663) analyzed preservice teachers' response to four tasks, focusing on the presence of three teaching competencies:

NSF award ID/title	Domain	Sample size	Outcome and measure	Key findings related to argumentation
				a. professional noticing of student mathematical reasoning and strategies b. the ability to assess the validity of student reasoning and strategies c. the ability to select student strategies for class discussion
	Math: general, rational numbers	37 late elementary preservice teachers	Teacher practice: short answer	Results suggest that supporting preservice teachers' ability to notice the mathematically significant aspects of student reasoning and strategies can help them better assess the validity of student reasoning and strategies and that selecting strategies with the purpose of engaging students in justifications can advance their conceptual understanding. A third study (#5651) examined preservice teachers' written explanations that they wrote to support their own solutions and to critique solutions generated by their students. The findings suggest that although overall preservice teachers' explanations written to critique students were weaker, preservice teachers who were stronger at explaining their own solutions also were stronger at critiquing students' explanations ($r = 0.501$, $p < .05$). The authors also identify six criteria preservice teachers used to evaluate students' explanations: 1. Attention to correctness of results or strategy 2. Attention to organization 3. Attention to foundations (articulation of the problem and process) 4. Attention to explanation's communicative power 5. Attention to justifications 6. Attention to generality

[a] Study included findings for both student and teacher participants.

Appendix C. Coding Structure

These codes should be viewed as "select all that apply" (not mutually exclusive categories).

Coding field	Response options
1. What problems or topics related to scientific and mathematical argumentation were studied?	
Research focus	• INTERVENTION: developing or testing specific interventions for teachers (e.g., professional development workshops) or students (e.g., curriculum resource) • NATURALISTIC: studying argumentation as practiced by teachers and students in business-as-usual contexts
Research participant type	• In-service teachers • Preservice teachers • PreK–12 students
Grade level	• PreK • Early elementary (Grades K–2) • Late elementary (Grades 3–5) • Middle school (Grades 6–8) • High school (Grades 9–12) • Not reported/general K–12
Studied student cultural, demographic, or linguistic diversity	• DIVERSITY AS CONTEXT: Study explicitly noted underserved communities or student populations underrepresented in STEM as a context for the research (e.g., "The study district was a majority Hispanic student population."). • DIVERSITY AS FOCUS: Study explicitly studied how student engagement in argumentation varies across diverse cultural, demographic, or linguistic backgrounds (e.g., how to support English learners specifically). If either, type of diversity mentioned or studied: • English learners • Socioeconomic status • Race/ethnicity • Gender • Urbanicity • Students with disabilities
Major content area	• Mathematics • Science
Specific mathematics content area	• NUMBER SENSE: seeing relations between different representations of numbers (e.g., placing numerals such as "5" on a number line), performing mental math estimations, and applying numbers in real-world contexts (generally preK or K) • BASIC ARITHMETIC: addition, subtraction, multiplication, and/or division of whole numbers

Coding field	Response options
	• RATIONAL NUMBERS: fractions, decimal numbers, percentages, ratios/proportions, including operations applied to rational numbers (e.g., multiplication of fractions) • GEOMETRY: lines, angles, classifying shapes, plotting points on a coordinate plane, two-dimensional or three-dimensional geometry • ALGEBRA: linear and quadratic equations, relationships, polynomials, systems of equations, functions, algebraic expressions. Symbols represent quantities and express generalized relations. • PROBABILITY AND STATISTICS: characterizing statistical variability, summarizing different types of distributions, calculating simple probabilities of events, interpreting conditional probabilities, making statistical inferences • PRECALCULUS/CALCULUS: advanced algebra and trigonometry topics needed for calculus, derivatives, integrals • GENERAL PROBLEM SOLVING: interpreting and solving word problems • GENERAL: general mathematical argumentation; not restricted to any particular content area or unspecified
Specific science content area	• CHEMISTRY: elements and compounds composed of atoms, molecules, and ions • PHYSICS/PHYSICAL SCIENCE: nature and properties of matter, including mechanics, motion, heat, light, electricity, magnetism, and sound • BIOLOGY/LIFE SCIENCE: life and living organisms, including their structure, interactions, physiological mechanisms, development, and evolution • EARTH SCIENCE: study of our planet's physical characteristics, from earthquakes to raindrops and floods to fossils • SCIENTIFIC INQUIRY: designing and conducting empirical investigations to test and refine hypotheses/explanatory models • GENERAL: general scientific argumentation; not restricted to any particular content area or unspecified
2. What theoretical perspectives on argumentation did researchers use to conceptually frame their work?	
Definition of argumentation	• TOULMIN'S MODEL: Explicitly referenced Toulmin's model of argumentation, consisting of data, warrants, and claims. • STRUCTURAL, NON-TOULMIN: Explicitly references "argument" or "argumentation" and defines components of an argument (e.g., explanation, reasoning, evidence) without referencing Toulmin. • DIALOGIC/SOCIOCULTURAL: Argumentation is a social practice involving discourse to persuade others, defend claims, critique others' claims, and achieve consensus within communities of practice. • PROOF: Explicitly references mathematical "proof" (does not need to explicitly reference "argument" or "argumentation"). • JUSTIFICATION: Explicitly references mathematical "justification" (does not need to explicitly reference "argument" or "argumentation"). • USE OF SCIENTIFIC EVIDENCE: Explicitly references the use of scientific evidence (e.g., supporting explanations with evidence; does not need to explicitly reference "argument" or "argumentation"). • OTHER: Study defines argumentation in another way not previously listed (use sparingly; this code is meant to help identify potential other more specific codes).

Coding field	Response options
Stated importance of argumentation (Jimenez-Aleixandre & Erduran, 2007; McNeill et al., 2017)	The article noted that engaging students in argumentation is important because of the following: • DISCLIPINARY PRACTICE: Argumentation is a key disciplinary practice of professional scientists and mathematicians. • CONTENT LEARNING: Argumentation activities are in the service of student learning of specific disciplinary content knowledge. • EDUCATIONAL STANDARDS: Argumentation is framed as a key instructional goal in educational standards documents (e.g., NGSS, Common Core Mathematics). • EVERYDAY REASONING/COMMUNICATION: Constructing and evaluating arguments is important to students' everyday reasoning and communication. • EQUITY: Culturally and linguistically responsive approaches to argumentation can engage and support nondominant students' equitable participation in math and science classrooms.
3. What methods were used to measure (a) student argumentation or (b) teachers' practices, skills, or beliefs for teaching argumentation?	
Measurement method of student argumentation	• SHORT-ANSWER RESPONSES (e.g., in response to structured assessment questions and to be coded with a rubric) • ESSAYS AND WRITTEN PROOFS (e.g., longer form written arguments, such as a scientific report or mathematical proof) • MULTIPLE-CHOICE TEST (e.g., with correct and incorrect answers) • QUESTIONNAIRE/SURVEY (e.g., Likert scale capturing beliefs about argumentation) • WHOLE-CLASS DISCUSSIONS (e.g., classroom of students engaged in discussion with each other) • SMALL-GROUP DISCUSSIONS (e.g., small groups of two or more students engaged in discussion with each other) • RESEARCHER-STUDENT INTERVIEWS (e.g., one-on-one interviews with students) • TEACHER-STUDENT INTERACTIONS (e.g., analysis of student arguments in response to teacher questioning) • ONLINE STUDENT DISCUSSIONS (e.g., analysis of student discussions in an online platform, potentially asynchronous)
Measurement method of teacher practices, skills, and beliefs	• SHORT-ANSWER RESPONSES (e.g., in response to structured questions outside of an in-depth researcher interview context) • TEACHER REFLECTIONS (i.e., teachers' unstructured reflection on practice, pedagogy, or learning about argumentation, such as journaling) • LESSON PLANS (e.g., analysis of teachers' lesson plans for supporting student argumentation) • MULTIPLE-CHOICE TEST (e.g., with correct and incorrect answers assessing teachers' pedagogical content knowledge for supporting student argumentation) • QUESTIONNAIRE/SURVEY (e.g., Likert scale capturing beliefs about argumentation) • WHOLE-CLASS DISCUSSIONS (e.g., teacher moves, such as questioning in whole-class discussions)

Coding field	Response options
	• SMALL-GROUP DISCUSSIONS (e.g., teacher moves, such as questioning in interactions with small groups of two or more students) • TEACHER-STUDENT INTERACTIONS (e.g., teacher moves, such as questioning in one-on-one interactions with individual students) • RESEARCHER-TEACHER INTERVIEWS (e.g., one-on-one interviews with teachers) • CLINICAL SIMULATIONS (e.g., structured teaching simulations with research staff trained to act as students in standardized, scripted ways) • PROFESSIONAL DEVELOPMENT OBSERVATIONS (e.g., analysis of teachers discussing pedagogical strategies for facilitating argumentation) • ONLINE STUDENT DISCUSSIONS (e.g., analysis of teacher moves in an online discussion forum with students, potentially asynchronous)
Teacher measurement construct	• PRACTICE (e.g., teacher moves in classrooms or clinical simulations) • SKILLS OR KNOWLEDGE (e.g., pedagogical content knowledge) • BELIEFS (e.g., about which students are ready to engage in argument)
Teacher instructional context	• WRITTEN: Teacher is supporting students' written arguments in essays, short-answer responses, proofs, etc. • SPOKEN: Teacher is supporting students' spoken arguments in classroom discussions, pair interactions, etc.
4. What new knowledge was produced about argumentation in naturalistic, business-as-usual settings that might aid in the development of new interventions?	
Knowledge learned in naturalistic, business-as-usual contexts	• STUDENT CHALLENGES: identifying specific aspects of argumentation that students struggle the most with and could benefit from targeted support • STUDENT COMPETENCIES: identifying nascent forms of argumentation that students exhibit that interventions could further build on • TEACHER CONCEPTUALIZATIONS: identifying teacher beliefs about argumentation that might shape how they teach or use interventions about supporting student argumentation • TEACHER PRACTICE: identifying how teachers currently support student argumentation in their classrooms • CURRICULUM/INSTRUCTIONAL TOOLS: identifying existing curriculum resources or instructional tools already in use for supporting argumentation
5. What instructional scaffolds and student-focused interventions were developed and tested to support student argumentation?	
Student supports (Wecker & Fischer, 2014; Wilson-Lopez et al., 2020)	• WRITING PROMPTS: Prior to writing arguments, students complete a graphic organizer, template, or question prompts (using visual tools or not). • DISCUSSION PROMPTS: prompting of activities or reflection questions for contributing to a discussion • VISUAL TOOLS: tools for supporting argumentation involving the representation of statements or propositions (claims, arguments, etc.) and their relations in an external visual representation • EXPLICIT ARGUMENTATION INSTRUCTION: Students explicitly learn the features of an argument through definitions of components, exemplar texts, or evaluation tools, such as rubrics and checklists.

Coding field	Response options
	• EXAMPLES: Students learn about an example scenario or data for developing initial arguments or evaluating others' arguments. • WRITTEN FEEDBACK: Students receive written feedback from peers, teachers, or technology systems to improve their written arguments. • CRITIQUE/CONTRASTING CLAIMS: Students hear, critique, or contrast others' claims, taking into account the evidence or warrants for that claim. • TECHNOLOGY SUPPORTS: Students interact with technology to support their argumentation (e.g., automated feedback, graphic tools, game-based learning). • PRESENTATION/PERSUASION: Students (a) convince a given audience of the argument's claim, (b) explain to a given audience why the claim is true, or (c) demonstrate the argument's validity to a given audience. • ACTIVITY STRUCTURE: Students engage in particular activity structures, such as card sorts or Science Seminars/Socratic circles designed to support argumentation.
6. What professional development and other teacher-focused interventions for teaching argumentation was developed and tested?	
PD products (CADRE, 2014)	• STAND-ALONE: stand-alone instruction, manuals, guides, or other information • WORKSHOPS: workshops or meetings for practicing teachers • CE CREDITS: continuing education credits for practicing teachers • PRESERVICE: preservice curricula for university courses • NETWORKS: networks or communities of practice (e.g., online discussion forums) • COACHING: supervision, coaching, or mentoring
PD duration	• 1 DAY: single PD session lasting 1 calendar day or less • 1 WEEK: PD lasting 1 week or less but more than 1 calendar day • 1 MONTH: PD lasting 1 month or less but more than 1 week • MULTIPLE MONTHS: PD lasting more than 1 month • UNCLEAR: duration unclear/not reported
PD activities (Lynch et al., 2019)	• EXAMPLE STUDENT WORK: Teachers study examples of students' work (e.g., example written proofs, recorded dialogue). • OWN STUDENTS' WORK: Teachers study examples of their own students' work. • DEMONSTRATION: Teachers observed a video of live demonstration/modeling of instruction to support argumentation. • SELF-OBSERVATION: Teachers watch video from their own classroom practice. • CASE STUDIES/VIGNETTES: Teachers read and analyze hypothetical case studies or vignettes of classroom argumentation (distinct from real examples). • SOLVED PROBLEMS: Teachers solved problems or worked through student materials during the PD (e.g., experiencing parts of the curriculum from the students' perspective). • LESSONS PLANS: Teacher developed curricula or lesson plans during the PD. • TEACHER REFLECTION: Teachers hear other teachers reflect on their own practice about argumentation.

Coding field	Response options
Instructional strategies and curriculum resources (McNeill et al., 2016)	• STRUCTURAL INSTRUCTIONAL STRATEGIES: Teachers learned strategies for strengthening how students use evidence and explicitly linking evidence to claims (e.g., prompting students to use evidence, asking about counterevidence). • DIALOGIC INSTRUCTIONAL STRATEGIES: Teachers learn strategies for creating classroom norms and prompting argumentative discourse (e.g., creating a culture of questioning and justification, engaging students with competing claims). • CURRICULUM: Teachers were taught how to use specific structured curriculum activities, technological tools, or lesson plans for teaching argumentation, either in general or linked to specific disciplinary topics. • OTHER: Teachers learn other types of instructional strategies not previously listed (use sparingly; this code is meant to help identify potential other more specific codes). • NONE: No specific instructional strategies were taught, or unspecified.
7. What research methods were used to study teacher-focused and student-focused interventions?	
Study design	• CORRELATIONAL: analysis of data collected at one point in time to examine relationships between two or more variables (e.g., between students' argumentation performance and conceptual knowledge) • OBSERVATIONAL/DESCRIPTIVE: classroom observations, interviews, or other related methods for understanding how students and teachers engage with the intervention (usually qualitative data but could include descriptive quantitative analysis, e.g., counts, frequencies) • PRE-POST: study of pre-post gains in one group of individuals who received the same intervention (e.g., student performance before and after the intervention) • LONGITUDINAL: study of change over time through receiving an intervention (must measure for more than two time points to distinguish from pre-post) • QUASI-EXPERIMENTAL GROUP DESIGN: nonrandomly assigned intervention and comparison groups being compared • EXPERIMENTAL: subjects randomly assigned to a treatment condition (intervention group) or business-as-usual/alternative treatment condition (comparison group)
8. What types of evidence were generated about the implementation and outcomes of argumentation interventions and resources?	
Student implementation and outcomes	• INTERVENTION IMPLEMENTATION: evidence about how students engaged with the intervention (e.g., chat logs, classroom observations) • ARGUMENTATION IMPROVEMENT: evidence that the intervention improved students' argumentation enactment or skills • CONTENT KNOWLEDGE IMPROVEMENT: evidence that the (argumentation-related) intervention improved students' disciplinary content knowledge
Teacher implementation and outcomes	• INTERVENTION IMPLEMENTATION: evidence about how teachers engaged with the intervention or viewed it (e.g., discussions during PD sessions, interviews with teachers about their experiences with the resources) • CLASSROOM PRACTICE CHANGE: evidence that the intervention changed how teachers supported student argumentation in the classroom • BELIEF/KNOWLEDGE CHANGE: evidence about how the intervention changed teachers' beliefs about argumentation (e.g., about what "argumentation" means, why it is important, how to support it in the classroom) or instructional knowledge about argumentation (e.g., pedagogical content knowledge)

Coding field	Response options
9. What types of products and content did projects disseminate to researcher and educator audiences?	
Product type (CADRE, 2014)	• Conference presentation slides or poster • Conference paper • Journal articles • Academic journals • Practitioner journals • Journals (not specified) • Websites • Workshops • Newsletters • Commercial products or publications • CDs/DVDs • Popular media • Reports (not articles or books) • Social media • Blogs • Books/book chapters • Webinars • White or working papers

About the American Institutes for Research

Established in 1946, with headquarters in Arlington, Virginia, the American Institutes for Research® (AIR®) is a nonpartisan, not-for-profit organization that conducts behavioral and social science research and delivers technical assistance to solve some of the most urgent challenges in the U.S. and around the world. We advance evidence in the areas of education, health, the workforce, human services, and international development to create a better, more equitable world. The AIR family of organizations now includes IMPAQ, Maher & Maher, and Kimetrica. For more information, visit AIR.ORG.

AIR® Headquarters
1400 Crystal Drive, 10th Floor
Arlington, VA 22202-3289
+1.202.403.5000 | **AIR.ORG**