International Journal of Mathematical Education in Science and Technology

ISSN: 0020-739X (Print) 1464-5211 (Online) Journal homepage: https://www.tandfonline.com/loi/tmes20

Proof and argumentation in K-12 mathematics: a review of conceptions, content, and support

Tye G. Campbell, Justin D. Boyle & Shande King

INTERNATIONAL JOURNAL OF MATHEMATICAL EDUCATION IN SCIENCE AND TECHNOLOGY
https://doi.org/10.1080/0020739X.2019.1626503

Proof and argumentation in K-12 mathematics: a review of conceptions, content, and support

Tye G. Campbell [a], Justin D. Boyle[a] and Shande King[b]

[a]Department of Curriculum and Instruction, The University of Alabama, Tuscaloosa, USA; [b]Department of Teacher Education, The University of Tennessee, Knoxville, USA

ABSTRACT

Current international curriculum frameworks and research advocate for K-12 mathematics to emphasize proof and argumentation. This review synthesizes literature related to conceptions, content, and supportive actions for proof and argumentation across four grade-level bands: K-2, 3–5, 6–8, and 9–12. Our systematic search and subsequent screening process yielded 76 articles for coding and analysis. The findings suggest (a) a variety of conceptualizations that are imbalanced amongst the grade bands, (b) there is more research related to proof and argumentation in middle (6–8) and high school (9–12) compared to elementary school (K-5), (c) the type of content used to solicit argumentation is not well-balanced, and (d) supportive actions include aspects related to classroom culture, individual accountability, and conjecturing before justifying.

ARTICLE HISTORY
Received 12 November 2018

KEYWORDS
Proof; argumentation; justification; mathematics; systematic literature review

1. Introduction

Proof and argumentation are important process standards in the teaching and learning of mathematics. Schoenfeld (2009) described it as the 'lifeblood' of the subject, and others similarly claim it is an integral component of learning mathematics (Hanna & Jahnke, 1996; Stylianides, Bieda, & Morselli, 2016). Recently, proof has received increased attention in curriculum frameworks for school mathematics around the world (Stylianides, Stylianides, & Weber, 2017). Notably, the National Mathematics Curriculum in England purports that students of all ages should 'reason mathematically by following a line of enquiry, conjecturing relationships and generalizations, and developing an argument, justification or proof using mathematical language' (Department for Education, 2013, p. 3). Similarly, the Common Core State Standards in the United States purports students of all ages should 'understand and use stated assumptions, definitions, and previously established results in constructing arguments' (Common Core State Standards Initiative, 2010, para. 4). Proof and argumentation are increasingly being emphasized as process standards that even elementary students can and should engage with (Stylianides, 2016). In light of this global emphasis on proof and argumentation in school mathematics, we aim to review literature to inform the field about: (a) differing conceptualizations of proof and argumentation

CONTACT Tye G. Campbell tgcampbell1@crimson.ua.edu Department of Curriculum and Instruction, The University of Alabama, Graves Hall, Tuscaloosa, AL 35487, USA

across four K-12 grade bands, (b) the mathematical content used to study proof and argumentation across each grade band, and (c) common instructional supports to promote the development of arguments for all K-12 students.

Since this review focuses on proof in school mathematics, we frame our study through the lens of the United States National Council of Teachers of Mathematics' (NCTM) (2000) specific recommendations for proof across four grade bands: K-2, 3–5, 6–8, and 9–12. NCTM (2000) created principles and standards for students in U.S. schools, but their recommendations have made a global impact and is often cited in proof-related literature in contexts other than the United States (e.g. Komatsu, 2010; Paneque, Cobo, & Fortuny, 2017; Patkin, 2012; Sen & Guler, 2015). Their recommendations also provide the most comprehensive set of standards for proof in K-12 mathematics by clearly explaining the teaching and learning processes for proof at each grade band. The three consistent recommendations that NCTM (2000) made across the grade levels include that students should: (a) use forms of reasoning that do not rely on the teacher or other authority figures; (b) create their own conjectures and explore their validity; and (c) maintain skepticism about mathematical properties. These three consistent recommendations are held constant across each grade band, but the specific mathematical practices are further specified within the aforementioned grade bands. Therefore in this review, we focus on the four NCTM (2000) grade bands and compare the similarities and differences between the recommendations within and across each grade band against what has been researched.

We situate our review in the context of two recent literature reviews performed by Stylianides et al. (2017) and one working group session by Staples et al. (2016). Stylianides et al. (2016) reviewed PME proceedings papers written between 2005 and 2015 related to proof and argumentation. They chose to focus on proof and argumentation because '(1) argumentation and proof are closely related, and (2) considering both argumentation and proof helps draw attention to a wider range of important processes related to proving than when considering them separately' [p. 316]. Aligning with this sentiment, our review similarly considers articles related to proof and argumentation. In their review, they noted themes related to student conceptions, classroom-based research, and teacher knowledge. Following this review, Stylianides, et al. (2017) reviewed literature related to proof from three differing perspectives. The three perspectives included proving as a form of problem-solving, proving as convincing, and proving as a socially-embedded activity. Staples et al. (2016) formed a working group at PME-NA to determine the perspectives of what the field calls argumentation, justification, and proof. Through their interactions, they noted there was great variability in how the field understood argumentation and proof. One of the purposes for this paper is to further add to this conversation by providing distinctions about how proof and argumentation is conceptualized within and across each grade band using Stylianides, et al.'s (2017) three perspectives.

In light of the recent reviews of literature and working groups, the novelty of our work lies in examining the distinction of conceptualizations of proof and argumentation for each grade band, the content used to solicit proof and argumentation across each grade band, and supportive actions for increasing argumentative capacities for K-12 students. Analyzing these important aspects of proof and argumentation across the grades can inform and guide future research and practice in school mathematics. The research questions that guided this review were as follows:

(1) How do scholars conceptualize proof and argumentation across each grade band?
(2) How are each grade band and content represented in research focusing on proof and argumentation?
(3) How does the literature suggest K-12 students should be supported in creating arguments?

2. Methodology

We used a systematic review methodology (Cooper, 2017; Hannes & Claes, 2007) to find and analyze research studies on K-12 proof and argumentation. The protocol for a systematic literature review entails the following: (1) create a list of search terms and choose databases in which the search will be run, (2) run a search and gather all articles which contain the search terms, and (3) use a screening process to choose articles based on a predetermined set of criteria. After consulting with an education research librarian, we ran an initial search in three databases: ERIC EBSCOhost, PsycINFO, and Education Full Text (H.W. Wilson). Due to the release of the NCTM (2000) recommendations for reasoning and proof, we considered articles written after the year 2000. Therefore, the following limiters were applied to the initial search: (a) published between January 1, 2000 and May 16, 2018; and (b) peer-reviewed. We created a bank of search terms using words commonly found in curriculum documents (Common Core State Standards Initiative, 2010; Department for Education, 2013; National Council of Teachers of Mathematics, 2000), literature reviews (Stylianides, 2016; Stylianides et al., 2017), working groups (Staples et al., 2016), or research studies which claim to examine proof and argumentation in school mathematics. While it is likely impossible to consider all terms related to proof and argumentation, the chosen bank of terms are well-founded in the literature and are used in conjunction with the words *proof* and *argumentation*. We retrieved articles that included the search terms appearing anywhere within the article. Table 1 shows the combinations of search terms used and the total output of each search with the limiters.

After retrieving an initial 8094 articles, we transported each article from the research databases to Refworks, a tool for creating bibliographies. Then, we removed all duplicates, which left 5263 articles. After deleting duplicates, we transported the citations to Google Sheets for screening purposes. We screened the title and abstract of all 5263 articles using

Table 1. Summary of database search.

Word search	ERIC EBSCOhost	PsychINFO	Education full text (H.W. Wilson)
Proof + math*	822	767	928
Prove + math*	293	1055	240
Proving + math*	182	148	139
Argument*+math*	788	1018	547
Justify + math*	177	221	102
Justification + math*	249	265	153
Total			8094
Total after duplicates			5263

Note: Any letters included after * were included in the search.

inclusion criteria. Our inclusion criteria were as follows: (1) *content* – the focus of the study must be on proof or argumentation; (2) *participants* – research subjects must be K-12 students; (3) *empirical*– the study must only report empirical findings; (4) *publication type* – the publication must be a journal article; (5) *language* – the article must be written in English.

Screening involved testing for overall inclusion based on three phases: (1) the article title met inclusion criteria; (2) the abstract of the article met inclusion criteria; and (3) the article met inclusion criteria after a full scan of the article. After the first phase, 902 articles remained for the abstract screen. After the second phase, 119 articles remained for a full scan. Upon completion of the third phase, 73 articles remained for inclusion in the systematic review. The third author double coded 10% of the articles at phase 1 and phase 2 of the screening process to calculate inter-rater reliability (as calculated by dividing the number of agreements by the sum of agreements and disagreements and multiplying by 100) with a cutoff set for 85% agreement. For phase 1, the inter-rater reliability was 90.7%, and for phase 2 the inter-rater reliability was 91.1%. After screening articles for inclusion, we followed the same screening process to evaluate references of the 73 articles. The ancestral search yielded 3 more articles for a total of 76 articles included in this review. All articles included in the review are listed in the appendix.

The 76 articles were rigorously coded for thematic analyses (see Table 2 for coding scheme). We coded for several predetermined indicators to answer our research questions. For analyses, we categorized themes that emerged via a constant comparison approach (Glaser & Strauss, 2009), and we also used descriptive statistical analyses to describe trends in the research. We further describe our analyses within each section of the results.

3. Analysis and results

3.1. Conceptualizations of proof and argumentation

To understand how scholars conceptualized proof and argumentation across the grades, we analyzed each article according to the participants' grade band. The grade bands were based on the trajectory of students in U.S. schools. Therefore, studies from other countries that did not follow this trajectory were correlated to the U.S. grade bands based on the student ages. First, we considered the types of tasks that scholars used to solicit argumentation. We noted whether the tasks allowed for general arguments or arguments for specific cases. Then, we considered the theoretical framework, specific quotations, methods, and findings to determine how scholars conceptualized proof and argumentation in their work. To categorize scholars' conceptualizations of proof and argumentation, we utilized Stylianides, et al.'s (2017) framework which consists of proving as a form of problem-solving, proving as convincing, and proving as a socially-embedded activity. Their framework, including the central tenets and critical questions/issues from each conceptualization, is detailed in Table 3. We begin each section by providing an overview of NCTM's (2000) recommendations at each specific grade band. Then, we report our thematic findings and compare our findings with the recommendations in the discussion section. Table 4 provides a summary of our findings.

Table 2. Article coding.

What was the participants' grade level? 1 = K-2, 2 = 3–5, 3 = 6–8, 4 = 9–12
How were the participants described? 1 = gifted, 2 = average, 3 = academically at risk, 4 = otherwise at risk, 5 = underachieving/below grade level, 6 = learning disabled, 7 = Low SES, 8 = Middle SES, 9 = High SES, 10 = Other (specify)
What sexes were represented in the sample? 1 = males, 2 = females, 3 = mixed, 4 = no sex information given
What content was used to solicit argumentation? 1 = Algebra, 2 = Number Theory, 3 = Geometry, 4 = Precal/Trig/Cal 5 = Other (specify), 6 = Unknown
What was the duration of the study? 1 = One day or less, 2 = 1 week or less, 3 = one month or less, 4 = one marking period or less, 5 = one semester (or summer) or less, 6 = one school year or less, 7 = more than one school year (specify)
List theoretical/conceptual framework used
How are arguments coded?
Instructional intervention to solicit argumentation (specify)
Was any type of technology used to solicit argumentation? 1 = yes, 2 = no. Specify the type of technology.
How were students working on argumentative tasks? 1 = individually, 2 = collaboratively, 3 = NA
The article uses the words (1) proof, (2) argumentation, or (3) both
How was proof/argumentation defined? Specify by direct quote if possible.
Methods used: 1 = quantitative, 2 = qualitative, 3 = mixed methods
Specify the methods used
Explain the findings

Table 3. Stylianides, et al.'s (Department for Education [2013])perspectives of proof and argumentation.

	Proving as problem-solving	Proving as convincing	Proving as a socially-embedded practice
Central tenets	Focuses on the process of proof rather than the product and tends to investigate students' understandings of logical aspects of proof	Focuses on convincing oneself and peers of the truth of a mathematical claim and places emphasis on students' interpretation of what is meaningful and acceptable.	Focuses on the social aspects of proof such as collaborating with others or justifying their arguments to others, and it tends to focus on the activity of proof rather than understanding.
Critical questions/issues	Which competencies are necessary to successfully engage in proof? What processes do students use to create mathematical arguments?	What types of arguments convince students? How do their convictions align or misalign with the field of mathematics?	Issues: Creating social norms for improving proof practices, seeking classroom environments conducive to proof activities.

Table 4. Conceptualization of proof and argumentation.

	K-2	3–5	6–8	9–12
Type of argument task allows				
General argument (argument across all cases)	80% (4/5)	57% (8/14)	71% (24/34)	93% (39/42)
Argument for one or multiple cases	20% (1/5)	43% (6/14)	29% (10/34)	7% (3/42)
Conceptualization of argumentation and proof				
Proof as problem-solving	60% (3/5)	36% (5/14)	41% (14/34)	71% (30/42)
Proof as convincing	0% (0/5)	21% (3/14)	26% (9/34)	21% (9/42)
Proof as socially-embedded practice	40% (2/5)	43% (6/14)	32% (11/34)	7% (3/42)

3.1.1. Grade band K-2

NCTM (2000) recommended pattern-recognition and classification skills as the two most important elements for reasoning and proof at the K-2 grade level. Students at this level are expected to create empirical arguments, or arguments based on examples. K-2 students should be encouraged to conjecture and explain their thinking by stating reasons. They might also start to justify their answers by using short chains of deductive reasoning. Students should slowly begin to use mathematical properties and relationships to create arguments.

3.1.1.1. Task for soliciting argumentation. Between January 2000 and May 2018, five studies (out of 76) were conducted at the K-2 grade band. At this grade band, four studies used tasks that warranted a justification of a single case or multiple cases (Kosko, 2016; Krummheuer, 2013; Strom, Kemeny, Lehrer, & Forman, 2001; Whitenack & Knipping, 2002) and one study examined students' abilities to generalize across cases (Barrett, Clements, Klanderman, Pennisi, & Polaki, 2006). Whitenack and Knipping's (2002) study is an example of students justifying a single case. The students created arguments to explain why 31-15 = 16. The focus of this study was on the explanations that students created to convince their classmates of the truth of an arithmetic problem. Barrett et al. (2006) was the single study in this grade band to go beyond arguing empirically to examine a generalization across cases. Barrett et al. (2006) researched students' abilities to form as many rectangles as possible with a perimeter of 24. The 2nd grade students in their study used straws as models to create the rectangles to form a conjecture and justify the conjecture. Their study provided an existence case that K-2 students have the ability to generalize across cases.

3.1.1.2. Conceptualization of argumentation and proof. Three studies conceptualized proof as a form of problem-solving (Barrett et al., 2006; Kosko, 2016; Krummheuer, 2013). For instance, Kosko (2016) sought to understand how elementary students used given information to create mathematical arguments. His focus was on the types of references students made to a given statement and the impact of the reference on students' success at creating an argument. This is indicative of proof as problem-solving because it focuses on the competencies and processes for successfully engaging in argumentation. No studies conceptualized proof as a form of convincing in this grade band, while two studies were categorized as a socially-embedded practice (Strom et al., 2001; Whitenack & Knipping, 2002). Whitenack and Knipping's (2002) research portrayed the classroom community as validators of truth. Students in their study iteratively created new explanations until the classroom reached agreement on the validity of an argument indicating the focus was on students' participation in the mathematical community. The articles in this grade band exhibited a focus on enhancing argumentative capacities through examining processes related to proof and argumentation, while also emphasizing the social role of the classroom community.

3.1.2. Grade band 3–5

NCTM (2000) asserted that students in grades 3–5 should understand that examples do not count as warrants for viable arguments. They should move from creating arguments about individual cases and start to create general arguments. Students should learn to critique their own and other's reasoning. They should base their arguments on mathematical properties, structures, and relationships. Teachers should facilitate students' thinking within a mathematical community and continually develop, test, and apply conjectures.

3.1.2.1. Task for soliciting argumentation. Fourteen (out of 76) research studies were coded at the 3–5 grade band. Six of the 14 studies used tasks that elicited empirical argumentation while the remaining eight studies used tasks requiring a general justification. Mueller and Yankelewitz (2014) used a task wherein students were asked to construct an argument for a single fraction bar. In contrast, Komatsu (2016) created a task where students were to create a conjecture and justification for the sum of a two-digit number and the integer whose digits are in reverse order (e.g. 41 + 14). This type of task requires students to conjecture and prove over a class of objects rather than justifying a single case.

3.1.2.2. Conceptualization of argumentation and proof. Five (of the 14) studies conceptualized proof as a form of problem-solving at the 3–5 grade band, while focusing on processes for specific forms of reasoning. For instance, Yim, Song, and Kim (2008) and Komatsu (2016) both investigated how students refined their conjectures and arguments when met with a counter-example. Three studies focused on proving as convincing by interviewing students to understand how they accepted claims as true in mathematics. For instance, in Flores (2006) study, elementary through high school students were asked how they knew what they learned in mathematics was true. Specifically, the study focused on the proof schemes that convinced students of truth, and it was found many students relied on authority figures such as the teacher for conviction. The remaining six studies examined proof as a socially-embedded practice, and Rojas-Drummond and Peón's (2004) conceptualization encapsulates the social embeddedness indicative of similar studies at this grade band in their conceptualization of argumentation as 'the act of providing reasons to make admissible a certain position, opinion or conclusion, or to confront others' positions, opinions or conclusions' [p. 540]. Some scholars conceptualizing argumentation from this lens sought to understand how non-dominant or marginalized students participate in mathematics classrooms (Civil & Hunter, 2015; Rojas-Drummond & Peón, 2004). For instance, Civil and Hunter (2015) focused on the personal relationships and atmosphere that helped non-dominant students participate and create argumentative positions in the classroom. The 3–5 grade band included studies viewing proof as a socially-embedded practice (6 of 14), proof as convincing (3 of 14), and proof as problem-solving (5 of 14).

3.1.3. Grade band 6–8

NCTM (2000) recommends that students in grades 6–8 make conjectures, evaluate conjectures, examine structures and patterns, and construct arguments for their conjectures. Their arguments should contain enough evidence to convince another middle grade student who is not a member of their own learning community. At this level, proofs need not be formal, but they should share important features of a formal argument. For instance, their reasoning should be clear enough to be evaluated by others. Students should start to develop skepticism about pattern generalization, and they should use inductive and deductive reasoning.

3.1.3.1. Task for soliciting argumentation. Thirty-four (of the 76) articles were coded at the 6–8 grade band. Ten studies utilized tasks which required a justification for specific cases while twenty-four studies utilized tasks that required a general argument. Most tasks required students to create generalizations based on patterns. For instance, Lannin (2005) and Ellis (2007) facilitated a learning sequence where students observed patterns, created algebraic generalizations, and justified their generalizations. In contrast, Matteson, Capraro, Capraro, and Lincoln (2012) required students to create justifications based on their answer to a standardized test problem. This type of justification required students to argue for a specific answer choice.

3.1.3.2. Conceptualization of argumentation and proof. The 6–8 grade band contained studies that were fairly balanced across the three categories: 14 as problem-solving, 9 as convincing, and 11 as socially-embedded practice. The fourteen (of the 34) studies viewing proof as problem-solving tended to focus heavily on developing deduction skills, and several studies used dynamic geometry software as a tool for increasing students' abilities to create deductive arguments (e.g. Hadas, Hershkowitz, & Schwarz, 2000; Healy & Hoyles, 2002; Jones, 2000). For example, students in Jones (2000) study used a dynamic geometry software as a tool to deduce claims about quadrilaterals. He studied students' interpretations and how their arguments progressed while using the dynamic geometry software. Yopp (2017) designed an intervention for one middle grade student to engage in proof by contrapositive. For the purposes of his study, he defined argumentation rigorously as 'an argument for a mathematical claim that uses valid logic and prior results appropriately sequenced to support the truth of the claim' [p. 155].

The nine proof as convincing studies in the 6–8 grade band often utilized interviews and questionnaires to determine what students found convincing. For instance, Bieda and Lepak (2014) and Liu and Manouchehri (2013) had similar goals in examining what middle school students found convincing while the former used interviews and the latter used a survey. The remaining 11 studies in the category of proof as a socially-embedded practice contained a variety of goals. Mueller and colleagues (Mueller, 2009; Mueller & Maher, 2009; Mueller & Yankelewitz, 2014; Mueller, Yankelewitz, & Maher, 2012) focused on the social and environmental conditions that could aid students who were economically disadvantaged to engage in creating mathematical arguments. Others focused on sociocultural factors related to autonomy in proof (Brown, 2007; Fried & Amit, 2008) while, still, others examined learning opportunities that occurred when students created arguments in collaboration with others (Fukawa-Connelly & Silverman, 2015; Weber, Maher, Powell, & Lee, 2008).

3.1.4. Grade band 9–12

NCTM (2000) recommends that high school students should use sophisticated forms of reasoning. They should constantly ask *why*, and seek to justify mathematical properties across all content. Students should learn a variety of proving techniques such as mathematical induction, contrapositive proof, and proof by counterexample. They should also use multiple forms of representation in their proof construction including paragraph and two-column form. Students should use logic rather than relying on external authority, and critiquing others' arguments should be a natural part of classroom discussion.

3.1.4.1. Task for soliciting argumentation. Forty-two articles were coded at the 9–12 grade band. Nearly all studies (39/42) at this grade band utilized tasks that allowed students to create general arguments. Geometry was the main content for soliciting proof and argumentation, and students were often asked to generalize features of shapes. Though three studies used tasks that required justification of a single or multiple cases, the required justifications went beyond arguing for a particular solution. For instance, Tsamir and Sheffer (2000) studied students' formal argumentations for the case of zero divided by zero.

3.1.4.2. Conceptualization of argumentation and proof. The majority (30 of 42) of studies in the 9–12 grade band examined proof as a form of problem solving. Scholars also tended to conceptualize proof rigorously such as Patkin's (2012) conceptualization of proof as 'a logical sequence of claims leading to a conclusion grounded only on terms and fundamental assumptions (that do not need proving) as well as conclusions drawn earlier on the basis of the terms and fundamental assumptions' [p. 986]. The studies tended to examine proof from a cognitive standpoint, and many focused on the effects of some proof intervention (e.g. Chinnappan, Ekanayake, & Brown, 2012; Lee, Chen, & Chang, 2014; Stout et al., 2011; Yang & Lin, 2012). For example, Yang and Lin (2012) studied the effects of a reading comprehension intervention on students' abilities to create arguments. Others tended to focus on strategic thinking or the knowledge students used to create viable arguments (Magajna, 2013; Martinez & Pedemonte, 2014; Soldano & Arzarello, 2016).

Nine studies were categorized as a form of convincing and were often analyzed based on the proof schemes that students found most convincing (e.g. Lee, 2016; Lin, Yang, & Chen, 2004). The remaining three studies were categorized under proof as a socially-embedded practice. These studies sought to understand the benefits of collaboration in creating mathematical arguments (Cross, 2009; Furinghetti, Olivero, & Paola, 2001) and the instructional identities of students (Aaron & Herbst, 2012). Aaron and Herbst (2012) examined the social norms created between students and teachers and the effects it has on students' willingness to participate in argumentative practices. These findings portray an unbalanced research focus related to how proof is conceptualized in the 9–12 grade band. The current research emphasizes proving as problem-solving while proving as convincing and proving as a socially embedded practice are scarcely represented.

3.2. Grade level and content

In this section, we coded each article for the participant grade level and the type of content that was used to elicit argumentation. We considered four grade bands: (a) K-2, (b) 3–5, (c) 6–8, and (d) 9–12. Further, we coded each article according to eight content areas by analyzing the tasks used for soliciting argumentation. The content areas we coded were as follows: Arithmetic, Algebra, Geometry, Number Theory, Probability, Trigonometry/Precalculus, Calculus, and Physics. To discriminate between number systems (an Arithmetic subcategory) and Number Theory, we coded tasks which appealed to a general statement as Number Theory. For instance, 'the sum of two odd numbers is even' was coded under Number Theory. However, a statement requiring an argument of a single case such as 'three plus five is even' was coded under Arithmetic.

Table 5 shows the percentage of studies with participants at each grade band. The 9–12 grade band accounted for 55% of the studies while grades 6–8 accounted for 45%. The primary grades represented a much lower percentage of publications with grades 3–5 accounting for 18% of the research and grades K-2 accounting for 7% of the research. The percentage of research publications clearly increases as the age of students increases.

Table 5. Percentage of publications by grade band.

Grade band	Percentage of publications
K-2	7% (5/76)
3–5	18% (14/76)
6–8	45% (34/76)
9–12	55% (42/76)

Note: Percentages exceed 100% because some articles studied more than one grade band.

Table 6. Percentage of publications by content strand.

Content	Percentage of publications
Geometry	49% (37/76)
Algebra	24% (18/76)
Number theory	17% (13/76)
Arithmetic	13% (10/76)
Probability	4% (3/76)
Trigonometry/Precalc	3% (2/76)
Physics	1% (1/76)
Unknown*	7% (5/76)

Note: Percentages exceed 100% because some articles utilized more than one content strand. *The content strand was coded as 'Unknown' if the task and/or content were not specified in the article.

Table 7. Percentage of top four content areas in each grade band.

	Content			
Grade level	Arithmetic	Number theory	Algebra	Geometry
K-5	29% (5/17)	24% (4/17)	0%	24% (4/17)
6–8	15% (5/34)	15% (5/34)	24% (8/34)	35% (12/34)
9–12	2% (1/42)	14% (6/42)	24% (10/42)	64% (27/42)

To understand the types of content related to research on argumentation, we coded the tasks used to solicit argumentation by content area (see Table 6). Researchers relied on tasks involving Geometry in almost half of the studies. Algebra (24%), Number Theory (17%), and Arithmetic (13%) were moderately represented in the data. Subjects such as Trigonometry and Probability were scarcely represented, and no publications utilized Calculus tasks.

We also analyzed the content used across each grade band for the four most common content areas found in the research: Geometry, Algebra, Number Theory, and Arithmetic (see Table 7). Due to the scarcity of research studies in the primary grades, K-5 was considered as one grade band. Of the publications in the K-5 grade band, Arithmetic was the leading content used (29%). Researchers also relied heavily on Number Theory (24%) and Geometry (24%) at the K-5 grade band. Algebra was not used in any studies for the primary grades.

The 6–8 grade band consisted of studies which primarily used Geometry (35%) and Algebra (24%). Number Theory (15%) and Arithmetic (15%) were also moderately used. The 6–8 grade band was more balanced in content than the other grade bands. The majority of studies at the high school level used Geometry as the content area (64%). Algebra was well-represented (24%) with Number Theory used sparingly (14%). Arithmetic was only used in one study at the 9–12 grade band.

3.3. Supporting proof and argumentation

In this section, we considered each article holistically to understand how scholars and practitioners support students to create arguments. The articles were compared to develop themes of support. Some scholars focused on problematizing proof and argumentation. For instance, that students often rely on empirical arguments was a significant finding in the literature. We choose not to report on the problematics of learning proof and argumentation. Instead, we focused on research which provides empirical evidence for supporting students to create arguments. We also choose to focus on themes that are relevant to each grade band. The literature reviewed in this analysis suggests three themes for aiding K-12 students in argumentation: (1) an open, collaborative classroom culture of argumentation and critique; (2) individual accountability; and (3) allowing students to conjecture before justifying.

3.3.1. Classroom culture

Students' arguments are often heavily influenced by authority (Flores, 2006; Fried & Amit, 2008; Sen & Guler, 2015). That is, students rely on teachers, textbooks, or some other external person or object as a warrant for their claims. To increase student autonomy, some scholars suggest that students should work collaboratively with one another to create and critique arguments (Brown, 2007; Cross, 2009; Mueller & Maher, 2009; Mueller & Yankelewitz, 2014; Weber et al., 2008; Whitenack & Knipping, 2002). Mueller and Yankelewitz (2014) developed an instructional sequence wherein students made arguments about fractional representations with Cuisenaire rods. Students created collective arguments with their classmates, and they critiqued others' arguments without direct assistance from the teacher. In fact, those who used faulty reasoning were able to refine their arguments based on the modifications and verbal critiques made by classmates. Weber et al. (2008) similarly allowed students to debate with one another until they reached a consensual understanding.

Further, the literature suggested that an open, collaborative environment of critique increased students' participation in argumentation (Brown, 2007; Civil & Hunter, 2015; Rojas-Drummond & Peón, 2004). Civil and Hunter's (2015) study suggested that non-dominant students participated in argumentation more regularly when the culture of the classroom was collaborative and safe. Brown (2007) similarly found in a case study that one student's social position changed to active forms of participating and doing mathematics as a result of engaging in collective argumentation. Brown (2007) created a comprehensive strategy for engaging students in collective argumentation through six phases: represent, compare, explain, justify, agree, and validate. First, students 'represent' an argument on their own to be 'compared' with those in a small group. They 'explain' and 'justify' their work to others and 'agree' on a common argument to be shared with the class. Lastly, each group shares their collective arguments to be 'validated' by the classroom community.

3.3.2. Individual accountability

Though scholars suggested that a collaborative culture enhanced student arguments, many deemed individual accountability a necessity (Brown, 2007; Cáceres, Nussbaum, Marroquín, Gleisner, & Marquínez, 2018; Cross, 2009; Kosko, 2016; Stoyle & Morris, 2017). Cross (2009) designed a quasi-experimental study to test the effects of collaborative argumentation and writing individual arguments on students' conceptual understanding. She found that writing individual arguments combined with collaborative argumentation were significantly related to students' post-test performance when compared to a control group who received traditional instruction. She conjectured that individual accountability in writing individual arguments played a key role in students gaining conceptual understanding of mathematics topics. Stoyle and Morris (2017) similarly found that combining a collaborative atmosphere with individual argumentation aided students in conceptual understanding. In their study, students individually created blog posts with the purpose of explaining and justifying mathematical topics. Stoyle and Morris (2017) asserted that the asynchronous environment allowed students to reflect on their mathematical reasoning and provided a record of communication. Not only should students individually create arguments after creating collective arguments with others, but Brown (2007) suggested students should create individual arguments before collaborating with others. In doing this, students have ideas to bring to the group setting instead of relying on other members of the group to carry the weight.

3.3.3. Conjecturing and justifying

To bridge the gap between empirical investigations and deductive proving, some researchers suggested that students need to experience cognitive unity (Boero, Garuti, Lemut, & Mariotti, 1996) between conjecturing and proving (Fiallo & Gutiérrez, 2017; Martinez, 2014; Palla, Potari, & Spyrou, 2012). In other words, researchers suggest that conjecture and proof are highly interrelated, and aspects of the conjecture process can help students make connections for proving. Martinez and colleagues (Martinez, 2014; Martinez & Pedemonte, 2014; Martinez, Brizuela, & Castro Superfine, 2011) developed an Algebra calendar task wherein students created a conjecture about the difference of cross products of a four by four calendar square. Students only created correct conjectures after engaging in the proving process. Their study suggested that conjecturing and proving are bidirectional meaning engagement in one increases the sophistication of the other.

Some scholars suggest that leading students to faulty conjectures can increase students' proving abilities (Komatsu, 2010, 2016; Komatsu, Tsujiyama, & Sakamaki, 2014, 2017). Komatsu and colleagues (Komatsu, 2010, 2016; Komatsu et al., 2014, 2017) strategically designed tasks wherein students refined their conjectures and justifications based on finding a counterexample to their original conjecture. In Komatsu's (2010, 2016) studies, students originally made a conjecture that the sum of all two digit numbers with the ones and tens digit flipped produced a number with the same ones and tens digit (e.g. $14 + 41 = 55$). When students were introduced to a counterexample (e.g. $91 + 19 = 110$), they refined their conjectures by deductive guessing. That is, without needing to check empirically, students were able to conjecture that the sum of these numbers would always be equal to eleven times the sum of the ones and tens digit of one of the numbers. Komatsu (2010) suggested that a major reason students were able to refine their conjectures is that they understood why their original conjecture was false. Another successful trait that these students exhibited is that they analyzed the part of their previous justifications that were still applicable to the counterexample. In this way, students made increasingly sophisticated and accurate conjectures and proofs.

4. Discussion

4.1. Conceptualizations of proof and argumentation

Scholars' conceptualizations of proof and argumentation, as determined by the tasks they used and the perspective in which their research was categorized, varied significantly within and across the grades. We found some alignment with the NCTM (2000) recommendations, but some unexpected findings emerged. Most scholars who did research in the early primary grades used tasks that did not require a general argument. Instead, students in their early primary years used argumentation to communicate with their classmates and explain their thinking across specific cases. K-2 scholars' conceptualizations agreed with NCTM (2000) in that a major purpose for proving in K-2 is to explain by stating reasons or examples. However, the recommendations place an emphasis on generalizing from examples which was not a primary consideration in the limited K-2 literature. The lack of generalization in the K-2 grade bands could be troublesome to continuity of proof and argumentation throughout schooling. Notably, there were no studies at the K-2 grade band which analyzed proving as convincing which signifies a gap that future research should explore. It is possible that other conceptualizations of proof were more prevalent because scholars believe communication and explanation are the primary functions of proof and argumentation at the K-2 band.

In the 3–5 grade band, the data is somewhat misaligned with NCTM's (2000) recommendation that students should consider classes of objects in their arguments. There was an increased emphasis on creating general arguments compared to the K-2 grade band, but many studies used tasks that warranted an explanation of one or multiple cases. There was a high prevalence of studies contained within the proving as a socially-embedded practice category which aligns with NCTM's (2000) call for critiquing arguments, but it is unclear whether these types of discourses were critical of general mathematical arguments. Instead, the discourse was more related to talking mathematically in general. Future research should explore justifying and critiquing arguments in collaboration with peers in grades 3–5. Similar to the K-2 grade band, few studies examined what convinced 3–5 students which might signify the need for more research on this conceptualization for elementary grades as a whole.

The 6–8 grade band mostly used tasks that required general arguments. However, there were still a significant number of studies that used tasks that did not require a generalization. Arguing for specific cases is important to conceptual understanding, but NCTM (2000) envisioned students creating informal general arguments by the time they reached the middle grades. The perspectives of research at the middle grades were more balanced than the other grade bands signifying a variety of research objectives amongst scholars. Surprisingly, some scholars believed proof should be a formal, rigorous exercise which is not an expectation listed in the recommendations for grade 6–8 students. It is reasonable to question whether middle grade students have the ability to conceptually understand arguments relying on formal axioms, theorems, and definitions. Research at the 6–8 grade band has made a major contribution to proof as a socially-embedded practice. Scholars such as Mueller and colleagues (Brown, 2007; Fried & Amit, 2008; Mueller, 2009; Mueller et al., 2012; Mueller & Maher, 2009; Mueller & Yankelewitz, 2014) have made important contributions that could profitably be recognized and built upon in different grade bands. The research provided by these scholars brought attention to issues related to autonomy and social factors in proof that were largely missing in other grade bands.

Nearly all scholars at the 9–12 grade band used tasks that required general justifications which is consistent with NCTM's (2000) recommendations. On the other hand, there was very little research conceptualizing proof as a socially-embedded practice which is problematic. Most of the research at the 9–12 band was cognitively situated and did not examine sociocultural factors of proof. This is a major gap considering NCTM's (2000) admonition that students should prove and critique in collaboration with others. Future research should examine proof as a socially-embedded activity building off the work of scholars such as Aaron and Herbst (2012). Though there was a fairly high prevalence of research in the other two categories, there is still much work to be done on proof as problem-solving and proof as convincing in the 9–12 grade band. Future research might examine specific forms of proof such as proof by induction as recommended by NCTM (2000).

As a whole, researchers' conceptualizations somewhat aligned to the recommendations for proof and argumentation. However, NCTM (2000) only acknowledged non-general arguments for K-2 students. Scholars found relevance for creating arguments for one or multiple cases from the early primary grades into the middle grades. Further, scholars have conceptualized proof and argumentation in a variety of ways. It is not necessary that scholars come to agreement on a conceptualization of argumentation and proof, but it is crucial that scholars provide a clear description of how they understand it in their work–a practice that many neglected. In the future, it might be beneficial for mathematics education stakeholders to further describe proof and argumentation at each grade level accounting for the conceptualizations held in this review.

4.2. Grade level and content

Our review suggests that Geometry remains the primary content for researching proof and argumentation. This could be due to several factors including traditional views of Geometry as the subject in which one does proof or limited curricular resources for proof in other subject areas. However, scholars often used tasks from Algebra and Number Theory to solicit argumentation as well. The recommendations called for creating arguments across all content areas. Tasks related to higher level mathematics content such as Precalculus, Trigonometry, and Calculus were particularly scarce in our data. Specifically, in the higher grades, scholars should devote much more attention to these content areas. The primary and middle grades were balanced in the types of content used to solicit argumentation. Based on the recommendations, scholars might place an equal emphasis on each content area in the upper grades.

Very little research was devoted to students' engagement in argumentation in primary grades. This is problematic because the recommendations clearly argue for creating arguments across each grade level. It is possible that the word search used for this review left out articles related to proof and argumentation in the primary grades. However, we used words directly related to the language used in the recommendations, literature reviews, and working groups. Therefore, it is reasonable to expect scholars to use language such as *proof*, *justification*, and *argument* in their work with primary grade students. If scholars define proof and argumentation differently in primary grades, it should be reflected in policy. We recommend that scholars might devote much more research to proof and argumentation in the primary grades.

5. Conclusion

In this review, we examined conceptualizations, content, and support for proof in K-12 contexts and used NCTM's (2000) globally recognized recommendations to understand the current state of research and how it might move forward. This review provided areas of alignment and misalignment with the recommendations, and it also provided suggestions for future research. It is necessary for all stakeholders to follow the example of Staples et al.'s (2016) working group and continue to have conversations about the state of research in proof and argumentation. Our review suggests it might be necessary to discuss these important processes on a grade-level basis. Still, scholars must continue to engage in discourse across grade levels to ensure a continuum of proof and argumentation consistent throughout school mathematics. These discourses will lead to increased understanding to ensure effective practices for the teaching and learning of proof and argumentation.

Disclosure statement

No potential conflict of interest was reported by the authors.

ORCID

Tye G. Campbell http://orcid.org/0000-0001-7827-8430

References

References marked with an asterisk signify articles included in the systematic review.

*Aaron, W. R., & Herbst, P. (2012). Instructional identities of geometry students. *The Journal of Mathematical Behavior*, *31*(3), 382–400.

*Barrett, J. E., Clements, D. H., Klanderman, D., Pennisi, S.-J., & Polaki, M. V. (2006). Students' coordination of geometric reasoning and measuring strategies on a fixed perimeter task: Developing mathematical understanding of linear measurement. *Journal for Research in Mathematics Education*, *37*, 187–221.

*Bieda, K. N., & Lepak, J. (2014). Are you convinced? Middle-grade students' Evaluations of mathematical arguments. *School Science and Mathematics*, *114*(4), 166–177.

Boero, P., Garuti, R., Lemut, E., & Mariotti, M. A. (1996). Challenging the traditional school approach to theorems: A hypothesis about the cognitive unity of theorems. In L. Puig, & A. Gutiérrez (Eds.), *Proceedings of the 34th PME Conference* (pp. 113–120). Valencia: PME.

*Brown, R. (2007). Exploring the social positions that students construct within a classroom community of practice. *International Journal of Educational Research*, *46*(3), 116–128.

*Cáceres, M., Nussbaum, M., Marroquín, M., Gleisner, S., & Marquínez, J. T. (2018). Building arguments: Key to collaborative scaffolding. *Interactive Learning Environments*, *26*, 355–371.

*Chinnappan, M., Ekanayake, M. B., & Brown, C. (2012). Knowledge use in the construction of geometry proof by Sri Lankan students. *International Journal of Science and Mathematics Education*, *10*(4), 865–887.

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

Common Core State Standards Initiative. (2010). Common core state standards for mathematics. Retrieved from http://www.corestandards.org/Math/
Cooper, H. (2017). *Research synthesis and meta-analysis: A step-by-step approach*. Los Angeles, CA: Sage.
*Cross, D. I. (2009). Creating optimal mathematics learning environments: Combining argumentation and writing to enhance achievement. *International Journal of Science and Mathematics Education*, *7*(5), 905–930.
Department for Education. (2013). Mathematics: programmes of study: key Stages 1–2 (national curriculum in England). Retrieved from https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/239129/PRIMARY_national_curriculum_-_Mathematics.pdf
*Ellis, A. (2007). Connections between generalizing and justifying: Students' reasoning with linear relationships. *Journal for Research in Mathematics Education*, *38*, 194–229.
*Fiallo, J., & Gutiérrez, A. (2017). Analysis of the cognitive unity or rupture between conjecture and proof when learning to prove on a grade 10 trigonometry course. *Educational Studies in Mathematics*, *96*(2), 145–167.
*Flores, A. (2006). How do students know what they learn in middle school mathematics is true? *School Science and Mathematics*, *106*(3), 124–132.
*Fried, M. N., & Amit, M. (2008). The co-development and interrelation of proof and authority: The case of Yana and Ronit. *Mathematics Education Research Journal*, *20*(3), 54–77.
*Fukawa-Connelly, T., & Silverman, J. (2015). The development of mathematical argumentation in an unmoderated, asynchronous multi-user dynamic geometry environment. *Contemporary Issues in Technology and Teacher Education*, *15*(4), 445–488.
*Furinghetti, F., Olivero, F., & Paola, D. (2001). Students approaching proof through conjectures: Snapshots in a classroom. *International Journal of Mathematical Education in Science and Technology*, *32*(3), 319–335.
Glaser, B. G., & Strauss, A. L. (2009). *The discovery of grounded theory: Strategies for qualitative research*. New Brunswick, NJ: Aldine Transaction.
*Hadas, N., Hershkowitz, R., & Schwarz, B. B. (2000). The role of contradiction and uncertainty in promoting the need to prove in dynamic geometry environments. *Educational Studies in Mathematics*, *44*(1–2), 127–150.
Hanna, G., & Jahnke, H. N. (1996). Proof and proving. In A. Bishop, M. Clements, C. Keitel-Kreidt, J. Kilpatrick, & C. Laborde (Eds.), *International handbook of mathematics education* (pp. 877–908). Dordrecht: Kluwer.
Hannes, K., & Claes, L. (2007). Learn to Read and Write systematic reviews: The Belgian Campbell group. *Research on Social Work Practice*, *17*, 748–753.
*Healy, L., & Hoyles, C. (2002). Software tools for geometrical problem solving: Potentials and pitfalls. *International Journal of Computers for Mathematical Learning*, *6*(3), 235–256.
*Jones, K. (2000). Providing a foundation for deductive reasoning: Students' interpretations when using dynamic geometry software and their evolving mathematical explanations. *Educational Studies in Mathematics*, *44*(1-2), 55–85.
*Komatsu, K. (2010). Counter-examples for refinement of conjectures and proofs in primary school mathematics. *The Journal of Mathematical Behavior*, *29*, 1–10.
*Komatsu, K. (2016). A framework for proofs and refutations in school mathematics: Increasing content by deductive guessing. *Educational Studies in Mathematics*, *92*(2), 147–162.
*Komatsu, K., Jones, K., Ikeda, T., & Narazaki, A. (2017). Proof validation and modification in secondary school geometry. *The Journal of Mathematical Behavior*, *47*, 1–15.
*Komatsu, K., Tsujiyama, Y., & Sakamaki, A. (2014). Rethinking the discovery function of proof within the context of proofs and refutations. *International Journal of Mathematical Education in Science and Technology*, *45*(7), 1053–1067.
*Kosko, K. W. (2016). Making use of what's given: Children's detailing in mathematical argumentative writing. *The Journal of Mathematical Behavior*, *41*, 68–86.
*Krummheuer, G. (2013). The relationship between diagrammatic argumentation and narrative argumentation in the context of the development of mathematical thinking in the early years. *Educational Studies in Mathematics*, *84*, 249–265.

*Lannin, J. K. (2005). Generalization and justification: The challenge of introducing algebraic reasoning through patterning activities. *Mathematical Thinking and Learning*, *7*(4), 231–258.
*Lee, C. Y., Chen, M. J., & Chang, W. L. (2014). Effects of the multiple solutions and question prompts on generalization and justification for non-routine mathematical problem solving in a computer game context. *Eurasia Journal of Mathematics, Science and Technology Education*, *10*(2), 89–99.
*Lee, K. (2016). Students' proof schemes for mathematical proving and disproving of propositions. *The Journal of Mathematical Behavior*, *41*, 26–44.
*Lin, F. L., Yang, K. L., & Chen, C. Y. (2004). The features and relationships of reasoning, proving and understanding proof in number patterns. *International Journal of Science and Mathematics Education*, *2*(2), 227–256.
*Liu, Y., & Manouchehri, A. (2013). Middle school children's mathematical reasoning and proving schemes. *Investigations in Mathematics Learning*, *6*(1), 18–40.
*Magajna, Z. (2013). Overcoming the obstacle of poor knowledge in proving geometry tasks. *Center for Educational Policy Studies Journal*, *3*(4), 99–116.
*Martinez, M. V. (2014). Algebra and proof in high school: The case of algebraic proof as discovery. *J Res Math Educ*, *3*(1), 30–53.
*Martinez, M. V., Brizuela, B. M., & Castro Superfine, A. (2011). Integrating algebra and proof in high school mathematics: An exploratory study. *The Journal of Mathematical Behavior*, *30*(1), 30–47.
*Martinez, M. V., & Pedemonte, B. (2014). Relationship between inductive arithmetic argumentation and deductive algebraic proof. *Educational Studies in Mathematics*, *86*(1), 125–149.
*Matteson, S., Capraro, M. M., Capraro, R. M., & Lincoln, Y. S. (2012). Extricating justification scheme theory in middle school mathematical problem solving. *Investigations in Mathematics Learning*, *5*(1), 38–62.
*Mueller, M., & Maher, C. (2009). Learning to reason in an informal math after-school program. *Mathematics Education Research Journal*, *21*(3), 7–35.
*Mueller, M., & Yankelewitz, D. (2014). Fallacious argumentation in student reasoning: Are there benefits? *European Journal of Science and Mathematics Education*, *2*(1), 27–38.
*Mueller, M., Yankelewitz, D., & Maher, C. (2012). A framework for analyzing the collaborative construction of arguments and its interplay with agency. *Educational Studies in Mathematics*, *80*(3), 369–387.
*Mueller, M. F. (2009). The co-construction of arguments by middle-school students. *The Journal of Mathematical Behavior*, *28*(2–3), 138–149.
National Council of Teachers of Mathematics. (2000). *Principles and standards for school mathematics*. Reston, VA: NCTM.
*Palla, M., Potari, D., & Spyrou, P. (2012). Secondary school students' understanding of mathematical induction: Structural characteristics and the process of proof construction. *International Journal of Science and Mathematics Education*, *10*(5), 1023–1045.
*Paneque, J. J., Cobo, P., & Fortuny, J. M. (2017). Intelligent tutoring and the development of argumentative competence. *Technology, Knowledge and Learning*, *22*(1), 83–104.
*Patkin, D. (2012). High school students' perceptions of geometrical proofs proving and refuting geometrical claims of the 'for every . . . ' and 'there exists' type. *International Journal of Mathematical Education in Science and Technology*, *43*(8), 985–998.
*Rojas-Drummond, S. M., & Peón, M. (2004). Exploratory talk, argumentation and reasoning in Mexican primary school children. *Language and Education*, *18*(6), 539–557.
Schoenfeld, A. H. (2009). The soul of mathematics. In D. A. Stylianou, M. L. Blanton, & E. J. Knuth (Eds.), *Teaching and learning proof across the grades: A K-16 perspective* (pp. xii–xvi). New York, NY: Routledge.
*Sen, C., & Guler, G. (2015). Examination of secondary school seventh graders' proof skills and proof schemes. *Universal Journal of Educational Research*, *3*(9), 617–631.
*Soldano, C., & Arzarello, F. (2016). Learning with touchscreen devices: Game strategies to improve geometric thinking. *Mathematics Education Research Journal*, *28*(1), 9–30.
Staples, M., Newton, J., Kosko, K., Cirillo, M., Weber, K., Bieda, K., & Conner, A. (2016). Conceptions and consequences of what we call argumentation, justification, and proof. In M. Wood, E. Turner, M. Civil, & J. Eli (Eds.), *Proceedings of the 38th conference of the North-American chapter of the*

international group for the psychology of mathematics education (pp. 1704–1712). Tuscon: PME-NA.
*Stout, T., Dwyer, J., Byerly, R., & Wilhelm, J. (2011). Assessing the learning of proofs in high school. *International Journal for Mathematics Teaching and Learning*, *18*. Retrieved from http://www.cimt.org.uk/journal/dwyer3.pdf
*Stoyle, K. L., & Morris, B. J. (2017). Blogging mathematics: Using technology to support mathematical explanations for learning fractions. *Computers & Education*, *111*, 114–127.
*Strom, D., Kemeny, V., Lehrer, R., & Forman, E. (2001). Visualizing the emergent structure of children's mathematical argument. *Cognitive Science*, *25*(5), 733–773.
*Stylianides, A. J. (2016). *Proving in the elementary classroom*. Oxford: Oxford University Press.
Stylianides, A. J., Bieda, K. N., & Morselli, F. (2016). Proof and argumentation in mathematics education research. In A. Gutiérrez, G. C. Leder, & P. Boero (Eds.), *The second handbook of research on the psychology of mathematics education* (pp. 315–351). Rotterdam: Sense Publishers.
Stylianides, G. J., Stylianides, A. J., & Weber, K. (2017). Research on the teaching and learning of proof: Taking stock and moving forward. In J. Cai (Ed.), *Compendium for research in mathematics education* (pp. 237–266). Reston, VA: National Council of Teachers of Mathematics.
*Tsamir, P., & Sheffer, R. (2000). Concrete and formal arguments: The case of division by zero. *Mathematics Education Research Journal*, *12*(2), 92–106.
*Weber, K., Maher, C., Powell, A., & Lee, H. S. (2008). Learning opportunities from group discussions: Warrants become the objects of debate. *Educational Studies in Mathematics*, *68*(3), 247–261.
*Whitenack, J. W., & Knipping, N. (2002). Argumentation, instructional design theory and students' mathematical learning: A case for coordinating interpretive lenses. *The Journal of Mathematical Behavior*, *21*, 441–457.
*Yang, K. L., & Lin, F. L. (2012). Effects of reading-oriented tasks on students' reading comprehension of geometry proof. *Mathematics Education Research Journal*, *24*(2), 215–238.
*Yim, J., Song, S., & Kim, J. (2008). Mathematically gifted elementary students' revisiting of Euler's polyhedron theorem. *The Mathematics Enthusiast*, *5*(1), 125–142.
*Yopp, D. A. (2017). Eliminating counterexamples: A grade 8 student's learning trajectory for contrapositive proving. *The Journal of Mathematical Behavior*, *45*, 150–166.

Appendix

Article (in alphabetical order by title)
Lavy, I. (2006). A case study of different types of arguments emerging from explorations in an interactive computerized environment. *The Journal of Mathematical Behavior, 25*(2), 153–169.
Mueller, M., Yankelewitz, D., & Maher, C. (2012). A framework for analyzing the collaborative construction of arguments and its interplay with agency. *Educational Studies in Mathematics, 80*(3), 369–387.
Komatsu, K. (2016). A framework for proofs and refutations in school mathematics: Increasing content by deductive guessing. *Educational Studies in Mathematics, 92*(2), 147–162.
Healy, L., & Hoyles, C. (2000). *A study of proof conceptions in algebra. Journal for Research in Mathematics Education, 31*, 396–428.
Martinez, M. V. (2014). Algebra and proof in high school: The case of algebraic proof as discovery. *Journal of Research in Mathematics Education, 3*(1), 30–53.
Stylianou, D. A. (2013). An examination of connections in mathematical processes in students' problem solving: Connections between representing and justifying. *Journal of Education and Learning, 2*(2), 23.
Fiallo, J., & Gutiérrez, A. (2017). Analysis of the cognitive unity or rupture between conjecture and proof when learning to prove on a grade 10 trigonometry course. *Educational Studies in Mathematics, 96*(2), 145–167.
Bieda, K. N., & Lepak, J. (2014). Are you convinced? Middle-grade students' evaluations of mathematical arguments. *School Science and Mathematics, 114(*4), 166–177.
Whitenack, J. W., & Knipping, N. (2002). Argumentation, instructional design theory and students' mathematical learning: a case for coordinating interpretive lenses. *The Journal of Mathematical Behavior, 21*(4), 441–457.
Stout, T., Dwyer, J., Byerly, R., & Wilhelm, J. (2011). Assessing the learning of proofs in high school. International Journal for Mathematics Teaching & Learning.
Stoyle, K. L., & Morris, B. J. (2017). Blogging mathematics: Using technology to support mathematical explanations for learning fractions. *Computers & Education, 111*, 114–127.
Cáceres, M., Nussbaum, M., Marroquín, M., Gleisner, S., & Marquínez, J. T. (2018). Building arguments: key to collaborative scaffolding. *Interactive Learning Environments, 26*(3), 355–371.
Evens, H., & Houssart, J. (2004). Categorizing pupils' written answers to a mathematics test question: 'I know but I can't explain.' *Educational Research, 46*, 269–282.
Tsamir, P., & Sheffer, R. (2000). Concrete and formal arguments: The case of division by zero. *Mathematics Education Research Journal, 12*(2), 92–106.
Ellis, A. B. (2007). Connections between generalizing and justifying: Students' reasoning with linear relationships. *Journal for Research in Mathematics Education, 38*, 194–229.
Komatsu, K. (2010). Counter-examples for refinement of conjectures and proofs in primary school mathematics. *The Journal of Mathematical Behavior, 29*(1), 1–10.
Cross, D. I. (2009). Creating optimal mathematics learning environments: Combining argumentation and writing to enhance achievement. *International Journal of Science and Mathematics Education, 7*(5), 905–930.
Yang, K. L., & Lin, F. L. (2012). Effects of reading-oriented tasks on students' reading comprehension of geometry proof. *Mathematics education research journal, 24*(2), 215–238.
Lee, C. Y., Chen, M. J., & Chang, W. L. (2014). Effects of the multiple solutions and question prompts on generalization and justification for non-routine mathematical problem solving in a computer game context. *Eurasia Journal of Mathematics, Science & Technology Education, 10*(2), 89–99.
Yopp, D. A. (2017). Eliminating counterexamples: A Grade 8 student's learning trajectory for contrapositive proving. *The Journal of Mathematical Behavior, 45*, 150–166.
Pedemonte, B., & Balacheff, N. (2016). Establishing links between conceptions, argumentation and proof through the ck¢-enriched Toulmin model. *The Journal Of Mathematical Behavior, 41*, 104–122.
Sen, C., & Guler, G. (2015). Examination of Secondary School Seventh Graders' Proof Skills and Proof Schemes. *Universal Journal of Educational Research, 3*(9), 617–631.
Komatsu, K., Fujita, T., & Jones, K. (2018). Explanatory unification by proofs in school mathematics. *For the Learning of Mathematics, 38*(1), 31–37.
Rojas-Drummond, S., & Peón, M. (2004). Exploratory talk, argumentation and reasoning in Mexican primary school children. *Language and Education, 18*(6): 539–557.
Brown, R. (2007). Exploring the social positions that students construct within a classroom community of practice. *International Journal Of Educational Research, 46*(3), 116–128.
Matteson, S., Capraro, M. M., Capraro, R. M., & Lincoln, Y. S. (2012). Extricating justification scheme theory in middle school mathematical problem solving. *Investigations in Mathematics Learning*, 5(1), 38–62.
Mueller, M., & Yankelewitz, D. (2014). Fallacious argumentation in student reasoning: Are there benefits?. *European Journal of Science and Mathematics Education, 2*(1), 27–38.
Lannin, J. K. (2005). Generalization and justification: The challenge of introducing algebraic reasoning through patterning activities. *Mathematical Thinking and Learning, 7*(3), 231–258.

(*continued*).

Continued.

Article (in alphabetical order by title)
Sriraman, B. (2004). Gifted ninth graders' notions of proof: Investigating parallels in approaches of mathematically gifted students and professional mathematicians. *Journal for the Education of the Gifted, 27*(4), 267–292.
Dorit Patkin (2012) High school students' perceptions of geometrical proofs proving and refuting geometrical claims of the 'for every . . . ' and 'there exists' type. *International Journal of Mathematical Education in Science and Technology, 43*(8), 985–998.
Flores, A. (2006). How do students know what they learn in middle school mathematics is true? *School science and mathematics, 106*(3), 124–132.
Kuchemann, D., & Hoyles, C. (2006). Influences on Students' Mathematical Reasoning and Patterns in Its Development: Insights from a Longitudinal Study with Particular Reference to Geometry. *International Journal of Science And Mathematics Education, 4*(4), 581–608.
Aaron, W. R., & Herbst, P. (2012). Instructional identities of geometry students. *The Journal of Mathematical Behavior, 31*(3), 382–400.
Martinez, M. V., Brizuela, B. M., & Superfine, A. C. (2011). Integrating Algebra and Proof in High School Mathematics: An Exploratory Study. *Journal of Mathematical Behavior, 30*(1), 30–47.
Paneque, J. J., Cobo, P., & Fortuny, J. M. (2017). Intelligent tutoring and the development of argumentative competence. *Technology, Knowledge And Learning, 22*(1), 83–104.
Mariotti, M. A. (2000). Introduction to proof: The mediation of a dynamic software environment. *Educational Studies in Mathematics, 44*(1–2), 25–53.
Chinnappan, M., Ekanayake, M. B., & Brown, C. (2012). Knowledge use in the construction of geometry proof by Sri Lankan students. *International Journal of Science and Mathematics Education, 10*(4), 865–887.
Weber, K., Maher, C., Powell, A., & Lee, H. S. (2008). Learning opportunities from group discussions: Warrants become the objects of debate. *Educational Studies in Mathematics, 68*(3), 247–261.
Mueller, M., & Maher, C. (2009). Learning to reason in an informal math after-school program. *Mathematics Education Research Journal, 21*(3), 7–35.
Soldano, C., & Arzarello, F. (2016). Learning with touchscreen devices: game strategies to improve geometric thinking. *Mathematics Education Research Journal, 28*(1), 9–30.
Georgios Kosyvas (2016) Levels of arithmetic reasoning in solving an open- ended problem, International *Journal of Mathematical Education in Science and Technology,* 47(3), 356–372.
Kosko, K. W. (2016). Making use of what's given: Children's detailing in mathematical argumentative writing. *The Journal of Mathematical Behavior, 41*, 68–86.
Ugurel, I., Morali, H. S., Karahan, O., & Boz, B. (2016). Mathematically gifted high school students' approaches to developing visual proofs (VP) and preliminary ideas about VP. *International Journal of Education In Mathematics, Science And Technology, 4*(3), 174–197.
Yim, J., Song, S., & Kim, J. (2008). Mathematically gifted elementary students' revisiting of Euler's polyhedron theorem. *The Mathematics Enthusiast, 5*(1), 125–142.
Olivero, F., & Robutti, O. (2007). Measuring in dynamic geometry environments as a tool for conjecturing and proving. *International Journal of Computers for Mathematical Learning, 12*(2), 135–156.
Rubel, L. H. (2007). Middle School and High School Students' Probabilistic Reasoning on Coin Tasks. *Journal for Research In Mathematics Education, 38*(5), 531–557.
Liu, Y., & Manouchehri, A. (2013). Middle school children's mathematical reasoning and proving schemes. *Investigations in Mathematics Learning, 6*(1), 18–40.
Haja, S., & Clarke, D. (2011). Middle school students' responses to two-tier tasks. *Mathematics Education Research Journal, 23*(1), 67–76.
Magajna, Z. (2013). Overcoming the obstacle of poor knowledge in proving geometry tasks. *Center for Educational Policy Studies Journal, 3*(4), 99–116.
Marta Civil & Roberta Hunter (2015) Participation of non-dominant students in argumentation in the mathematics classroom. *Intercultural Education, 26*(4), 296–312.
Komatsu, K., Jones, K., Ikeda, T., & Narazaki, A. (2017). Proof validation and modification in secondary school geometry. *The Journal of Mathematical Behavior, 47*, 1–15.
Marrades, R., & Gutiérrez, Á. (2000). Proofs produced by secondary school students learning geometry in a dynamic computer environment. *Educational Studies in Mathematics, 44*(1–2), 87–125.
Jones, K. (2000). Providing a foundation for deductive reasoning: Students' interpretations when using dynamic geometry software and their evolving mathematical explanations. *Educational studies in mathematics, 44*(1–2), 55–85.
Herbst, P., & Brach, C. (2006). Proving and doing proofs in high school geometry classes: What is it that is going on for students? *Cognition and Instruction, 24*(1), 73–122.
Almeida, D. (2001). Pupils' proof potential. *International Journal of Mathematical Education in Science and Technology, 32*(1), 53–60.
Martinez, M. V., & Pedemonte, B. (2014). Relationship between inductive arithmetic argumentation and deductive algebraic proof. *Educational Studies in Mathematics, 86*(1), 125–149.

(*continued*).

Continued.

Komatsu, K., Tsujiyama, Y., & Sakamaki, A. (2014). Rethinking the discovery function of proof within the context of proofs and refutations. *International Journal of Mathematical Education in Science and Technology, 45*(7), 1053–1067.

Palla, M., Potari, D., & Spyrou, P. (2012). Secondary school students' understanding of mathematical induction: Structural characteristics and the process of proof construction. *International Journal of Science and Mathematics Education, 10*(5), 1023–1045.

Healy, L., & Hoyles, C. (2002). Software tools for geometrical problem solving: Potentials and pitfalls. *International Journal of Computers for Mathematical Learning, 6*(3), 235–256.

Furinghetti, F., Olivero, F., & Paola, D. (2001). Students approaching proof through conjectures: Snapshots in classrooms, International Journal of Mathematical Education in Science and Technology, 32(3), 319–335.

Barrett, J. E., Clements, D. H., Klanderman, D., Pennisi, S. J., & Polaki, M. V. (2006). Students' coordination of geometric reasoning and measuring strategies on a fixed perimeter task: Developing mathematical understanding of linear measurement. *Journal for Research in Mathematics Education, 37*(3), 187–221.

Miyazaki, M., Fujita, T., & Jones, K. (2017). Students' understanding of the structure of deductive proof. *Educational Studies in Mathematics, 94*(2), 223–239.

Hoyles, C., & Küchemann, D. (2002). Students' understandings of logical implication. *Educational Studies in Mathematics, 51*(3), 193–223.

Lee, K. (2016). Students' proof schemes for mathematical proving and disproving of propositions. *The Journal of Mathematical Behavior, 41*, 26–44.

Hanna, G., Jahnke, H. N., DeBruyn, Y., & Lomas, D. (2001). Teaching mathematical proofs that rely on ideas from physics. *Canadian Journal of Math, Science & Technology Education, 1*(2), 183–192.

Mueller, M. F. (2009). The co-construction of arguments by middle-school students. *The Journal of Mathematical Behavior, 28*(2–3), 138–149.

Fried, M. N., & Amit, M. (2008). The co-development and interrelation of proof and authority: The case of Yana and Ronit. *Mathematics Education Research Journal, 20*(3), 54–77.

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

Lin, F. L., Yang, K. L., & Chen, C. Y. (2004). The features and relationships of reasoning, proving and understanding proof in number patterns. *International Journal of Science and Mathematics Education, 2*(2), 227–256.

Lachmy, R., & Koichu, B. (2014). The interplay of empirical and deductive reasoning in proving 'if' and 'only if' statements in a dynamic geometry environment. *The Journal of Mathematical Behavior, 36*, 150–165.

Stylianides, A. J. (2007). The notion of proof in the context of elementary school mathematics. *Educational Studies in Mathematics, 65*, 1–20.

Krummheuer, G. (2013). The relationship between diagrammatic argumentation and narrative argumentation in the context of the development of mathematical thinking in the early years. *Educational Studies in Mathematics, 84*(2), 249–265.

Hadas, N., Hershkowitz, R., & Schwarz, B. B. (2000). The role of contradiction and uncertainty in promoting the need to prove in dynamic geometry environments. *Educational Studies in Mathematics, 44*(1–2), 127–150.

Leung, A., & Lopez-Real, F. (2002). Theorem justification and acquisition in dynamic geometry: A case of proof by contradiction. *International Journal of Computers for Mathematical Learning, 7*(2), 145–165.

Wong, W. K., Yin, S. K., Yang, H. H., & Cheng, Y. H. (2011). Using computer-assisted multiple representations in learning geometry proofs. *Journal of Educational Technology & Society, 14*(3).

Strom, D., Kemeny, V., Lehrer, R., & Forman, E. (2001). Visualizing the emergent structure of children's mathematical argument. *Cognitive Science, 25*(5), 733–773.