

REVIEW

Open Access



# Professional development programs to improve science teachers' skills in the facilitation of argumentation in science classroom—a systematic review

Raphael Wess<sup>1,2\*</sup> , Burkhard Priemer<sup>1</sup>  and Ilka Parchmann<sup>2</sup>

## Abstract

Argumentation is crucial to fostering scientific reasoning and problem-solving in science education. However, researchers and teachers still report problems facilitating argumentation in the classroom. This may be influenced by the design of the corresponding professional development programs (PDPs) and the focus of the underlying research. To describe the state of the research regarding science teacher PDPs on facilitating argumentation, we systematically reviewed publications from the last 20 years in terms of their design, with selected low- and high-inference characteristics, for example, in terms of the addressed professional competence and the argumentation framework. The results illustrate a broad spectrum of teacher PDPs on argumentation in terms of formal-structural aspects (e.g., sample size) and interests (e.g., methodology). We found, for example, that pre-service teachers' argumentation PDPs are less frequent than in-service teachers' argumentation PDPs and that research rarely focuses on situation-specific competencies, such as professional vision or decision-making. Additionally, we report challenges in analyzing the argumentation framework and discuss possible reasons for this. We critically evaluate these and other findings, point to fruitful directions for further research and reviews, and inform practitioners of professional development of argumentation.

**Keywords** Science teacher, (Scientific) argumentation, Professional competence, Professional development, Systematic review

## Introduction

Research on argumentation has received substantive interest in science education, since constructing scientific arguments can foster science learning (*arguing to learn*) and be an important scientific skill of its own (*learning to argue*) (Asterhan & Schwarz, 2016; Driver et al., 2000; von Aufschnaiter et al., 2008). Argumentation

is commonly seen as a discursive act that involves the “complex process of reasoning utilized in situations that require content knowledge to construct and/or critique proposed links between claim and evidence” (Osborne et al., 2016, p. 825). This gives argumentation the potential for enabling collaborative and critical discourse based on evidence and makes it a cornerstone of science communication and the generation of scientific knowledge (Jiménez-Aleixandre & Erduran, 2007). Acknowledging these benefits, it has also found its way into recent policy documents (Hazelkorn et al., 2015; Kultusministerkonferenz, 2020a; Kultusministerkonferenz, 2020b; National Research Council, (2012)). Mutually reinforcing the prominent role of argumentation has led to a growing body of research in science education.

\*Correspondence:

Raphael Wess  
raphael.wess@hu-berlin.de

<sup>1</sup> Physics Education, Humboldt-Universität Zu Berlin, Newtonstraße 15,  
12489 Berlin, Germany

<sup>2</sup> Chemistry Education, Leibniz Institute for Science and Mathematics  
Education at Kiel University, Kiel, Germany



© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

Different approaches have been undertaken to analyze and classify argumentation and its related skills (e.g., by conceptualizing a pedagogical content knowledge (PCK) of argumentation) (McNeill et al., 2016) or operationalizing argumentation competence (Rapanta et al., 2013). However, argumentation is a highly complex process involving numerous activities, and therefore can lead to a variety of different understandings of what counts as argumentation (Rapanta et al., 2013). Hence, the underlying conceptualization of argumentation used in professional development programs (PDPs) and in research studies often differs. There is evidence that this ambiguity makes it hard for teachers to obtain a consensus understanding of argumentation and its aims (Katsh-Singer et al., 2016), and to facilitate it in their classrooms (Choi et al., 2021; McNeill & Knight, 2013; McNeill & Pimentel, 2010; Sampson & Blanchard, 2012).

We want to analyze this issue by systematically reviewing previous argumentation PDPs and their associated research. In recent decades, a considerable number of argumentation PDPs aiming to improve facilitation in science lessons have been developed and evaluated, but a summative overview is missing regarding the following three aspects.

First, there has been no systematic and comparable analysis of how research on PDPs in argumentation is framed theoretically in terms of the underlying argumentation conception. Nor has there been appropriate analysis of the professional competence this research addresses, especially with regard to the situated competencies of teachers. Second, it has not been investigated how argumentation PDPs are described, and how the results relate to selected influential characteristics of effective professional development (PD). Third, we want to review the influences that the authors of the research papers attribute to the PDP results.

Our aim in relation to the last two aspects is not to make a whole new inventory of characteristics from which to draw conclusions on effective PD strategies in general. There are studies on effective PDP characteristics in education (Borko et al., 2010; Desimone, 2009; van Driel et al., 2012) as well as for argumentation (Aydeniz, 2019; Weiss et al., 2022; Zembal-Saul & Vaishampayan, 2019), which are relevant for developing argumentation skills that enable its facilitation. The cited three argumentation reviews focused on argumentation approaches regarding inquiry or on a comparison of the two approaches, yet they did not provide a comprehensive summative overview of different argumentation approaches regarding professional competence and the underlying argumentation frameworks. To fill this gap, we want to specifically analyze published research studies on argumentation PDPs regarding these two aspects.

This is accompanied by an analysis of general PD characteristics and their attributed influences by the different authors to identify research trends or desiderata at this high resolution level.

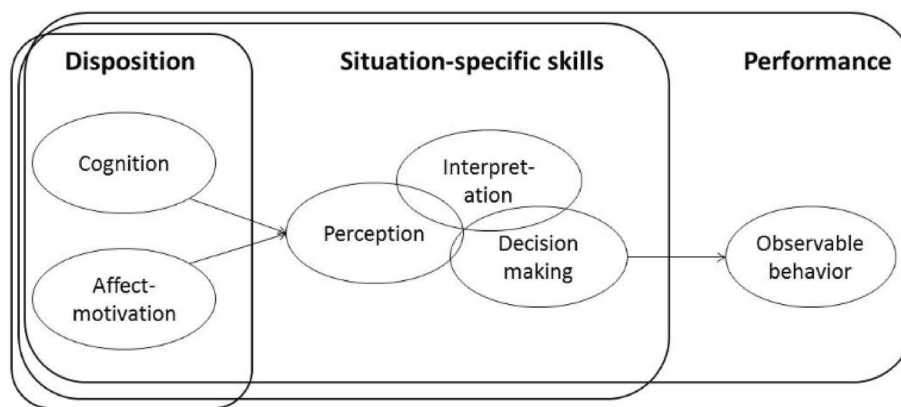
Overall, our review may be helpful in providing recommendations for yet unattended research foci in argumentation and for designing new argumentation PDPs. To address these issues, we first define professional competence, its development, and its role in PDPs before focusing on these for argumentation in science teacher education. Then, we describe the methods used to search for and review PDPs on argumentation. Next, we present the results of the review based on the categories related to the aforementioned topics. In our discussion, we contextualize our results. This paper concludes with recommendations for researchers by outlining new perspectives for argumentation PDPs, and for practitioners to improve the development of future science teacher PDPs that aim to enhance argumentation.

## Theoretical background

### Professional competence

When describing the relevant knowledge and concrete skills and abilities of a teacher regarding his or her professional knowledge (e.g., concerning argumentation), the term (professional) competence is often used. One of the most influential definitions describes competence “as the ability to meet individual or social demands successfully, or to carry out an activity or task” (OECD, 2002, p. 8). Despite the ambiguity of the term competence and the resulting variety of uses (Schneider, 2019), the term still best summarizes the goals and requirements of teachers’ abilities to successfully orchestrate learning arrangements (Kunter et al., 2013).

In the following, we rely on the model of Blömeke et al. (2015) that describes competence (i.e., the ability to successfully meet certain demands) as a continuum. Blömeke et al. (2015) emphasize that a commonly assumed dichotomy between dispositional skills (e.g., knowledge and affective-motivational traits, as latent variables), and performance (as an observable variable) is not specific enough to describe the affordances required to perform successfully in complex, real-life classroom situations. They point out the importance of situation-specific characteristics, making competence a “latent cognitive and affective-motivational underpinning of domain-specific performance in varying situations” (Blömeke et al., 2015, p. 3). The development of professional competence is, hence, more of a continuous, mutually influencing process. Accordingly, the authors divide this process into three areas: disposition, situation-specific skills, and concrete performance (see Fig. 1).



**Fig. 1** Continuum of professional competence (Blömeke et al., 2015)

Disposition includes cognitive prerequisites and motivational-affective variables. This is in line with Kunter et al.'s (2013) model, which more precisely differentiates the subdomains of cognition and affect-motivation. Cognition is understood as professional knowledge, which is an abstract psychological construct describing the knowledge of a teacher relevant to professional fields of action. Among the motivational-affective traits, Kunter et al. (2013) identified variables relating to beliefs, motivational orientations, and self-regulation skills. These dispositional variables are regarded as prerequisites for situation-specific skills, which, in turn, represent the repertoire of actions available for the selection and execution of a teacher's concrete actions in a teaching situation.

This view of competence as a continuum or process variable is in accordance with the conclusion that (professional) competence and its facets are changeable and thus learnable (Kunter et al., 2013). This indicates that the development of professional competence is a central task of teacher education. This can also be assumed for argumentation, since the ability to argue and its relevant skills have been described as a competence (Osborne et al., 2013; Rapanta et al., 2013). With this in mind, argumentation viewed as a competence can be addressed within the PD of teachers and hence can be learned by them.

#### Development of professional competence

By PDPs on argumentation, we mean all training in argumentation competence that is related to the profession of a (science) teacher with the aim to facilitate argumentation in the science classroom. The institutional design of PD varies from country to country. Regardless of the phasing and their respective weighting within national curricula, PDPs globally share common goals (e.g., promoting argumentation) or, generally speaking, the overall goal of developing professional competence in the teaching profession.

The view of in-service training as a one-way knowledge transfer is now considered outdated, and viewing competence as a dichotomy, rather than a continuum, is now viewed as an oversimplification (Blömeke et al., 2015). The view of competence as a continuum is supported by the evidence that effective measures for PD should not solely focus on the transfer of dispositional aspects (e.g., professional knowledge) but also take into account many different situation-specific aspects. Many frameworks exist that try to conceptualize effective PD concerning these different aspects (i.e., "effective" PD improvements in teachers' competence) (Borko et al., 2010; Desimone, 2009; Opfer & Pedder, 2011). PDPs are effective if teachers become able to increase their knowledge and competence and / or have specific changes in their attitudes and beliefs. They can then use this in learning situations to plan, determine, and improve their instruction, thus increasing student learning (Desimone, 2009). Desimone (2009) proposed five core features of effective PDPs: *content focus*, *active learning*, *coherence*, *duration*, and *collective participation*.

For Desimone, *content focus* is the most influential characteristic. Based on previous research, she advocates for the "link between activities, that focus on subject matter content and how students learn that content with increases in teacher knowledge and skills, improvements in practice, and, to a more limited extent, increases in student achievement" (Desimone, 2009, p. 184). *Active learning* considers opportunities for teachers to engage in PDPs. This definition covers a range of activities (e.g., making own practical experiences in the classroom, participating in feedback, and debating, discussing, and reviewing student work). *Coherence* is seen as the extent to which the activities of a PDP are consistent with the participating teachers' knowledge and beliefs. *Duration* means that changes in teachers' abilities require time. The duration of a PDP over a specific period is

important for effectiveness, although there is no “tipping point” for PDPs to be effective (Desimone, 2009). Considering the duration of a PDP more closely, Guskey and Yoon (2009) suggested assessing contact hours. Contact hours refer to the amount of time the participants interacted face-to-face with a trainer. The information gathered from their review showed that PDPs had positive effects when contact hours exceeded 30 h (20 h in Desimone, 2009). The last feature, *collective participation*, is seen as an arrangement based on shared participants’ characteristics (e.g., same school or grade) that provides momentum for potential interaction or discourse. Similarly, in another PDP review, Borke et al. (2010) emphasized collective participation, active learning possibilities, and the effect of school student learning on the intended content.

Thus, there is consensus among researchers that, inter alia, the duration of the PDP and active learning (e.g., in the sense of practical activities) are important factors in the development of teachers’ professional competence.

#### Scientific argumentation as professional competence

The aforementioned thoughts on teacher PD and the importance of practical activities for it to be beneficial also consistently hold for the PD of argumentation competence.

Considering our research aim, one should keep in mind the different uses of terms in the field. A popular distinction is made between *argument* as a product and *argumentation* as a discursive process (Osborne et al., 2004). In simple terms, an argument is the referent to a claim backed by evidence that is—based on scientific principles or reasoning—justified with structure and content in varying complexity (McNeill & Krajcik, 2011; Osborne et al., 2004). The discursive process or exchange of arguments in a specific context is labeled as argumentation. Given this understanding, the term *arguing* refers to a participant being actively involved in argumentation by providing, elaborating on, or judging (parts of) an argument. This results in a common understanding of (scientific) argumentation as a dialogic process with the aim of persuasion or consensus based on a scientific issue, which is based on and results in an exchange of arguments in a scientific context (e.g., in a science classroom) (Jiménez-Aleixandre & Erduran, 2007).

It is important to note that argumentation can still be operationalized in various ways, and cover a wide range of activities (Aydeniz, 2019; Rapanta et al., 2013). A means of investigating an argumentation framework lies in its ability to allow a comparison of the results of different argumentation studies.

*First, it is important for researchers to understand that analytic frameworks [...] are tools created for specific tasks to investigate specific questions. Frameworks, therefore, are not fully interchangeable, and the foci of each framework require consideration before comparing the results of various studies. (Sampson & Clark, 2008, p. 469)*

Rapanta et al. (2013) suggest that defining argumentation as a competence should consider the choice in research to analyze and/or assess argumentation. Analysis refers to different approaches in terms of arguments. The authors propose the following three distinctions to summarize this variety in the educational context: argument as a *form*, argument as a *strategy*, and argument as a *goal*. Argument as a form refers to the investigation of structural aspects (e.g., as provided by the claim-evidence-reasoning [CER] framework) (McNeill & Krajcik, 2011) or Toulmin’s Argument Pattern [TAP] (Erduran et al., 2004). Argument can also be seen as a strategy that results in the analysis of (dialogical) argument moves that in turn result in highly context-specific use and identification (Rapanta et al., 2013). In the argument as a goal perspective, these aspects are summarized, emphasizing the aim of argumentation (e.g., reaching a consensus) (Berland & Lee, 2012) or individual vs. communal understanding (González-Howard & McNeill, 2019).

Assessment, in turn, is divided into metacognitive, metastrategic, and epistemological knowledge. Metacognitive knowledge refers mainly to declarative knowledge (“know-what”). This includes knowledge about the structure of an argument and its conceptual or epistemic quality. Metastrategic refers to procedural knowledge (“know-how”). This includes studies that examine or train a) specific argumentative discourse elements rather than others or b) the implementation of certain argumentative strategies that presuppose (a high level of) metacognitive knowing. Epistemological knowing focuses on knowing about knowledge in general and in relation to a persons’ knowledge (“know-be”-skills). In terms of argument quality, the study thematizes the a) relevance, b) sufficiency, or c) acceptability of an argument. This refers mainly to studies involving problem-solving or conceptual change through collaboration (collaborative learning). Since we want to cover a variety of PDPs on argumentation, we rely on the extensive yet detailed assumptions made for argumentation as a competence by Rapanta et al. (2013).

#### Previous reviews of PDPs on argumentation and the need for a new one

Since the importance of argumentation and its development is commonly known, some work has been done to review the research on argumentation and associated

PDPs. Previous reviews on argumentation in science education have focused on the theoretical side to sharpen argumentation as competence (Rapanta et al., 2013) or examine general research trends based on linguistic or epistemic criteria (Erduran et al., 2015). In addition, on the practical side, previous reviews have examined argumentation to assess student argumentation (Sampson & Clark, 2008), to foster scientific literacy in K-12- (Cavagnetto, 2010), or to group argumentation studies thematically in K-8-contexts (Bağ & Çalik, 2017).

Additionally, some reviews explicitly consider the PD of argumentation (Aydeniz, 2019; Weiss et al., 2022; Zembal-Saul & Vaishampayan, 2019; Zohar, 2007). Zohar (2007) suggested that the elements for successful PDP on argumentation effort a turn in orchestrating classroom discourse and considering the importance of scientific reasoning in the classroom. Roughly a decade later, Zembal-Saul and Vaishampayan (2019) reemphasized the need for a greater focus on scientific practices and discourse in light of reform for any successful PDPs on argumentation. With this, they highlighted the importance that future teachers should also experience these scientific practices and discourse styles they later teach children.

Weiss et al. (2022) characterized immersive argument-based inquiry (ABI) learning environments in school education. Weiss et al.' review focused mostly on ABI-related learning environments, but it also identified the teacher actions within a learning environment as one of three central categories, containing the following typical actions as common elements of immersive ABI learning environments:

- Encouraging argumentation
- Providing resources
- Asking questions
- Sharing authority
- Communicating norms
- Modeling dialogue or language use
- Emphasizing important ideas

They additionally identified student actions (e.g., small-group work and engaging in argumentation) as well as generative opportunities (e.g., student authorship of the initial activity) as the other two commonly occurring elements.

Additionally, Aydeniz (2019) reviewed the studies of two curriculum design projects—the IDEAS project and the Argument-Driven-Inquiry (ADI) project—and presented two main findings regarding pre-service chemistry teachers' experiences of teaching science through argumentation: First, pre-service teachers lack a "sophisticated epistemology related to the way chemical knowledge gets constructed, evaluated and critiqued due to the

experiment-driven nature of chemistry" (Aydeniz, 2019, p. 25). Second, he highlights the "importance of mastery experiences and argumentation-based teaching resources in teachers' ability to facilitate argumentation in classroom" (Aydeniz, 2019, p. 25). With regard to them, he suggested design principles for argumentation tasks and design principles for group argumentation:

Design principles for argumentation tasks (Aydeniz, 2019, p. 26):

- Any argumentation-based learning task should make the goal of the activity explicit to the students
- Learning tasks should engage students in culturally relevant and academically important problems
- Learning tasks should provide explicit scaffolding for students' cognitive engagement with the chemical ideas and social engagement with members of the learning community
- Learning tasks should engage students in epistemic questions about the nature of chemical knowledge (e.g., "What is the evidence behind claims to knowledge?" or "Is the evidence provided relevant and sufficient?")
- Structures must be put in place for epistemic and cognitive divisions of labor so that all students are equally engaged with the construction, evaluation, and critique of knowledge
- Teachers must create conditions for all students to equally exercise their epistemic authority for making meaningful contributions to the shared knowledge-building activities

Design principles for group argumentation (Aydeniz, 2019, p. 26):

- The instructor should explicitly communicate the goals of the group assignment to the students
- The teacher should introduce the students to the linguistic, mechanistic, epistemic, and social aspects of argumentation
- Teachers should use assessment strategies that promote both positive interdependence and individual accountability

Overall, the above-mentioned reviews provide insight into different aspects of scientific argumentation, with Aydeniz (2019) outlining challenges and providing design principles for PDPs on argumentation for science teachers and Weiss et al. (2022) discussing potential steps to build further understanding of teachers' actions. Yet, their foundation is based solely on three approaches (IDEAS, ADI, and ABI) of the many different existing argumentation programs, so the suitability of these

results for other PDPs remains unclear. Thus, it seems useful to investigate the characteristics of argumentation PD, because even though teachers seem to acknowledge the importance of argumentation in general, their understanding of what counts as argumentation still differs (Katsh-Singer et al., 2016).

Additionally, they still seem to struggle to implement the argumentation processes outlined above in their teaching practice (Choi et al., 2021; McNeill & Knight, 2013; Pimentel & McNeill, 2013; Sampson & Blanchard, 2012). However, and perhaps most importantly, no review has addressed the recently developed models of professional competence that emphasize the importance of situated competencies of teachers (Blömeke et al., 2015). It remains unclear how and if this situated perspective has been represented in previous publications on argumentation PDPs. We argue that this is one beneficial approach for concretizing the generally stated need for support for teachers in facilitating argumentation in the classroom (Henderson et al., 2018).

In conclusion, we argue that a review that specifically analyzes professional competence and the underlying argumentation frameworks of research studies investigating PDPs for science teachers on argumentation is a reasonable attempt to investigate the merits of research on argumentation PDPs for science teachers. This will be accompanied by reviewing active learning and the duration of PDPs on argumentation, and identifying the reasons the authors of the associated research attribute to the results regarding argumentation PDPs. With this, we strive to provide an overview of the state of argumentation PDPs and to examine future research directions. Motivated to shed light on this, the following research objectives emerged.

### Research objectives

(RO1) Identifying the focus of accompanying research on teacher PDPs regarding argumentation in terms of a) the development of their professional competence and b) the underlying argumentation framework.

(RO2) Examining the description of PDPs on argumentation in the accompanying studies, and how PDPs on argumentation attend to the two selected influential characteristics of effective PD: active learning and duration.

(RO3) Analyze, how the characteristics of RO2 account for the reported results, and what influences do the authors themselves highlight in terms of the PDP results.

### Methods

Since we aim to examine the focus of research studies accompanying PDPs on argumentation to evaluate the characteristics of these PDPs and their influences on the

results, we decided to use a systematic review approach. The aim of a systematic review is the summarization of existing research, with the help of given criteria, to provide an overview and conclusions about a specific topic (Newman & Gough, 2020). Methodologically, we followed the steps of PRISMA (Moher et al., 2009, 2015) and the recommendations of Newman and Gough (2020) for reviews in education. This section begins with further elaboration of the research objectives and the categories emerging from these objectives. On this basis, we report the identification of search criteria, the search process, and the coding of the individual studies based on the categories in the coding manual.

### Operationalization of research objectives and category development

We chose our categories based on the theoretical assumptions made previously. The research focus (RO1) was examined in terms of general study characteristics, professional competence, and argumentation framework (see Table 1). The general characteristics included the methodology used (quantitative, qualitative, or mixed-methods design) and the identification of the dependent variable(s) (Bağ & Çalik, 2017). Based on the dependent variable(s), the professional competence on which a study focused was identified. This allowed us to make inferences about the different PDP studies and sort them by their contribution to the development of professionals' argumentation competence. Professional competence was categorized, according to Blömeke et al. (2015), as disposition, situation-specific skills, and performance. *Disposition* refers to a teachers' latent traits (e.g., professional knowledge, beliefs, motivation, and self-regulation) (Kunter et al., 2013). *Situation-specific skills* cover research focusing on the perception and interpretation of a situation, or, as commonly represented in science education research, the concept of professional vision (Goodwin, 1994; Seidel & Stürmer, 2014; van Es & Sherin, 2002a, 2002b). Decision-making, another part of situation-specific skills, was coded for studies that investigated and therefore addressed teachers' ability to choose from different possible reactions in a situation. *Performance* covered studies examining an observable concrete action undertaken in the classroom (e.g., concrete instruction within an argumentation task or a recorded teacher-student talk used for further discourse analysis).

Finally, we aimed to examine the underlying argumentation framework since it provides insight into the design and conceptualization of PDPs' argumentation activities. By focusing on underlying argumentation frameworks, we try to unpack the theory of argumentation used and do not refer to frameworks used to *teach* argumentation.

**Table 1** Categories for classifying the research focus – overview

Domain	Category	Source	Operationalisation
General characteristics of study	Methodology	-	<ul style="list-style-type: none"> <li>• Quantitative</li> <li>• Qualitative</li> <li>• Mixed methods</li> </ul>
	Dependent variable (DV)	-	<ul style="list-style-type: none"> <li>• <i>Free text, summarized</i></li> </ul>
Professional competence	Model of professional competence	Blömeke et al. (2015)	<ul style="list-style-type: none"> <li>• Disposition</li> <li>• Situation-specific</li> <li>• Performance</li> </ul>
	Focal issues	Sampson and Clark (2008)	<ul style="list-style-type: none"> <li>• Structure</li> <li>• Content</li> <li>• Justification</li> </ul>
Argumentation framework	Argumentation competence based on analysis and assessment	Rapanta et al. (2013)	Analysis: <ul style="list-style-type: none"> <li>• Form</li> <li>• Strategy</li> <li>• Goal</li> </ul> Assessment: <ul style="list-style-type: none"> <li>• Metacognitive</li> <li>• Metastrategic</li> <li>• Epistemological</li> </ul>

We tried to obtain information on the argumentation focus by using the frameworks of Sampson and Clark (2008) and Rapanta et al., (2013). We chose Sampson and Clark’s framework because we assumed that the three focal issues—*structure*, *content*, and *justification*—are not only suitable to describe the assessment of student argumentation but also to categorize the main argumentation focus of a research study. This is reasonable since the assessment of student argumentation is quite likely to be an explicit or at least implicit topic of any PDP on argumentation. Additionally, we considered the description of argumentation competence by Rapanta et al., (2013) as a promising second approach for a combined examination of the structural and content aspects of the broad field of PDP on argumentation. Using this framework, we conducted an argument analysis (*form*, *strategy*, and *goal*) and assessed the approaches of argumentation knowledge (*metacognitive*, *metastrategic*, and *epistemological*) of the studies. In summary, an analysis of the frameworks used in PDPs provides important information on how PDPs are tied to theory.

Note that all categories in Table 1—methods, professional competence, dependent variable, and argumentation focus—are study variables, so they are related to *the research focus* of the studies accompanying a PDP. It is, in general, possible to identify all the professional competencies addressed in a PDP, but this requires checking the course material and concrete instructions that are often not included in research papers. Due to economic reasons and based on anticipating a small response rate of the authors of the research papers, we decided against an additional analysis of the PDP curricula. We believe that a focus on the study variables still indicates the core

aims of a PDP and hence reflects the area with the most emphasis. This procedure does not exclude the possibility that more than the reported competencies were addressed in the PDPs, which is assumably likely.

To examine the PDP characteristics (RO2), we extracted the provided description of the PDP activities from each individual study. We analyzed this information by applying categories that address two of the above-mentioned features of effective professional teacher development—*active learning* and *duration* (Desimone, 2009). The feature *active learning* was operationalized here in the presence of active teaching situations with school students (i.e., student inclusion). The extent of a PDP was measured by *duration* and—where applicable—by contact hours. We chose these categories and operationalizations because the possibility of engaging in practice situations has particular importance to teachers’ PD as well as the duration of a PDP. We are aware that there are other gradual nuances for how to conceptualize active learning (e.g., the use of video or text vignettes or role-playing classroom situations), but we only chose to examine if the PDPs included the “most authentic” practical activity—teaching with real students. To contextualize this feature, we classified all studies based on low inference characteristics, such as the participating science teachers’ sample size, career point, and school systems.

Table 2 summarizes which characteristics were analyzed regarding RO2. Again, please note that this review follows a general approach, so no (in-depth) analysis of the concrete use of the described activities, instructions, and materials used in PDPs was conducted; conclusively, we do not investigate Desimone’s

**Table 2** Characteristics of the PDP analysed

Domain	Category	Source	Operationalisation
<b>General characteristics – PDP</b>	PDP description	-	<i>Free text, summarized</i>
	Participants – Number	-	<i>Free text, number</i>
	Participants – Career point	-	<ul style="list-style-type: none"> <li>• Pre-Service teachers</li> <li>• In-Service teachers</li> <li>• Both</li> <li>• Unclear / Not reported</li> </ul>
	Participants – School form	-	<ul style="list-style-type: none"> <li>• Primary school form</li> <li>• Secondary school form</li> <li>• Both</li> <li>• Unclear / Not reported</li> </ul>
<b>Duration<sup>a</sup></b>	Extent	Desimone (2009)	<i>Free text</i>
	Contact hours	Guskey and Yoon (2009)	<i>Free text</i>
<b>Active Learning<sup>a</sup></b>	Practical experience (here: Inclusion of students)	Desimone (2009)	<ul style="list-style-type: none"> <li>• Yes</li> <li>• No</li> </ul>

<sup>a</sup> Feature to examine effective PDPs based on Desimone (2009)

(2009) categories of *collective participation*, *coherence*, and *content focus* of PDPs.

To investigate the results and the attributed reasons (RO3), we first extracted the results from the studies. After filtering the results for those related to a change in teachers' competence based on the PDP, we classified them based on the direction of change (positive, negative, or neutral). We then reviewed the abstract and discussion sections of the studies to identify the reasons for the reported changes that the authors attributed to teachers' competence related to argumentation. Finally, we grouped these reasons into patterns of influential argumentation PD characteristics. Table 3 summarizes which characteristics were analyzed regarding RO3.

Based on the results for both RO2 and RO3, we can provide an overview of how two generally influential PD characteristics are considered in argumentation PDPs and what reason(s) for the influential results the authors attribute.

### Study selection

To obtain a broad view of the research regarding argumentation PDPs, we included studies relating to all aspects of argumentation. Furthermore, we only wanted to include high-quality papers, so we decided to assess study quality using the criterion of peer review. Although the peer review procedure is not completely bias-free and has its own limitations as a quality criterion (Lee et al., 2013), we consider it an appropriate selection criterion for the inclusion of research examining PDPs. Additionally, we limited our search to publications written in English and studies published from 2000 to 2020 (20 years). We are aware that these limitations have raised the potential to miss out on studies that would have met the following search criteria, but we believe that the scope of our review will not be impaired.

### Search criteria

Based on our research objectives and the outlined categories, we developed the following criteria for the inclusion of the research papers:

**Table 3** Categories for the results and associated reasons

Domain	Category	Source	Operationalisation
<b>General characteristics – Study results and associated reasons</b>	Results teacher learning	-	<i>Free text, summarized</i>
	Change in teacher competence	-	<ul style="list-style-type: none"> <li>• Yes</li> <li>• No</li> </ul>
	Direction of change in teacher competence	-	<ul style="list-style-type: none"> <li>• Positive</li> <li>• Neutral</li> <li>• Negative</li> </ul>
	Reason(s) for result(s)	-	<i>Free text, summarized</i>
	Systematisation for reason(s)	-	<i>Inductively developed groups from the category Reason(s) for result(s), see table in results</i>



- (1) Study focuses (at least partially) on the concept of scientific argumentation in PDPs
- (2) Study reports findings based on a PDP or at least one PD activity relating to argumentation
- (3) Study reports on PDPs with pre-service or in-service science teachers
- (4) Study published is peer-reviewed
- (5) Study language is English
- (6) Study is accessible (e.g., no paywall)

### Search procedure

To identify the relevant literature for this review, we searched the databases Scopus, ERIC, and Educational Research Abstracts Online (ERA) in February and March 2020. Based on the research objectives for this review, search terms covering the topics PD, argumentation, and science teacher were developed (Table 4). Table 5 shows the final search terms used for each database.

Note that the search terms are slightly different for each database because of database-specific search operations. Some words in the search terms for the SCOPUS and ERA databases contain an asterisk (\*) to search for all possible variations of a lemma. For example, the lemma *argu\** searches for the words *argument*, *argumentation*, *argue*, *argued*, and *arguing*. For the search in the

ERIC database, the search terms had to be modified since the asterisk function is not provided.

The entire search and selection process is presented in Fig. 2. The database search yielded 993 papers. Additional papers were identified through other sources (e.g., using the “snowball method” by checking the references of the identified studies or referring to a description of a PDP in another study). These were also included in the screening process ( $N=4$ ). After clearing for duplicates, 746 studies remained.

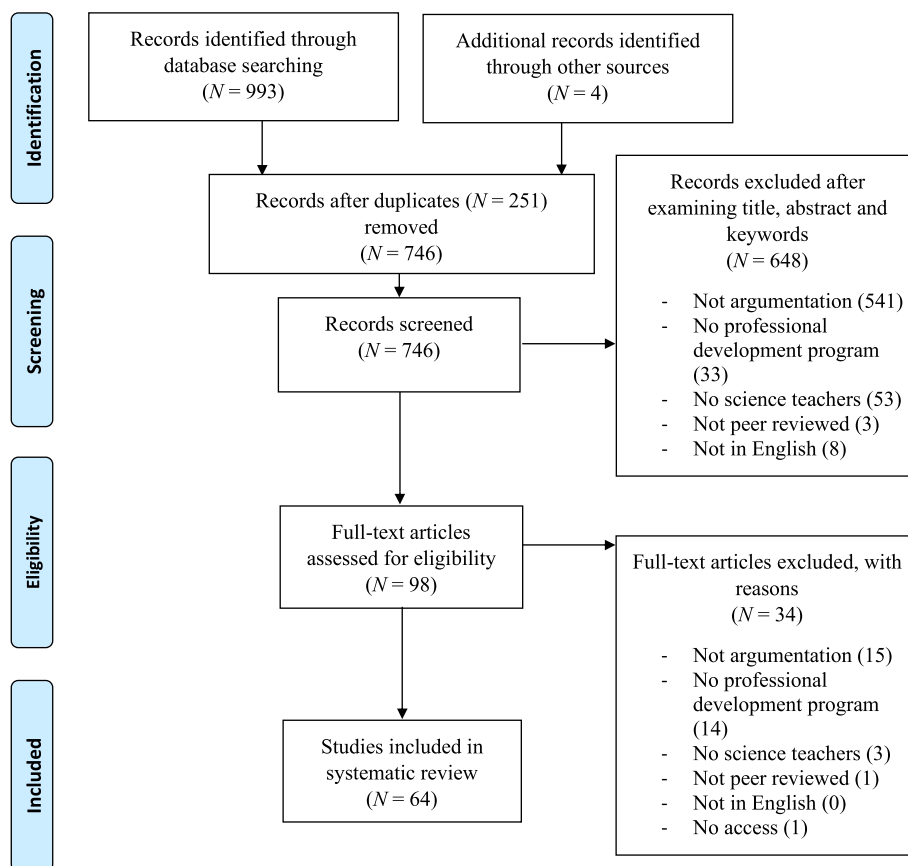
From March to the end of June 2020, we conducted two rounds of screening, in which we first applied the inclusion criteria for suitability to the titles, abstracts, and keywords. Applying the aforementioned criteria, this screening led to the exclusion of 647 papers. The studies were excluded when one inclusion criterion was not fulfilled. As can be seen in Fig. 2, the main reason for exclusion was the absence of scientific argumentation. Often, the word *argu\** returned studies using argumentation in a different way (e.g., to position a paper on a controversial topic: “In this paper, I argue that...” or “We want to present an argument for...”). In the second round of screening, we applied the same criteria to 98 articles that were read completely. Thirty-four articles did not match the inclusion criteria in this round. The main reason for exclusion was also the use of argumentation in another way, followed by studies

**Table 4** Criteria for developing the search term

Criterion	Example keywords	Search terms
Argument	argument, (scientific) argumentation, argue, arguing	scien*, argu*
Science teacher	science teacher, science teachers, pre-service science teacher, in-service science teacher, teach, teaching	scien*, teach*
Professional development	professional development, teacher development, teacher learning, teacher training	professional development, teacher development, teacher learning, teacher training

**Table 5** Databases and search term

Database	Full search term	Date of last search	results	Comment
SCOPUS	TITLE-ABS-KEY ( argu* AND scien* AND teach* AND (“teacher training” OR “professional development” OR “teacher learning” OR “teacher development”) AND PUBYEAR > 2000	06.03.2020	489	Since 2001 set as criterion
ERIC	(argument OR argumentation OR argue) AND (scientific OR science) AND (teach OR teacher) AND (“teacher training” OR “professional development” OR “teacher learning” OR “teacher development”)	06.03.2020	243	Since 2001 (“last 20 years”) and only peer reviewed set as criteria
ERA	[All: argu*] AND [All: scien*] AND [All: teach*] AND [All: “teacher training”] OR [All: “professional development”] OR [All: “teacher development”] OR [All: “teacher learning”] AND [DatabaseType: Educational Research Abstracts Online]	05.03.2020	261	No criteria for further selection in search engine possible, manually exclusion of studies before 2001



**Fig. 2** Flowchart diagram of paper selection process according to PRISMA (Moher et al., 2009)

that mostly provided a general theoretical framing of PDPs on argumentation but did not explicitly present the results of one. The whole analysis of studies, including the various review steps from the 997 initial studies to the 64 included studies, was published and is openly available (Wess, 2023).

This led to the final inclusion of 64 studies reporting on different PDPs for (scientific) argumentation for science teachers. This number does not represent 64 different PDPs but rather studies that are based on a science teacher PDP relating to argumentation. Using this approach, different studies with different research foci sometimes emerged from the same PDP (Christodoulou & Osborne, 2014\*; Osborne et al., 2013), or different studies reported different phases of ongoing PDPs (Fishman et al., 2017\*; Osborne et al., 2019\*). As an overview of the research context is one of the main aims of our review, we decided against further clustering here. In addition, some PDPs build on or re-use the materials of previous PDPs (Larraín et al., 2017\*). Since we were investigating more general characteristics and not the concrete instructions or materials, we did not differentiate when the existing material was (re)used or modified.

### Study analysis / intercoding

The next step was to extract relevant information from the 64 articles, including basic study information and the aforementioned categories. A detailed coding manual for the above-mentioned categories can be found in the published study data repository (Wess, 2023). For the higher inference categories, intercoding was conducted by two raters (Gwet, 2012). The first rater was the corresponding author of this paper, who works in the field of science education research and teacher PD. The second rater was the second author; a physics education expert whose research focus is scientific argumentation. We randomly generated a subsample of the 64 studies, and 11 studies were simultaneously coded by the two raters. To correct for the chance of guessing the same codes, we used Cohen's Kappa (Gwet, 2012) to estimate intercoder reliability (Table 6).

Without considering the argumentation categories (which will be addressed in the results section), the other relevant high-inferential categories of this review had Kappa values in the range of  $K=0.63-1.0$ , which is considered to be good to excellent agreement. One exception is the professional competence category relating

**Table 6** Interrater-Reliability (Cohen's Kappa) for each category ( $N=11$ )

	Domain	Category	Cohen's $\kappa$
<b>Block – RO 1</b>	Professional Competence	Disposition	1.0
		Situation-Specific skills	0
		Performance	.63
	Argumentation Competence – Focal Issues	Structure	N/A
		Content	N/A
		Justification	N/A
	Argumentation competence – Analysis	Form	.01
		Strategy	.30
		Goal	0
	Argumentation competence – Assessment	Metacognitive Knowledge	.01
Metastrategic Knowledge		.28	
Epistemological Knowledge		0	
<b>Block – RO 2</b>	Active learning	Practical experience with students	.81
<b>Block – RO 3</b>	Reasons for results (authors)	Systematisation of reasons	.69

to situation-specific skills ( $K=0$ ), despite having a percentage agreement of 81%. This is due to the missing occurrence of situation-specific skills in the randomly generated subsample of studies, which led to a rating of zero for each case and a low difference in ratings. In this situation, Cohen's Kappa is prone to error, which resulted in this exceptional Kappa value. Based on this circumstance and the considerable percentage of agreement, we concluded that the rating of this category is not coincidental and therefore worth considering in this review.

Besides this, for all the other categories for which the Kappa values indicated at least good agreement in the subsample, we assumed that our coding manual worked and provided reasonable results. In the few cases within these categories for which the codes did not match, discussion between the two raters resolved the discrepancies, and a consensus was reached.

## Results

The results are presented in three parts, each addressing the respective research objective. In all three parts, the descriptive findings for the mentioned categories are reported (Tables 7, 8, and 9). The full table is a single spreadsheet with study data for all three research objectives, and all studies can be openly accessed in the published study dataset (Wess, 2023).

### Results for RO1

We categorized the studies we reviewed by *methodology*, *dependent variable*, and professional competence to obtain an overview of the context of the PDPs on argumentation. Where applicable, the number of occurrences was counted (Table 7).

**Table 7** Research focus characteristics of included studies<sup>a</sup>

Study characteristic	Description	N (%)
Methodology	- Quantitative	7 (10.9%)
	- Qualitative	40 (62.5%)
	- Mixed methods	17 (26.6%)
Professional Competence (based on DV)	- Disposition	48 (75.0%)
	- Situation-specific	3 (4.7%)
	- Performance	35 (54.7%)
Argumentation	Argumentation framework & competence	Dismissed during review process

<sup>a</sup> Percentages may not total 100% due to rounding. Values of professional competence may add up to more than 64 because some studies fit multiple categories of study characteristics

Despite addressing many different dependent variables, the results show that most research settings use a qualitative approach and mainly examine dispositional characteristics (e.g., professional knowledge or beliefs). In approximately half of the studies, an examination of concrete performance was also considered in the research focus.

With the two chosen argumentation frameworks and the corresponding categories, we were, in most cases, neither able to find a consensus on the focal issues nor on the analysis or assessment of argumentation competence. This was caused by the circumstance that the given and/or selected information in the studies was in most cases insufficient to identify an underlying theoretical framework of argumentation, or the information given was indistinguishable for the categories of the two argumentation frameworks used by Rapanta et al., (2013) and Sampson and Clark (2008).

**Table 8** PDP characteristics of included studies<sup>a</sup>

PD characteristic	Description	N (%)
<b>General characteristics</b>	Participants – N	1 – 120
	Participants – career point	
	- In-service	42 (65.6%)
	- Pre-service	22 (34.4%)
	Participants – school form	
	- primary	21 (32.8%)
	- secondary	26 (40.6%)
	- both	5 (7.8%)
	- unclear / not reported	12 (18.8%)
<b>Duration</b>	- Extent	30–60 min – 4 years, single to multiple meetings
	- Contact hours	24 (37.5%)
	> 30 h	9 (37.5%) <sup>b</sup>
<b>Active Learning</b>	Practical experience—Inclusion of students	
	- Yes	44 (68.8%)
	- No	20 (31.3%)

<sup>a</sup> Percentages may not total 100% due to rounding

<sup>b</sup> Percentage based on a subsample of  $N=24$  studies where contact hours were reported / calculable; for the full sample ( $N=64$ ) it would be 14.1%

**Table 9** Reported change of teacher competence based on PD and direction of change. (Percentages may not total 100% due to rounding. Values of Direction of change may add up to more than 64 because some studies fit multiple categories)

Characteristic	Description	N (%)
<b>Result</b>	Change in teacher competence?	
	- Yes	54 (84.4)
	- No	10 (15.6)
	Direction of change	
	- Positive	50 (92.6)
	Positive and other	21 (42.0) <sup>a</sup>
	- Neutral	3 (5.6)
	- Negative	1 (1.9)

<sup>a</sup> Percentage based on a subsample of  $N=50$  studies where positives results were reported; for the full sample ( $N=64$ ) it would be 32.8%

When using the framework of Sampson and Clark (2008), we faced the following problem: The focal issues are not disjunctive, and we encountered multiple studies, where no clear distinction in reference to a framework was possible. The focal issues were not developed to differ between certain argumentation frameworks but rather to characterize given argumentation frameworks. This resulted in not considering them for rating; hence, no interrater reliability could be calculated.

We attempted to resolve the fact that Sampson and Clark's (2008) framework was unsuitable by using the description of argumentation competence provided by Rapanta et al., (2013). Even though we used the same assumptions that were reported in their paper, and partly their exact coding manual, our agreement was comparatively poor. Here, the missing distinguishability of the analysis and assessment categories influenced our coding. There is varying information density in terms of the argumentation reported in the studies. Most studies do not provide any or sufficient information to clearly identify the underlying concept of argumentation. We illustrate this exemplarily based on three studies in the Additional file 1: (APP1).

### Results for RO2

We categorized the PDP characteristics for the description provided, country, participants, duration via extent and contact hours, and active learning in the sense of practical experiences with students. Where applicable, the number of occurrences was counted (Table 8). A list of the countries in which the PDPs were conducted can be found in the Additional file 1: (APP2).

### General characteristics—participants

The number of participants in the PDP studies varied (from  $N=1$  to  $N=120$ ). Some studies report results for the development of one person (Christodoulou & Osborne, 2014\*; Kilinc et al., 2017\*; Zaccarelli et al., 2018\*) or—at the other extreme—there was a full cohort of teachers being taught argumentation skills (Kaya, 2013\*). Two studies had more than 100 participants, but these followed an experimental study design, dividing their participants into intervention and control groups and offering PDPs only to the (smaller) intervention group (Cinici, 2016\*; Hasnunidah et al., 2020\*). In most case studies, only the number of participant(s) in the study was reported. If this was the case, it is marked with an asterisk (\*) (e.g.,  $N=1^*$ ) behind the number in the overview, indicating that there were more participants in the PDP that have not been investigated in the study (Christodoulou & Osborne, 2014\*; Kilinc et al., 2017\*). As can also be seen from the Table 8, there are PDPs for science teachers in primary and secondary education and for in-service and pre-service teachers.

### Duration and active learning

The duration of the PDPs varied widely. PDPs lasted from 30–60 min (1 module) (Bayram-Jacobs et al., 2019\*) to multiple day-long meetings over 4 years (Chen et al., 2017\*). Most of the studies lasted between a few weeks and one year. Additionally, the number of meetings in the PDP period varied in the different studies, ranging from

one session to multiple meetings, and using synchronous and asynchronous formats.

Twenty-four of the 64 studies reported the contact hours or provided sufficient information to calculate them. Of these studies, 9 studies had an overall time of contact greater than 30 h (Baker et al., 2009\*; Choi et al., 2015\*; Cigdemoglu & Köseoğlu, 2019\*; Crippen, 2012\*; Kapon et al., 2009\*; Karisan & Topcu, 2016\*; Sarıbaş et al., 2019\*; Soysal & Yılmaz-Tuzun, 2019\*; Ünal Çoban et al., 2016\*), which was set as a cut-off criterion for the long-term benefit of PDPs (Guskey & Yoon, 2009). Finally, as shown in Table 8, approximately two-thirds of the PDPs included practical activities with students.

Summarized for the PDP characteristics, nearly in all sizes (Participants N, duration), school forms, and career stages, an argumentation PDP was conducted. However, there is still a small disparity regarding the pre-service teachers in the number of PDPs investigated. Concerning active learning, the majority of the PDPs included practice phases.

### Results for RO3

We categorized the PDP results based on a change in teacher competence and the direction of change. Where applicable, the number of occurrences was counted (Table 9). We then investigated the influence of the criteria for RO2 on the results. Finally, we provided a systematization for the reasons mentioned in the studies based on thematic analysis.

### Results of the PDPs

Most studies reported positive results ( $N=50$ ). Some of these studies reported also a neutral or negative outcome ( $N=21$ ) as the main result. Fewer studies exclusively reported a summation of neutral (Kapon et al., 2009\*; Osborne et al., 2013; Sánchez-Martín et al., 2017\*) or negative findings (Kilinc et al., 2017\*). In a few studies ( $N=10$ ), no change in teacher competence was reported; hence, the category change in teacher competence was not applicable (e.g., when the main result focused on the description of different dialogic patterns of teachers) (Anderson Quaderer & McDermott, 2018\*) or mapping science teachers' beliefs about argumentation (Katsh-Singer et al., 2016).

### Influence of the PDP criteria on the results

Because the vast majority of studies reported positive results, an identification of the influence of our investigated PDP criteria of RO2 was impeded. For example, considering duration, no influence was identifiable. In addition, looking more specifically at contact hours, nine studies exceeded the cut-off criterion of more than 30 h. On the one hand, these studies reported—as

expected—mainly positive results ( $N=5$ ). However, six of these nine studies also reported exclusively (e.g., Osborne et al., 2013) or additionally neutral results (e.g., Crippen, 2012\*) concerning their central research interests. On the other hand, studies with contact hours of less than 30 also demonstrated mostly positive effects. This result also applies to the category of active learning. Independent of including reflection or practical experiences in the PDPs, the majority of studies reported positive results. The same applied when comparing studies that had multiple reflection and practice sessions or did not.

Overall, it can be seen that, with the criteria we have chosen, no causally plausible conclusions can be drawn about the results of the training due to the large number of positive results. Based on this result, it can be argued that our chosen criteria do not influence the outcome of the reported PDP results. To fully investigate RO3, we extracted the information available from the studies and the reasons the authors attributed to their results.

### Reasons for results (by Authors)

Based on a thematic analysis of the mentioned reasons for the results in teacher competence by the study authors, we identified different reasons and counted how often they were mentioned in the 64 studies (see Fig. 3).

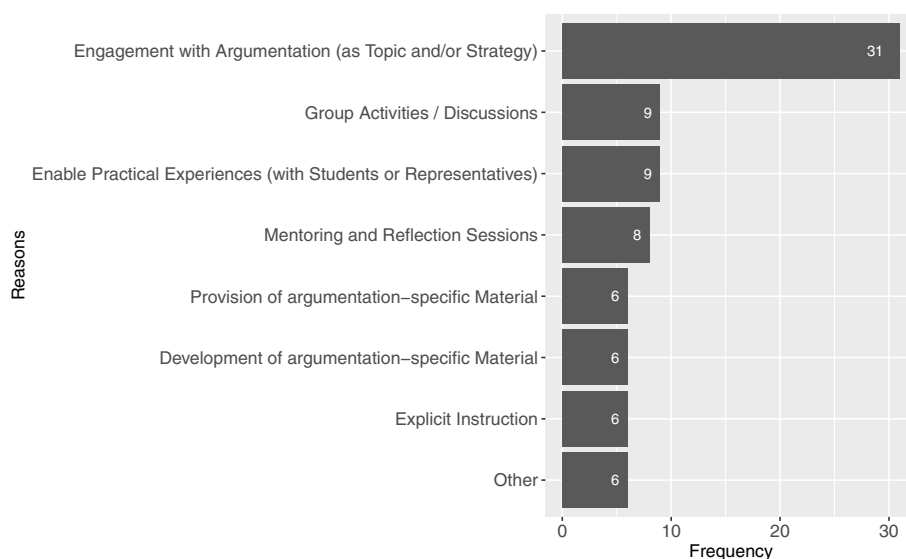
As one can see, and perhaps not surprisingly, almost half of the studies attributed engagement with argumentation as a reason for the results of their argumentation PDP. Furthermore, group activities, practical experiences, and reflection sessions in the PDP are frequently mentioned as reasons. Yet, punctually other reasons are also provided that refer to the material and instruction of the PDP or general PDP conditions.

### Discussion

Based on our results, we now interpret and contextualize them to identify patterns in our analyzed studies and provide suggestions on how research on argumentation PDPs was framed and designed. We identified three central topics that guide our discussion: 1) the variety of designs of PDPs on argumentation and the research accompanying it; 2) the visibility of underlying theoretical frameworks of PDPs on argumentation; and 3) the influence of PDP characteristics that lead to (positive) effects of PDPs.

### Variety of designs of PDPs on argumentation and accompanying research

Based on our results, research on PDPs on argumentation has built a strong foundation over the last 20 years. The diversity in the formal-structural characteristics of studies (e.g., the methodology) and in the design of PDPs (e.g., target groups and duration)—combined



**Fig. 3** Grouped reasons for results given by the study authors

with an increase of argumentation PDP studies in the last 20 years (see APP3)—indicates a general growth in the field and accompanying research. At first, this variety seems counterintuitive based on the reported struggles teachers have with implementing argumentation in their classrooms (Choi et al., 2021; McNeill & Knight, 2013; Pimentel & McNeill, 2013) and science teachers' need for support (Henderson et al., 2018). However, we argue that growing research around scientific argumentation was necessary to first identify and describe the observed problems in more detail. For example, it is now known that teachers have different opinions regarding what counts as argumentation (McNeill et al., 2016), and that their understanding of argumentation can change even during PD (Lazarou et al., 2017). The latter indicates that PDPs can help solve at least some of the problems encountered.

Looking more closely at the variety in the PDPs, it seems that there is an imbalance regarding the group of pre-service teachers. Unlike in-service teachers, pre-service teachers have specific characteristics and needs: they are not as entrenched in teaching practice and are still developing their identities as teachers (Beauchamp & Thomas, 2009; van Lankveld et al., 2017). Furthermore, they are, on the one hand, likelier to be exposed to current science education topics in their studies, but on the other hand, have less practical experience in designing and guiding learning situations in the classroom. Considering the specific characteristics and needs of this group and the differences from in-service teachers, it seems useful to take this target

group further into account when developing, providing, and investigating PDPs on argumentation.

More closely examining the variety in the research accompanying argumentation PDPs, many studies have focused on a broad range of dispositional competencies. For example, the effect of argumentation on the development of teacher content knowledge in terms of argumentation was explored by de Sá Ibraim and Justi (2016\*, 2012\*). Additionally, changes in the content knowledge of chemistry (Kaya, 2013\*) and physics topics (Kapon et al., 2009\*) using argumentation account for this. Other studies have examined beliefs about (specific parts or strategies of) argumentation (Hand et al., 2018\*; Katsh-Singer et al., 2019\*). Some studies have examined self-efficacy related to argumentation (Aydeniz & Ozdilek, 2016\*).

Additionally, observable concrete performance was investigated in-depth in 35 of the 64 studies. Here, the research focus ranged from the effect of the PDPs on concrete instruction or teacher moves (e.g., using the provision of scaffolds in the classroom) (Beland et al., 2015\*; Larraín et al., 2017\*) or examining classroom talk (Chen et al., 2017\*; Christodoulou & Osborne, 2014\*; Kim & Hand, 2015\*).

Thus, research studies have considered both dispositional characteristics and performance in argumentation training, but the combining element of situation-specific skills has hardly been explicitly taken into account as a dependent variable. Only three studies explicitly focused their research on one of the situation-specific skills (Berson et al., 2018\*; Rosaen et al., 2010\*; Tekbiyik, 2015\*). Berson et al. (2018\*) investigated the use of instructional

practices that facilitate scientific discourse and argumentation using a qualitative approach. They had 44 in-service teachers in their PDP, in which they analyzed video recordings of their own practical experiences to investigate examples of “good practice”. A similar approach was used by Rosaen et al. (2010\*) in their PDP for five pre-service teachers; they explicitly referred to the situation-specific skill of professional vision. Tekbiyik (2015\*) conducted a 4-h workshop with 90 pre-service teachers (59 in post-test) to examine the decision-making processes concerning the use of nuclear energy.

All three studies reported positive results, and as these three studies and Blömeke et al. (2015) point out, these skills seem crucial for teacher learning. In particular, the positive effects of the use of video (as in two of the reviewed studies) are in line with other research on the potential of video use in teachers’ PD (Gaudin & Chalès, 2015; Tripp & Rich, 2012) (e.g., by enabling the use of authentic classroom situations that effectively activate observers’ knowledge) (Goldman et al., 2007), and promote self-reflection abilities (Hollingsworth & Clarke, 2017) or teachers’ self-efficacy (Gröschner et al., 2018).

The description of PDPs in some studies we reviewed left the impression that activities related to situation-specific skills are also, or at least implicitly, addressed. For example, many studies have investigated a teachers’ ability to implement / facilitate / negotiate / incorporate argumentation in a classroom (Osborne et al., 2004; Saribaş et al., 2019\*; Shemwell et al., 2015\*; Simon et al., 2008\*) or a teachers’ use of a specific curriculum or strategy (Marco-Bujosa et al., 2017\*). However, situation-specific skills remain implicit, and the explicit research focus is based on performance. In particular, the extensive and well-documented PDPs (de Sá Ibrahim & Justi, 2016\*; Fishman et al., 2017\*; Osborne et al., 2004, 2013, 2019\*) that included discussion sessions or used video vignettes in their PDPs strengthen this conclusion. It appears likely that the central aspects of situation-specific skills, noticing, reasoning about specific in-situation behavior, and reflection on decision-making processes have been taken into account in these PDPs—but were not an explicit research topic.

Altogether, situation-specific skills, such as professional vision and in-situation decision-making, raise the potential to make teachers aware of the possibility of seeing and acting differently in a relevant situation based on (their own) authentic classroom situations (Sherin et al., 2011). However, as our results show, an explicit research focus on situation-specific skills, such as professional vision and decision-making around science teachers’ PDPs on argumentation, is underrepresented. Thus, situation-specific skills (i.e., noticing and reasoning about argumentation in science education or—as a first step—recognizing

and thinking about relevant situations in which argumentation takes place or can take place in the science classroom) represent a promising but still insufficiently considered branch of research. We argue that research on this would help teachers transfer the skills learned via PDPs into their classrooms and enhance science teachers’ repertoires to facilitate argumentation—a problem that research is still facing (Henderson et al., 2018).

Therefore, we conclude: *There is diversity in the characteristics of PDPs on argumentation for science teachers and in addressed research topics in terms of formal-structural aspects, but a) more attention should be paid to the PD of pre-service teachers, and b) there is a need for explicit, systematic research on situation-specific skills in argumentation PDPs.*

### The visibility of underlying theoretical frameworks of PDPs on argumentation

We initially assumed that the identification of focal issues based on the work of Sampson and Clark (2008) and the allocation of the activities to the argumentation competencies based on Rapanta et al.’s (2013) work would allow us to draw conclusions on this topic. It was pointed out that a description of the underlying argumentation framework is important to allow a comparison of the different study results (Sampson & Clark 2008). However, we were not able to accomplish this task due to the challenges in reviewing the studies. Our struggles with the argumentation frameworks are based on multiple reasons, as outlined earlier in the Results section of this paper. Summarized, our challenges confirm the need for better documentation:

*Another important message, related to the first, underscores how much information readers need to interpret the results of a study; it is simply not enough to say that a given intervention supports students in creating “high”-or “low”-quality arguments. An audience needs very specific details about the nature of the analytic foci as well the underlying assumptions about “what counts” as quality to interpret findings. Explicit sharing of these details among researchers will improve communication and comparison of results across studies. (Sampson & Clark, 2008, p. 469)*

This conclusion regarding the need for better reporting after a review of the assessment of student argumentation is also applicable to the development of teacher PDP on argumentation. Perhaps the occurrence of only seven quantitative studies in our review of the research focus also relates to this. Presuming that quantitative measures need a valid and reliable underlying operationalization of the construct/variable of interest, perhaps quantitative

approaches and the prerequisites for it (e.g., instrument development and validation) need more attention, especially when one person looks at our struggles to describe the PDPs underlying argumentation framework. A first step could be a description and classification of the argumentation concept in the seven identified quantitative studies for science teacher PDPs on argumentation (Fishman et al., 2017\*; Hasnunidah et al., 2020\*; Kaya, 2013\*; Murphy et al., 2018\*; Osborne et al., 2013, 2019\*; Sánchez-Martín et al., 2017\*).

Thus, we recommend that research papers include a detailed description of the underlying argumentation framework in a holistic, clear, and systematic way to enable comparisons.

#### **PDP characteristics that lead to positive effects of PDPs**

We were able to show that there is broad variety in PDPs regarding characteristics, such as the participants or the duration. Further, practice experiences were prominent in the majority of the PDPs on argumentation. We were pleasantly surprised by the large number of PDPs enabling active learning in terms of the integration of practical parts in science teachers' PD of argumentation. However, we cannot infer any causal relationship between the contribution of these investigated categories of PDPs and the reported results.

For example, although duration is a considerable factor influencing the effectivity of PDPs (Borko et al., 2010; Desimone, 2009), viewing it as a content-independent criterion that causally links to the results has its fallacies. Too wide is the range of different PDPs in terms of duration and number of meetings. Even the more specialized measure of contact hours, used as a standalone criterion to provide effective, sustainable PDPs, is too short-sighted. First, of the 64 studies, only 24 reported or provided enough information to calculate contact hours, so the information was mostly unavailable from the researched studies. Only 9 of the 24 studies had an overall time of contact greater than 30 h, which has been set as a cut-off criterion for long-term benefits of PDPs (Guskey & Yoon, 2009). Still, of the nine studies identified, not all reported a positive effect on teacher learning in argumentation, despite the occurrence of many positive results. Therefore, studies not fulfilling this criterion also had positive effects.

Osborne et al., (2013) and Kilinc et al. (2017\*), or the results of the training may remain below expectations due to insufficient supervision (Osborne et al., 2013). The experiences shared in these two studies about PDP on argumentation are in line with other literature regarding teacher PD: The effectiveness of learning processes through practical experiences depends strongly on a) the mentoring before, during, and after the internship

(Gergen, 2019) as well as, not only the possibility for but also the quality of reflection of one's own teaching–learning sequences (Herzog & von Felten, 2001). Therefore, using practical experience or duration as a single measure for effective teacher PD is not useful.

Based on this, we can confirm that the results of the PDPs reported in the studies were not automatically influenced by the sole availability of influential PDP criteria, such as long duration and active, yet unreflected or unevaluated learning. Our chosen resolution of duration, especially contact hours, and practical experience, is suitable to show if an argumentation PDP had one of these components at all. They do not discern the quality of the PDP session(s) or the practical experience(s). More sensitive operationalizations of the sessions (e.g., analyzing tasks, material, and instruction), or the practical experiences (e.g., investigation of the preparation, mentoring, or concrete tasks) enable a more in-depth analysis of the quality of the PDP.

Yet, we want to critically note that this is partially impeded by two circumstances. First, as shown for the contact hours, at least for our 64 selected studies, there is a varying information density or absence of relevant information in some studies. Second, the sheer number of positive results of PDPs for argumentation still impedes a comparison to determine good practice indicators. Concerning the second aspect, we cannot exclude the prevalence of p-bias based on the large number of positive results. Since one of our inclusion criteria was “peer review,” we have no indication of studies on argumentation PDPs that may have been rejected for publishing due to reporting null or negative results. This imposes the question of a prevailing publication bias in peer-reviewed studies, as has been recognized in other fields of academic research—for example, for social sciences, see Franco et al. (2014).

Besides illustrating that these “hard” PDP characteristics do not allow any causal conclusions of the results of a PDP, we have shown with our categorization of the reasons the study authors attribute that there are many variables to improve argumentation. We can also see an overlap of our reasons extracted with other effective criteria in general—for example, practical activities and reflection phases as mentioned by Desimone (2009)—and beneficial argumentation-specific design principles—for example, explicit instruction and introduction of argumentation is also mentioned by Aydeniz (2019). However, the reasons themselves remain vague. One could argue here that our generalization of the reasons for RO3 is too low to resolve. However, with the varying amount of information given in the studies and our intent to cover *all* studies that all make argumentation an explicit goal of their studies, this was our best attempt to group



these diverse studies that all aim for enhancing argumentation using different contexts and aims with this resolution level, and still reach a considerable consensus. To obtain finer-grained results, we refer here back to our conclusion for the argumentation framework and cite a) a need for better reporting, and b) a more detailed sampling based on our findings. Perhaps one could further investigate the subsample of the 31 studies on “engaging with argumentation,” but then still, and we highlight this, even if the single activities were more precisely described, their impact on the result as one of many PDP activities remains unknown.

Overall, we conclude *that the identification of effective PDP characteristics is possible, but generalizations based on surface PDP characteristics are limited.* For example, practical phases can be the reasons for positive results, yet generally speaking, practical phases do not automatically lead to good results. This also applies to the other investigated categories for RO2 and RO3. Therefore, *future research should build on the solid fundament of previous argumentation PDPs to develop and investigate the latent trait of argumentation competence. Future research should more clearly relate the resulting different PDP activities used to develop science teachers’ argumentation skills to the actual PDP result at the same inference level and further investigate the share each of these activities has on the result.*

### Limitations

One may criticize the fact that we used Blömeke et al.’s (2015) model for studies published before 2015 and therefore have an indiscriminate use of the categories regarding professional competence, especially situation-specific skills. We rebut this critique because subordinated concepts, such as noticing (van Es & Sherin, 2002a, 2002b) or professional vision (Goodwin, 1994), were established long before the application in the model used here. Furthermore, we highlighted in the discussion the implicit presence of these components in PDPs, yet an explicit research focus in the related studies—and when looking for established concepts such as noticing and professional vision—still remains unattended.

We reviewed a subsample of the papers using more than one coder; the rest of the analysis was done by only one person. Even though this procedure seems legitimate (Hallgren, 2012), one could challenge the generalizability of this step. However, we did not only opt for a subsample because of time and efficiency reasons. Rather, we see our high agreement in the values of a randomized subsample as evidence of the functioning of our coding manual and ratings (except for the argumentation categories). Therefore, we assume that additional ratings with two

coders would not necessarily have resulted in a significant change.

Furthermore, our selection of included studies may be biased by country. Most studies were found in Western, English-speaking countries (see APP2). This may not be surprising, considering that we filtered for studies in English since it is one of the standard languages of science communication worldwide. However, educational contexts are often bound to national curricula. Therefore, this review possibly misses out on studies and PDPs that are written in a language other than English but otherwise would have suited the criteria.

Additionally, the studies included in the review span the entire K-12 curriculum (elementary, middle and high school). In our review we followed the approach to give a general overview of how the development of skill in the facilitation of argumentation has been realized, independent of the curriculum. Future research on this issue or on the type of school (urban, rural, suburban) can use our here gained insights to examine these objectives more closely.

Addressing the search terms for this review, one could argue that, for example, the imbalance in the group of pre-service teachers is grounded in the search terms, leading to more results for PDP and in-service teachers and not for pre-service teachers, where “continuous teacher development” is a more prominent term. However, since we explicitly included “pre-service teachers” in the search terms, as well as “teacher development” and “teacher training” as varieties of PD, we feel that we still have a representative sample for pre-service teachers. Hence, the found imbalance represents the existent number of studies and is not based on faulty or incomplete search terms.

Also, we want to share a critical thought on our approach to identifying the argumentation framework for a PDP on argumentation. We tried to describe the underlying argumentation framework of the studies considering the argumentation PDP based on the descriptions provided in the studies. We did this since we anticipated varying depths in the description of argumentation in the PDPs, yet we did not systematically take into account some studies that reported concrete argumentation instruction and the material of the PD. Perhaps a combined analysis of the argumentation framework, argumentation activities, tasks, and material of the PDP using the categorizations of Rapanta et al. (2013) and an examination of the relationship between the underlying theoretical argumentation framework and the argument instruction and material used in PD will lead to more satisfying results, as only a subsample of the 64 studies provided this information. Here, the teacher actions of Weiss et al. (2022), the design principles of Aydeniz (2019), the work of Zohar

(2007), and Zembal-Saul and Vaishampayan (2019), which teachers need to know to successfully introduce argumentation into the classroom, can be useful sources. Furthermore, the additional application of other classifications in a combined approach—for example, that of Erduran et al. (2015)—could deliver more holistic results for reviewing the underlying argumentation framework. However, this cannot be done by us at this point, as this is a whole new, time-intensive research task.

Finally, we reviewed only papers that considered science education. General approaches to argumentation PDPs in other fields of education were not considered. However, our presented approach could function as a template, and the results could provide valuable implications for conducting argumentation PDP reviews in other relevant fields.

## Conclusions

The need for this review arose based on teachers' challenges in facilitating scientific argumentation in classrooms, even though PDPs on argumentation exist. Based on our review of these PDPs in terms of selected PD characteristics and research focus, and the discussion and experience that occurred during the review process of the argumentation categories, we came to three central conclusions regarding the diversity of research of PDPs on argumentation, its accompanying research, the visibility of underlying theoretical frameworks of PDPs on argumentation, and the characteristics that lead to positive effects of PDPs. With these recommendations, we hope to provide helpful insight for developers, moderators, and researchers in designing, implementing, and researching PDPs on argumentation in science education.

## Abbreviations

ABI	Argument-Based Inquiry
ADI	Argument-Driven Inquiry
CER	Claim Evidence Reasoning framework
ERA	Educational Research Abstracts Online
ERIC	Education Resources Information Center
K	Cohens Kappa
N	Sample size
PCK	Pedagogical Content Knowledge
PD	Professional Development
PDP	Professional Development Program
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
SWH	Science Writing Heuristic
TAP	Toulmins Argument Pattern

## Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s43031-023-00076-3>.

**Additional file 1: APP1.** Example coding of the categorization of argumentation competence. **APP2.** Overview of countries where PDPs were enacted.

## Acknowledgements

Not applicable.

## Approval from authors

All authors confirm the approval of the manuscript for submission.

## Publishing of manuscript

All authors confirm, that the content of the manuscript has not been published, or submitted for publication elsewhere.

## Explanation of any issues relating to journal policies

The authors declare that they have no issues to report relating to the journal policies.

## Authors' contributions

RW conducted the database search, analyzed and interpreted the data and was a major contributor in writing the manuscript. RW and BP rated both studies with the given categories. BP and IP contributed by consulting on all sections of the manuscript, also revising and proof-reading it. All authors read and approved the final manuscript.

## Authors' information

Raphael Wess is a doctoral candidate in science education specializing in chemistry and physics education. He works as a research associate in the physics education department at the Humboldt-Universität zu Berlin and is a doctoral candidate in the chemistry education department of the IPN Kiel. His research focuses on the description and development of argumentation skills of science teachers. He holds a master degree in secondary education from the Humboldt-Universität zu Berlin.

Burkhard Priemer is professor for physics education at the Humboldt-Universität zu Berlin. His research focuses on the use of data, measurement uncertainties, and argument development in physics classroom.

Ilka Parchmann is professor for chemistry education at the Leibniz Institute for Science and Mathematics Education in Kiel. Her research focuses, inter alia, on concept development and investigation for the (further) qualification of teachers.

## Funding

This research received no external funding.

## Availability of data and materials

The data that support the findings of this study are openly available in the Open-Access-Publication-Server of the Humboldt-Universität zu Berlin (edoc) at <https://doi.org/10.18452/24795>.

## Declarations

### Ethics approval and consent to participate

Not applicable.

### Consent for publication

Not applicable.

### Competing interests

The authors declare that they have no competing interests.

Received: 26 January 2023 Accepted: 29 May 2023

Published online: 20 June 2023

## References

### Research study data

Wess, R. (2023). *Research Data for the Systematic Review "Professional Development Programs to Improve Science Teachers' Argumentation Skills" (Version 1) [Data set and code book]*. Humboldt-Universität zu Berlin (edoc). <https://doi.org/10.18452/24795>

## Literature

- Asterhan, C. S. C., & Schwarz, B. B. (2016). Argumentation for learning: Well-trodden paths and unexplored territories. *Educational Psychologist*, *51*(2), 164–187. <https://doi.org/10.1080/00461520.2016.1155458>
- Aydeniz, M. (2019). Teaching and learning chemistry through argumentation. In S. Erduran (Ed.), *Argumentation in Chemistry Education* (pp. 11–31). Royal Society of Chemistry.
- Bağ, H., & Çalık, M. (2017). A thematic review of argumentation studies at the K-8 Level. *Eğitim ve Bilim-Education and Science*, *42*(190), 1–23. <https://doi.org/10.15390/EB.2017.6845>
- Berland, L. K., & Lee, V. R. (2012). In Pursuit of Consensus: Disagreement and legitimization during small-group argumentation. *International Journal of Science Education*, *34*(12), 1857–1882. <https://doi.org/10.1080/09500693.2011.645086>
- Beauchamp, C., & Thomas, L. (2009). Understanding teacher identity: An overview of issues in the literature and implications for teacher education. *Cambridge Journal of Education*, *39*(2), 175–189. <https://doi.org/10.1080/03057640902902252>
- Blömeke, S., Gustafsson, J.-E., & Shavelson, R. J. (2015). Beyond Dichotomies: Competence Viewed as a Continuum. *Zeitschrift Für Psychologie*, *223*(1), 3–13. <https://doi.org/10.1027/2151-2604/a000194>
- Borko, H., Jacobs, J., & Koellner, K. (2010). Contemporary Approaches to Teacher Professional Development. In Peterson et al. (Ed.), *International Encyclopedia of Education (3rd ed)* (pp. 548–556). Elsevier. <https://doi.org/10.1016/B978-0-08-044894-7.00654-0>
- Cavagnetto, A. R. (2010). Argument to foster scientific literacy: A review of argument interventions in K–12 science contexts. *Review of Educational Research*, *80*(3), 336–371. <https://doi.org/10.3102/0034654310376953>
- Choi, A., Seung, E., & Kim, D. (2021). Science teachers' views of argument in scientific inquiry and argument-based science instruction. *Research in Science Education*, *51*, 251–268. <https://doi.org/10.1007/s11165-019-9861-9>
- Desimone, L. M. (2009). Improving impact studies of teachers' professional development: Toward better conceptualizations and measures. *Educational Researcher*, *38*(3), 181–199. <https://doi.org/10.3102/0013189X08331140>
- 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%3c287::AID-SC1163E3.0.CO;2-A](https://doi.org/10.1002/(SICI)1098-237X(200005)84:3%3c287::AID-SC1163E3.0.CO;2-A)
- Erduran, S., Ozdem, Y., & Park, J.-Y. (2015). Research trends on argumentation in science education: A journal content analysis from 1998–2014. *International Journal of STEM Education*, *2*(1), 1–12. <https://doi.org/10.1186/s40594-015-0020-1>
- 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>
- Franco, A., Malhotra, N., & Simonovits, G. (2014). Publication bias in the social sciences: Unlocking the file drawer. *Science*, *345*(6203), 1502–1505. <https://doi.org/10.1126/science.1255484>
- Gaudin, C., & Chaliès, S. (2015). Video viewing in teacher education and professional development: A literature review. *Educational Research Review*, *16*, 41–67. <https://doi.org/10.1016/j.edurev.2015.06.001>
- Gergen, A. (2019). Mentoring in schulpraktischen Phasen der Lehrerbildung. Zusammenfassung ausgewählter Forschungsbeiträge zur Mentorentätigkeit. In Degeling, M., Franken, N., Freund, S., Greiten, S., Neuhaus, D., Schellenbach-Zell, J. (Ed.), *Herausforderung Kohärenz: Praxisphasen in der universitären Lehrerbildung. Bildungswissenschaftliche und fachdidaktische Perspektiven* (pp. 329–339). Verlag Julius Klinkhardt. <https://doi.org/10.25656/01:17285>
- Goldman, R., Pea, R., Barron, B., & Derry, S. J. (2007). *Video research in the learning sciences*. Mahwah. Erlbaum.
- González-Howard, M., & McNeill, K. L. (2019). Teachers' framing of argumentation goals: Working together to develop individual versus communal understanding. *Journal of Research in Science Teaching*, *56*(6), 821–844. <https://doi.org/10.1002/tea.21530>
- Goodwin, C. (1994). Professional vision. *American Anthropologist*, *96*(3), 606–633.
- Gröschner, A., Schindler, A.-K., Holzberger, D., Alles, M., & Seidel, T. (2018). How systematic video reflection in teacher professional development regarding classroom discourse contributes to teacher and student self-efficacy. *International Journal of Educational Research*, *90*, 223–233. <https://doi.org/10.1016/j.ijer.2018.02.003>
- Guskey, T. R., & Yoon, K. S. (2009). What works in professional development? *Phi Delta Kappan*, *90*(7), 495–500. <https://doi.org/10.1177/003172170909000709>
- Gwet, K. L. (2012). *Handbook of Inter-Rater Reliability. The definitive guide to measuring the extent of agreement among multiple raters* (3rd ed.). Advanced Analytics LLC.
- Hallgren, K. A. (2012). Computing inter-rater reliability for observational data: An overview and tutorial. *Tutorials in Quantitative Methods for Psychology*, *8*(1), 23–34.
- Hazelkorn, E., Ryan, C., Beernaert, Y., Constantinou, C. P., Deca, L., Grangeat, M., Karikorp, M., Lazoudis, A., Casulleras, R. P., Welzel, M., Europäische Kommission, & Europäische Kommission. (2015). *Science education for responsible citizenship: report to the European Commission of the Expert Group on Science Education*. Publications Office of the European Union.
- 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. <https://doi.org/10.1002/tea.21412>
- Herzog, W., & von Felten, R. (2001). Erfahrung und Reflexion. Zur Professionalisierung der Praktikumsausbildung von Lehrerinnen und Lehrern. *Beiträge zur Lehrerbildung*, *19*(1), 17–28. <https://doi.org/10.25656/01:13447>
- Hollingsworth, H., & Clarke, D. (2017). Video as a tool for focusing teacher self-reflection: Supporting and provoking teacher learning. *Journal of Mathematics Teacher Education*, *20*(5), 457–475. <https://doi.org/10.1007/s10857-017-9380-4>
- Jiménez-Aleixandre, M. P., & Erduran, S. (2007). Argumentation in Science Education: An Overview. In M. P. Jiménez-Aleixandre & S. Erduran (Ed.), *Argumentation in Science Education. Perspectives from Classroom-Based Research* (pp. 3–29). Springer.
- 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>
- Kultusministerkonferenz (2020a). *Bildungsstandards im Fach Chemie für die Allgemeine Hochschulreife*. Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland (Ed.), [https://www.kmk.org/fileadmin/Dateien/veroeffentlichungen\\_beschluesse/2020a/2020a\\_06\\_18-BildungsstandardsAHR\\_Chemie.pdf](https://www.kmk.org/fileadmin/Dateien/veroeffentlichungen_beschluesse/2020a/2020a_06_18-BildungsstandardsAHR_Chemie.pdf). Accessed 28 June 2022
- Kultusministerkonferenz (2020b). *Bildungsstandards im Fach Physik für die Allgemeine Hochschulreife*. Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland (Ed.), [https://www.kmk.org/fileadmin/Dateien/veroeffentlichungen\\_beschluesse/2020b/2020b\\_06\\_18-BildungsstandardsAHR\\_Physik.pdf](https://www.kmk.org/fileadmin/Dateien/veroeffentlichungen_beschluesse/2020b/2020b_06_18-BildungsstandardsAHR_Physik.pdf). Accessed 28 June 2022
- Kunter, M., Klusmann, U., Baumert, J., Richter, D., Voss, T., & Hachfeld, A. (2013). Professional competence of teachers: Effects on instructional quality and student development. *Journal of Educational Psychology*, *105*(3), 805–820. <https://doi.org/10.1037/a0032583>
- Lazarou, D., Erduran, S., & Sutherland, R. (2017). Argumentation in science education as an evolving concept: Following the object of activity. *Learning, Culture and Social Interaction*, *14*, 51–66. <https://doi.org/10.1016/j.lcsi.2017.05.003>
- Lee, C. J., Sugimoto, C. R., Zhang, G., & Cronin, B. (2013). Bias in peer review. *Journal of the American Society for Information Science and Technology*, *64*(1), 2–17.
- 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., & 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. S. (2011). *Supporting Grade 5–8 Students in Constructing Explanations in Science: The Claim, Evidence, and Reasoning Framework for Talk and Writing*. Pearson.
- McNeill, K. L., & Pimentel, D. S. (2010). Scientific discourse in three urban classrooms: The role of the teacher in engaging high school students in argumentation. *Science Education*, *94*(2), 203–229. <https://doi.org/10.1002/sce.20364>

- Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., PRISMA Group\*. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *Annals of Internal Medicine*, 151(4), 264–269.
- Moher, D., Shamseer, L., Clarke, M., Ghersi, D., Liberati, A., Petticrew, M., Shekelle, P., Stewart, L. A., PRISMA-P Group. (2015). Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Systematic Reviews*, 4(1), 1. <https://doi.org/10.1186/2046-4053-4-1>
- Newman, M., & Gough, D. (2020). Systematic Reviews in Educational Research: Methodology, Perspectives and Application. In O. Zawacki-Richter, M. Kerres, S. Bedenlier, M. Bond, & K. Buntins (Eds.), *Systematic Reviews in Educational Research: Methodology, Perspectives and Application* (pp. 3–22). Springer Fachmedien. [https://doi.org/10.1007/978-3-658-27602-7\\_1](https://doi.org/10.1007/978-3-658-27602-7_1)
- National Research Council (2012). *Discipline-based education research: Understanding and improving learning in undergraduate science and education*. National Academies Press.
- OECD, Rychen, D. S., & Salganik, L. H. (2002). *Definition and Selection of Competencies (DESECO): Theoretical and Conceptual Foundations. Strategy paper*. Swiss Federal Statistical Office. <http://hdl.voced.edu.au/10707/156754> (= OECD, 2002)
- Opfer, V. D., & Pedder, D. (2011). Conceptualizing teacher professional learning. *Review of Educational Research*, 81(3), 376–407. <https://doi.org/10.3102/0034654311413609>
- Osborne, J., Erduran, S., & Simon, S. (2004). Enhancing the quality of argumentation in school science. *Journal of Research in Science Teaching*, 41(10), 994–1020. <https://doi.org/10.1002/tea.20035>
- 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., Simon, S., Christodoulou, A., Howell-Richardson, C., & Richardson, K. (2013). Learning to argue: A study of four schools and their attempt to develop the use of argumentation as a common instructional practice and its impact on students. *Journal of Research in Science Teaching*, 50(3), 315–347. <https://doi.org/10.1002/tea.21073>
- Pimentel, D. S., & McNeill, K. L. (2013). Conducting talk in secondary science classrooms: Investigating instructional moves and teachers' beliefs. *Science Education*, 97(3), 367–394.
- 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>
- 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>
- Schneider, K. (2019). What does competence mean? *Psychology*, 10(14), 1938–1958. <https://doi.org/10.4236/psych.2019.1014125>
- Seidel, T., & Stürmer, K. (2014). Modeling and measuring the structure of professional vision in preservice teachers. *American Educational Research Journal*, 51(4), 739–771. <https://doi.org/10.3102/0002831214531321>
- Sherin, M., Jacobs, V., & Philipp, R. (2011). Situation awareness in teaching: What educators can learn from video-based research in other fields. In M. Sherin, V. Jacobs, & R. Philipp (Eds.), *Mathematics teacher noticing* (pp. 81–95). Routledge.
- Tripp, T. R., & Rich, P. J. (2012). The influence of video analysis on the process of teacher change. *Teaching and Teacher Education*, 28(5), 728–739. <https://doi.org/10.1016/j.tate.2012.01.011>
- van Driel, J. H., Meirink, J. A., van Veen, K., & Zwart, R. C. (2012). Current trends and missing links in studies on teacher professional development in science education: A review of design features and quality of research. *Studies in Science Education*, 48(2), 129–160. <https://doi.org/10.1080/03057267.2012.738020>
- Van Es, E. A., & Sherin, M. G. (2002b). Learning to notice: Scaffolding new teachers' interpretations of classroom interactions. *Journal of Technology and Teacher Education*, 10(4), 571–596.
- van Es, E. A., & Sherin, M. G. (2002a). Learning to notice: Scaffolding New teachers' interpretations of classroom interactions. *Journal of Technology and Teacher Education*, 10(4), 571–596. <https://www.learntechlib.org/prima/ry/p/9171/>
- van Lankveld, T., Schoonenboom, J., Volman, M., Croiset, G., & Beishuizen, J. (2017). Developing a teacher identity in the university context: A systematic review of the literature. *Higher Education Research & Development*, 36(2), 325–342. <https://doi.org/10.1080/07294360.2016.1208154>
- von Aufschnaiter, C., Erduran, S., Osborne, J., & Simon, S. (2008). Arguing to learn and learning to argue: Case studies of how students' argumentation relates to their scientific knowledge. *Journal of Research in Science Teaching*, 45(1), 101–131. <https://doi.org/10.1002/tea.20213>
- Weiss, K. A., McDermott, M. A., & Hand, B. (2022). Characterising immersive argument-based inquiry learning environments in school-based education: A systematic literature review. *Studies in Science Education*, 58(1), 15–47. <https://doi.org/10.1080/03057267.2021.1897931>
- Zemal-Saul, C., & Vaishampayan, A. (2019). Research and practice on science teachers' continuous professional development in argumentation. In S. Erduran (Ed.), *Argumentation in Chemistry Education* (pp. 142–172). Royal Society of Chemistry.
- Zohar, A. (2007). Science Teacher Education and Professional Development in Argumentation. In S. Erduran & M. P. Jiménez-Aleixandre (Eds.), *Argumentation in Science Education: Perspectives from Classroom-Based Research* (pp. 245–268). Springer Netherlands. [https://doi.org/10.1007/978-1-4020-6670-2\\_12](https://doi.org/10.1007/978-1-4020-6670-2_12)

#### Literature used in review (\* indicates reviewed study)

- \*Anderson Quarderer, N., & McDermott, M. A. (2018). Examining science teacher reflections on argument-based inquiry through a critical discourse lens. *Research in Science Education*. <https://doi.org/10.1007/s11165-018-9790-z>
- \*Aydeniz, M., & Ozdilek, Z. (2016). Assessing and enhancing pre-service science teachers' self-efficacy to teach science through argumentation: challenges and possible solutions. *International Journal of Science and Mathematics Education*, 14(7), 1255–1273. <https://doi.org/10.1007/s10763-015-9649-y>
- \*Bagdonas, A., & Silva, C. C. (2015). Enhancing teachers' awareness about relations between science and religion. *Science & Education*, 24(9), 1173–1199. <https://doi.org/10.1007/s11191-015-9781-7>
- \*Baker, D. R., Lewis, E. B., Purzer, S., Watts, N. B., Perkins, G., Uysal, S., Wong, S., Beard, R., & Lang, M. (2009). The Communication in Science Inquiry Project (CISIP): A Project to Enhance Scientific Literacy through the Creation of Science Classroom Discourse Communities. *Faculty Publications: Department of Teaching, Learning and Teacher Education*, 119. <https://digitalcommons.unl.edu/teachlearnfacpub/119>
- \*Bayram-Jacobs, D., Henze, I., Evagorou, M., Shwartz, Y., Aschim, E. L., Alcaraz-Dominguez, S., Barajas, M., & Dagan, E. (2019). Science teachers' pedagogical content knowledge development during enactment of socioscientific curriculum materials. *Journal of Research in Science Teaching*, 56(9), 1207–1233. <https://doi.org/10.1002/tea.21550>
- \*Belland, B. R., Burdo, R., & Gu, J. (2015). A blended professional development program to help a teacher learn to provide one-to-one scaffolding. *Journal of Science Teacher Education*, 26(3), 263–289. <https://doi.org/10.1007/s10972-015-9419-2>
- \*Berson, E., Borko, H., Million, S., Khachatryan, E., & Glennon, K. (2018). Practice what you teach: a video-based practicum model of professional development for elementary science teachers. *Orbis Scholae*, 9(2), 35–53. <https://doi.org/10.14712/23363177.2015.79>
- \*Braund, M., Hewson, P. W., Scholtz, Z., Sadeck, M., & Koopman, R. (2011). Learning to teach argumentation: Facilitated reflection on a pre-service curriculum in South Africa. *Education as Change*, 15(1), 79–93. <https://doi.org/10.1080/16823206.2011.643630>
- \*Cavlazoglu, B., & Stuessy, C. (2018). Examining science teachers' argumentation in a teacher workshop on earthquake engineering. *Journal of Science Education and Technology*, 27(4), 348–361. <https://doi.org/10.1007/s10956-018-9728-2>
- \*Chen, Y.-C., Hand, B., & Norton-Meier, L. (2017). Teacher roles of questioning in early elementary science classrooms: a framework promoting student cognitive complexities in argumentation. *Research in Science Education*, 47(2), 373–405. <https://doi.org/10.1007/s11165-015-9506-6>

- \*Choi, A., Klein, V., & Hershberger, S. (2015). Success, difficulty, and instructional strategy to enact an argument-based inquiry approach: experiences of elementary teachers. *International Journal of Science and Mathematics Education*, 13(5), 991–1011. <https://doi.org/10.1007/s10763-014-9525-1>
- \*Christodoulou, A., & Osborne, J. (2014). The science classroom as a site of epistemic talk: A case study of a teacher's attempts to teach science based on argument. *Journal of Research in Science Teaching*, 51(10), 1275–1300. <https://doi.org/10.1002/tea.21166>
- \*Cigdemoglu, C., & Köseoglu, F. (2019). Improving science teachers' views about scientific inquiry. *Science & Education*, 28(3), 439–469. <https://doi.org/10.1007/s11191-019-00054-0>
- \*Cinicì, A. (2016). Balancing the pros and cons of GMOs: Socio-scientific argumentation in pre-service teacher education. *International Journal of Science Education*, 38(11), 1841–1866. <https://doi.org/10.1080/09500693.2016.1220033>
- \*Crippen, K. J. (2012). Argument as professional development: impacting teacher knowledge and beliefs about science. *Journal of Science Teacher Education*, 23(8), 847–866. <https://doi.org/10.1007/s10972-012-9282-3>
- \*de Sá Ibrahim, S., & Justi, R. (2016). Teachers' knowledge in argumentation: Contributions from an explicit teaching in an initial teacher education programme. *International Journal of Science Education*, 38(12), 1996–2025. <https://doi.org/10.1080/09500693.2016.1221546>
- \*Erduran, S., Ardac, D., & Yakmaci-Guzel, B. (2006). Learning to teach argumentation: case studies of pre-service secondary science teachers. *EURASIA Journal of Mathematics, Science and Technology Education*, 2(2), 1–4. <https://doi.org/10.12973/ejms/75442>
- \*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>
- \*Hand, B., Park, S., & Suh, J. K. (2018). Examining teachers' shifting epistemic orientations in improving students' scientific literacy through adoption of the science writing heuristic approach. In K.-S. Tang & K. Danielsson (Eds.), *Global Developments in Literacy Research for Science Education* (pp. 339–355). Springer International Publishing. [https://doi.org/10.1007/978-3-319-69197-8\\_20](https://doi.org/10.1007/978-3-319-69197-8_20)
- \*Hasnunidah, N., Susilo, H., Irawati, M., & Suwono, H. (2020). The contribution of argumentation and critical thinking skills on students' concept understanding in different learning models. *Journal of University Teaching & Learning Practice*, 17(1), 1–11. <https://ro.uow.edu.au/jutlp/vol17/iss1/6>
- \*Jordan, R. C., Brooks, W. R., DeLisi, J. R., Gray, S. A., & Berkowitz, A. R. (2015). Ecology nature of science: shared discussions and practices among ecologists and high school teachers. *Ecosphere*, 6(11), art223. <https://doi.org/10.1890/ES13-00386.1>
- \*Justi, R., & Mendonça, P. C. C. (2016). Discussion of the controversy concerning a historical event among pre-service teachers. *Science & Education*, 25(7), 795–822. <https://doi.org/10.1007/s11191-016-9846-2>
- \*Kang, E. J. S., McCarthy, M. J., & Donovan, C. (2019). Elementary teachers' enactment of the ngss science and engineering practices. *Journal of Science Teacher Education*, 30(7), 788–814. <https://doi.org/10.1080/1046560X.2019.1630794>
- \*Kapon, S., Ganiel, U., & Eylon, B. (2009). Scientific argumentation in public physics lectures: Bringing contemporary physics into high-school teaching. *Physics Education*, 44(1), 33–38. <https://doi.org/10.1088/0031-9120/44/1/004>
- \*Karisan, D., & Topcu, M. S. (2016). Contents exploring the preservice science teachers' written argumentation skills: the global climate change issue. *International Journal of Environmental & Science Education*, 11(6), 1347–1363.
- \*Kaya, E. (2013). Argumentation practices in classroom: pre-service teachers' conceptual understanding of chemical equilibrium. *International Journal of Science Education*, 35(7), 1139–1158. <https://doi.org/10.1080/09500693.2013.770935>
- \*Kilinc, A., Demiral, U., & Kartal, T. (2017). Resistance to dialogic discourse in SSI teaching: The effects of an argumentation-based workshop, teaching practice, and induction on a preservice science teacher. *Journal of Research in Science Teaching*, 54(6), 764–789. <https://doi.org/10.1002/tea.21385>
- \*Kim, S., & Hand, B. (2015). An analysis of argumentation discourse patterns in elementary teachers' science classroom discussions. *Journal of Science Teacher Education*, 26(3), 221–236. <https://doi.org/10.1007/s10972-014-9416-x>
- \*Knight-Bardsley, A., & McNeill, K. L. (2016). Teachers' pedagogical design capacity for scientific argumentation. *Science Education*, 100(4), 645–672. <https://doi.org/10.1002/sce.21222>
- \*Kulatunga, U., & Lewis, J. E. (2013). Exploration of peer leader verbal behaviors as they intervene with small groups in college general chemistry. *Chemistry Education Research and Practice*, 14(4), 576–588. <https://doi.org/10.1039/C3RP00081H>
- \*Larraín, A., Moreno, C., Grau, V., Freire, P., Salvat, I., López, P., & Silva, M. (2017). Curriculum materials support teachers in the promotion of argumentation in science teaching: A case study. *Teaching and Teacher Education*, 67, 522–537. <https://doi.org/10.1016/j.tate.2017.07.018>
- \*Lehesvuori, S., Hähkiöniemi, M., Jokiranta, K., Nieminen, P., Hiltunen, J., & Viiri, J. (2017). Enhancing dialogic argumentation in mathematics and science. *Studia Paedagogica*, 22(4), 55–76. <https://doi.org/10.5817/SP2017-4-4>
- \*Liu, S., & Roehrig, G. (2019). Exploring science teachers' argumentation and personal epistemology about global climate change. *Research in Science Education*, 49(1), 173–189. <https://doi.org/10.1007/s11165-017-9617-3>
- \*Loper, S., McNeill, K. L., González-Howard, M., Marco-Bujosa, L. M., & O'Dwyer, L. M. (2019). The impact of multimedia educative curriculum materials (MECMs) on teachers' beliefs about scientific argumentation. *Technology, Pedagogy and Education*, 28(2), 173–190. <https://doi.org/10.1080/1475939X.2019.1583121>
- \*Marco-Bujosa, L. M., McNeill, K. L., González-Howard, M., & Loper, S. (2017). An exploration of teacher learning from an educative reform-oriented science curriculum: Case studies of teacher curriculum use. *Journal of Research in Science Teaching*, 54(2), 141–168. <https://doi.org/10.1002/tea.21340>
- \*Martin, A. M., & Hand, B. (2009). Factors affecting the implementation of argument in the elementary science classroom. A longitudinal case study. *Research in Science Education*, 39(1), 17–38. <https://doi.org/10.1007/s11165-007-9072-7>
- \*McDonald, C. V., & Heck, D. (2012). 'How do we teach argumentation in the new Australian Curriculum?' Secondary science teachers' experiences in an argumentation-based professional development program. *Teaching Science*, 58(3), 22–28.
- \*Mchenry, N., Borger, L., & Liable-Sands, L. (2017). A novel approach to professional development in middle school science: instructional coaching by university professors to improve the instructional core. *Journal of Curriculum and Teaching*, 6(2), 59. <https://doi.org/10.5430/jct.v6n2p59>
- \*Morge, L., Toczek, M.-C., & Chakroun, N. (2010). A training programme on managing science class interactions: Its impact on teachers' practices and on their pupils' achievement. *Teaching and Teacher Education*, 26, 415–426.
- \*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>
- \*Ogunniyi, M. B. (2007). Teachers' stances and practical arguments regarding a science-indigenous knowledge curriculum: Part 1. *International Journal of Science Education*, 29(8), 963–986. <https://doi.org/10.1080/0950069060931020>
- \*Osborne, J. F., Borko, H., Fishman, E., Gomez Zaccarelli, F., Berson, E., Busch, K. C., Reigh, E., & Tseng, A. (2019). Impacts of a practice-based professional development program on elementary teachers' facilitation of and student engagement with scientific argumentation. *American Educational Research Journal*, 56(4), 1067–1112. <https://doi.org/10.3102/0002831218812059>
- \*Ramorogo, G., & Ogunniyi, M. (2010). Exploring teachers' conceptions of the rainbow using an argumentation-based intervention. *African Journal of Research in Mathematics, Science and Technology Education*, 14(1), 24–35. <https://doi.org/10.1080/10288457.2010.10740670>
- \*Roberts, R., & Gott, R. (2010). Questioning the evidence for a claim in a socio-scientific issue: An aspect of scientific literacy. *Research in Science & Technological Education*, 28(3), 203–226. <https://doi.org/10.1080/02635143.2010.506413>
- \*Rosaen, C. L., Lundeborg, M., Terpstra, M., Cooper, M., Niu, R., & Fu, J. (2010). Constructing videocases to help novices learn to facilitate discussions in science and English: How does subject matter? *Teachers and Teaching*, 16(4), 507–524. <https://doi.org/10.1080/13540601003754905>

- \*Sánchez-Martín, J., Zamora-Polo, F., Moreno-Losada, J., & Parejo-Ayuso, J. P. (2017). Innovative education tools for developing ethical skills in university science lessons. The case of the Moral Cross Dilemma. *Journal of Applied Ethics*, 8, 225–245.
- \*Saribaş, D., Ceyhan, G. D., & Lombardi, D. (2019). Bilimin doğasının öğretimi sırasında bilimsel pratiklere ve kanıta dayalı öğretime yakından bakış: Öğretmen yetiştirme programında bir çalışma. *İlköğretim Online*, 343–366. <https://doi.org/10.17051/ilkonline.2019.527626>
- \*Shemwell, J. T., Gwarjanski, K. R., Capps, D. K., Avargil, S., & Meyer, J. L. (2015). Supporting teachers to attend to generalisation in science classroom argumentation. *International Journal of Science Education*, 37(4), 599–628. <https://doi.org/10.1080/09500693.2014.1000428>
- \*Simon, S. (2008). Using Toulmin's Argument Pattern in the evaluation of argumentation in school science. *International Journal of Research & Method in Education*, 31(3), 277–289. <https://doi.org/10.1080/17437270802417176>
- \*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>
- \*Simon, S., & Johnson, S. (2008). Professional learning portfolios for argumentation in school science. *International Journal of Science Education*, 30(5), 669–688. <https://doi.org/10.1080/09500690701854873>
- \*Simon, S., Naylor, S., Keogh, B., Maloney, J., & Downing, B. (2008). Puppets promoting engagement and talk in science. *International Journal of Science Education*, 30(9), 1229–1248. <https://doi.org/10.1080/09500690701474037>
- \*Soysal, Y., & Yılmaz-Tuzun, O. (2019). Relationships between teacher discursive moves and middle school students' cognitive contributions to science concepts. *Research in Science Education*. <https://doi.org/10.1007/s11165-019-09881-1>
- \*Tekbiyik, A. (2015). The use of Jigsaw collaborative learning method in teaching socio-scientific issues: the case of nuclear energy. *Journal of Baltic Science Education*, 14(2), 17.
- \*Torres, N., & Christancho, J. G. (2018). Analysis of the forms of argumentation of teachers in training in the context of a socio-scientific issue. *Journal of Turkish Science Education*, 15(1), 23. <https://doi.org/10.12973/tused.10221a>
- \*Ünal Çoban, G., Akpınar, E., Baran, B., Kocagül Sağlam, M., Özcan, E., & Kahyaoğlu, Y. (2016). The Evaluation of "Technological Pedagogical Content Knowledge based Argumentation Practices" Training for Science Teachers. *Ted Eğitim ve Bilim*, 41(188). <https://doi.org/10.15390/EB.2016.6615>
- \*Vieira, R. D., da Bernardo, J. R., & R., Evagorou, M., & Melo, V. F. de. (2015). Argumentation in science teacher education: The simulated jury as a resource for teaching and learning. *International Journal of Science Education*, 37(7), 1113–1139. <https://doi.org/10.1080/09500693.2015.1022623>
- \*Yaman, F. (2018). Pre-service science teachers' development and use of multiple levels of representation and written arguments in general chemistry laboratory Courses. *Research in Science Education*. <https://doi.org/10.1007/s11165-018-9781-0>
- \*Zaccarelli, F. G., Schindler, A.-K., Borko, H., & Osborne, J. (2018). Learning from professional development: A case study of the challenges of enacting productive science discourse in the classroom. *Professional Development in Education*, 44(5), 721–737. <https://doi.org/10.1080/19415257.2017.1423368>
- \*Zemba-Saul, C. (2009). Learning to teach elementary school science as argument. *Science Education*, 93(4), 687–719. <https://doi.org/10.1002/sce.20325>

### Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Submit your manuscript to a SpringerOpen® journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► [springeropen.com](https://www.springeropen.com)