Proof in the Context of Middle Grades: Can We Label Middle School Arguments as Proof with a Capital P?

David A. Yopp, Rob Ely, Anne E. Adams, and Annelise W. Nielsen

Introduction

The concept of *proof* is central to the field of mathematics, but defining the term has challenged educators and researchers (see Czocher & Weber, 2020; Weber, 2014, for a discussion). In theory, the term seems easy to define. A *proof* is a sequence of statements involving logic and prior results. These statements are arranged in a logical order to support or refute a quantified claim. *Prior results* include axioms, definitions, and previously established results. The logic is mathematical logic, which is a metatheory about what types of logical inferences can be made and what methods and modes of reasoning (e.g., modus ponens and modus tollens) are acceptable. This metatheory also establishes what types of statements can be made (e.g., quantified statements that are either true or false), which statements are logically equivalent, and the norms for representing generality and inferences. This metatheory provides the "rules of the game" for mathematical proof and proving.

The challenge in using this definition is that the metatheory is tough to unpack in school mathematics and it does not account for all the types of arguments some mathematicians have accepted as proofs (Czocher & Weber, 2020). Moreover, it

D. A. Yopp (✉)
Departments of Mathematics and Curriculum and Instruction, University of Idaho, Moscow, ID, USA
e-mail: dyopp@uidaho.edu

R. Ely · A. W. Nielsen
Department of Mathematics, University of Idaho, Moscow, ID, USA
e-mail: ely@uidaho.edu; annelise@uidaho.edu

A. E. Adams
Department of Curriculum and Instruction, University of Idaho, Moscow, ID, USA
e-mail: aeadams@uidaho.edu

seems impractical and, arguably, imprudent to teach middle school students all the rules that mathematicians have learned to follow when producing a proof. Teachers pursue many instructional goals in the middle school mathematics—content, skills, and other mathematical practices—and have limited amounts of time to pursue these goals. Yet, many educators, researchers, and policy makers ask that middle-grade mathematics students produce arguments that can be taken as proofs, have elements of proof, or at least provide pathways toward proofs. In the USA, Common Core State Standards for Mathematics (National Governors Association Center for Best Practices & Council of Chief State School Officers, 2010) recommend that students use "stated assumptions, definitions, and previously established results in constructing arguments" (p. 3) and "build logical progression[s] of statements to explore the truth of their conjectures" (p. 3). Similar recommendations have motivated researchers to search students' mathematical arguments for elements of proof amidst the various types of reasoning students use when they are asked to prove. Researchers have noted proof-like behaviors among students who responded to proving tasks, such as attending to the general case and using definitions and prior results, and have sometimes likened these behaviors to the practices of trained mathematicians. Stylianides (2007), for example, found proof-like practices among elementary students' responses and likened them to the proof practices of mathematicians. Yet, noting "proof- and proving-like" behaviors among students does not necessarily mean that students have an understanding of a metatheory for mathematical proof and proving. In other words, the students who produce arguments akin to proofs may not understand why their reasoning is valid nor understand how their reasoning fits into a valid proof and proving scheme.

Students' lack of a proof metatheory is perhaps one reason the editors of this book asked us to use a broader definition of proof that more readily adapts to the reasoning of students in the middle grades. In this definition, which we will call "our working definition," a mathematical proof has all or most of the following characteristics, and proving is an activity that leads to such a product:

1. A proof is a convincing argument that convinces a knowledgeable mathematician that a claim is true.
2. A proof is a deductive argument that does not admit possible rebuttals.
3. A proof is a transparent argument *where a mathematician can fill in every gap* (given sufficient time and motivation), perhaps to the level of being a formal derivation.
4. A proof is a *perspicuous argument that provides the reader with an understanding of why a theorem is true*.
5. A proof is an *argument within a representation system satisfying communal norms.*
6. A proof is an *argument that has been sanctioned by the mathematical community* (Weber, 2014, p. 537).

In this book chapter, we use a version of this definition to analyze a collection of middle-grade student work presented to us by the editors to determine the degree to which the students demonstrated a level of proficiency with proof and proving as

described by this definition. We also consider the teachers' contributions during activities related to proof, proving, and justification from the classroom episode from which this student work was collected. Through this lens we aim to discuss whether or not students in the class exhibited characteristics of the above definition in their arguments and justification practices as well as whether or not these students were exposed these "rules of the game" for proof and proving in mathematics. Consequently, we discuss possible mismatches between this definition and the guidelines students were given for developing acceptable justification.

Ultimately, we will consider the possibilities for proof and proving in the middle grades and the implications of exposing students to the same definition of proof and proving that is used to analyze their work. In other words, we ask, "What can happen when students are presented with the same rules of the game for proof and proving in mathematics as those used to analyze their data?" But, first, we consider this above definition in the context of middle school mathematics and discuss factors to consider when adapting this definition into a rubric used to examine middle-grade students' proving activities.

Reframing the Definition of Proof for the Current Context

As described in the chapter "Overview of Middle Grades Data" (this volume), the current context involved a small seventh-grade class of 19 students in New England (USA) who engaged in the Number Trick task. This class was taught by a teacher who actively worked to engage students in argument and justification as part of the Justification and Argumentation: Growing Understanding of Algebraic Reasoning Project (the JAGUAR Project, funded by NSF, DRL 0814829). More will be said in the methods section about the materials the students received for understanding characteristics of justifications and argumentation, and later we will discuss how well these resources aligned with our working definition of mathematical proof. For now, we will reflect on the definition of mathematical proof we were provided and aspects we considered when adapting it to be an analytic tool.

In our opinion, the six characteristics in our working definition of mathematical proof attend to purposes and norms for proving in the community of mathematicians, and consequently, the definition can be difficult to apply to the work of school-age children, particularly when we do not know whether the children received explicit training in the norms and practices expressed in the definition or frequent experiences with these norms and purposes. Applying the definition also relies on the judgment and imagination of the reader, which is a source of subjectivity. Nonetheless, we believe it to be valuable to adapt the criteria in the definition to our current context for the sake of understanding the types of reasoning students used in their arguments and justifications and how their reasoning can be developed to better fit with the norms and practices of the broader mathematical community.

One characteristic that was particularly challenging to apply to our context— middle-grade students engaged in the Number Trick task—was the *convincingness*

characteristic. A knowledgeable mathematician or mathematics educator would already be convinced that Jessie's two computational approaches equate prior to reading a student's argument because Jessie's approaches can be equated using the distributive property. Thus, to ascertain convincingness, the analyst must imagine, "Were I unfamiliar with this claim, and the distributive property, would this argument convince me of its truth?" We wonder how reliable an analyst's image of a convincing argument can be in such a hypothetical situation.

Another source of subjectivity when applying the convincingness characteristic is "who gets to decide?" Who exactly are these "knowledgeable mathematicians," and how can we assume they would all agree? Can mathematics teachers of various mathematical backgrounds count as knowledgeable mathematicians? This question is particularly troubling if we view proof and proving as a means for students and teachers to take ownership of and authority over their mathematical knowledge and learning. Relying on the authority of more knowledgeable others to assess the teachers' and students' arguments seems to defeat this purpose of proof and proving.

The *representation system* and *community-sanctioned* characteristics (5 and 6) presented similar challenges. The *community-sanctioned* characteristic applies readily to classic proofs like those of the Pythagorean theorem and those showing $\sqrt{2}$ is irrational, but no such prototypical proofs exist for claims associated with the Number Trick task, such as "For all real numbers, the two computational approaches in the Number Trick task produce the same result." Therefore, it was unclear how to compare students' arguments in this context to arguments sanctioned by the mathematical community.

The *representation system* characteristic (5) presented similar challenges. This characteristic refers to norms for representing arguments, but whose norms are to be met? The mathematical norm for representing algebraic proofs is to use variables to express the general case. However, Yopp and Ely (2016) argued that noncanonical representations such as generic example referents can be leveraged in proofs provided that certain criteria were met. One critical criterion is that the example (a.k.a. referent) is not appealed to in any way that is specific to the example chosen. Here, specific means a trait not shared by all cases in the domain of the claim. To us, it is more important to assess how representations are leveraged in the argument than it is to assess whether or not the representations are canonical. But this view can be in tension with the social role of proof and proving. In other words, our view acknowledges that a student might produce a proof that her/his peers would not sanction as proof but a reader with an open mind about the power of alterative representation systems would sanction as proof.

Indeed, all six of the characteristics in the definition of proof depend on socially accepted norms and meanings. Even the characteristics that seem most objective such as deductive inference, *proof* characteristic (2), are subjective when applied to middle school students' work. Deduction is a well-defined method of reasoning where the arguer identifies a rule $p \rightarrow q$ and a case of p and concludes that because of the rule, the case of p also has the property (or properties) q. But in practice, we rarely find middle school reasoning expressed in this form. Instead, we, the authors, have found ourselves searching student arguments for deductive-like reasoning, as

we envision it, without any knowledge of whether or not the students who produced the arguments were conscious of the deductive inferences being made and the logical necessities such inferences produced. Are we sometimes using wishful thinking when we attribute deductive reasoning?

In fact, a student who appears to be reasoning deductively may be appealing to modes of reasoning that arise outside of mathematics courses. Various psychology frameworks describe "naturally occurring" and spontaneous modes of reasoning that can mimic deduction but are distinct from deduction (e.g., mental models, Johnson-Laird, 1983, and pragmatic reasoning schemes, Cheng & Holyoak, 1985). Even Harel and Sowder (1998) ground their analytic and deductive reasoning schemes not just in deductive inferences but in schema-based transformations of mathematical objects. We assert that it would be quite difficult to know whether or not deductive reasoning is truly present in any piece of student work unless the student is explicit about this mode of thought.

Another subjective aspect of applying the *deductive* characteristic is judging whether the general rules being leveraged by students are taken as general prior knowledge in the classroom community. Stylianides (2007) and Stylianides and Stylianides (2008) proposed, as part of a definition of proof for school mathematics, that "[proof in school mathematics] uses statements accepted by the classroom community (set of accepted statements) that are true and available without further justification" (Stylianides & Stylianides, 2008, p. 107). Although this criterion specifies the community, it also requires that the reader assessing the argument is knowledgeable in the background of that classroom community.

A final point we wish to make about the *deductive* characteristic is that sometimes the characteristic does not apply at all to a proof. For instance, a proof by exhaustion might contain no deductive reasoning. In some of the proofs we analyzed in this chapter, the students simply checked that the claim was true for all ten possible cases. This is a completely valid method of proving that uses no deductive reasoning.

Moreover, a proof by exhaustion also might have no explanatory power, which would also prevent us from applying the *explains why* characteristic (4) when evaluating it. Another factor when applying this proof characteristic (to any argument, not just to an exhaustion argument) is that a proof's explanatory power depends upon the reader. For instance, Hanna (1990) asserts that mathematical induction may not have this explanatory trait, but Stylianides, Sandefur, and Watson (2016) point out that many mathematical induction arguments do have this trait and can be distinguished from those that do not. Perhaps what we mean is that explanatory proofs! demonstrate why the defining properties of the mathematical objects specified in the conditions of the claim *must* also have the properties specified in the conclusion of the claim. If so, the *explains why* characteristic may be better described in terms of conceptual insights (Sandefur et al., 2013), which provide structural links between the conditions and conclusions.

In the context of the Number Trick task, an example of a conceptual insight that provides such a structural link is the distributive property, which transforms one expression into another equivalent expression. Yet students might see other

structures in the expressions that equate the approaches' outputs. The need for "balance" by adding twice as much to the doubled value in the case where doubling comes first is one such structural link that does not explicitly employ the distributive property. Moreover, what if it is clear that a student is searching for a conceptual insight but does not find one in the time allocated? This could be evidence that the student is aware that a proof of a general claim needs more than just an empirical check but that the student got "stuck" and was unable to find an insight that shows why the conditions imply the conclusion. We consider any search for a conceptual insight as evidence of deeper understanding of proof than expressed in an empirical argument.

Although our discussion so far has pointed out ambiguities and potential problems with assessing these six characteristics of proof, we wrap up with a more optimistic look at one of these, the *fillable gaps* characteristic (3). This characteristic turned out to be one of the most useful of the six and perhaps truest to the analytic methods we typically use. Inevitably, when analyzing middle-grade students' arguments, including arguments we collected in other projects, our group's discussions of these arguments turn to our own knowledge and practices. To make sense of the students' work, we often spend a great deal of time constructing our own proofs from the sparks of insights found in students' work. We find merit in novel approaches, and we discover proof paths that we overlooked. Of course, this type of analysis also presents challenges. Reconstructions, gap-fillings, and extensions may project reasoning onto student responses that the students did not intend. Perhaps we might also fail to value certain insights because we cannot see the proof path, even when one is present.

Methods

Our critical discussion in the introduction was not meant to dismiss our task as impossible. Instead, our critical discussion was an attempt to make sense of what we could assess and how we could assess it and to determine what we could not assess. From our ponderings, we developed a rubric (Fig. 1) for assessing the student data that maintains, in slightly altered form, the six characteristics provided to us and elaborates on them in the context of the Number Trick task.

The rubric applies only to arguments that attempt to address a general case. It does not apply to exhaustive proofs. We found that we needed to create two scores for each student because the Number Trick task had two prompts: whether the Number Trick worked for natural numbers from 1 to 10 and whether the Number Trick worked for all real numbers. For the first prompt, the second and fourth characteristics in Levels 1 and 2 had to be modified to accommodate proof by exhausting all cases, which may contain little or no explicit deduction and may provide no explanatory power.

We also wish to note that while our rubric articulates criteria for every characteristic at every level, the notion of cluster category does not mean that every argument

2	If we were to set our conviction aside, then the argument would <u>convince</u> us that the general claim, "for all real numbers, the two approaches produce the same result," is true.
	The argument illustrates <u>deductive-like</u> reasoning, leveraging what are assumed to be prior results to make inferences that we assume are logical necessities.
	The argument <u>needs little or no gap filling</u>, meaning all the important inferences needed are provided.
	The argument <u>provides a conceptual insight showing why</u> for every real number, the two approaches produce the same result.
	We were <u>able to understand the representations</u> used, see how these representations illustrate the two approaches, and see how the transformations on the representations illustrate how the structure of the approaches necessitate the conclusion that the approaches produce the same values for any number in the domain of the claim.
	We were <u>able to sanction</u> the argument as proof, based on our experiences as mathematicians.
1	If we were to set our conviction aside, then the argument would <u>not persuade</u> us that the general claim, "for all real numbers, the two approaches produce the same result", is true. The argument leaves room for doubt.
	The argument illustrates <u>a search for a conceptual insight</u> that could lead to <u>deductive-like</u> reasoning, yet the prior results and inferences made are vague.
	The argument <u>needs gap filling</u>, meaning some important inferences must be filled in by the reader.
	The argument indicates a search for a conceptual insight that could show a structural reason why for every real number, the two approaches produce the same result, even if the structural reason was not found or is unclear.
	We made <u>significant assumptions about</u> how the representations were used to illustrate the two approaches <u>and/or</u> how the transformations on the representations illustrate how the structure of the approaches necessitate the conclusion that the approaches produce the same values for any number in the domain of the claim.
	We were <u>not able to sanction</u> the argument as proof based on our experiences as mathematicians.
	Overall: The argument provides evidence that the arguer was aware that either all cases in the domain of the claim must be tested or awareness that a general argument based in mathematical structures of the conditions and the conclusion and awareness that inferential links between these structures must be illustrated.
0	Perhaps the argument provides empirical support for the general claim, perhaps illustrating only a proper subset of the cases in domain of the claim. Nevertheless, the argument fails to provide evidence that the arguer was aware that either all cases in the domain of the claim must be tested or that a general argument based in mathematical structures of the conditions and the conclusion and that inferential links between these structures must be illustrated.

Fig. 1 Proof rubric

that can be labeled as "proof" must satisfy all six characteristics. Weber (2014) explains that such an argument that satisfies all six characteristics is likely to generate wide agreement regarding its status as a proof, while arguments that only satisfy a subset may generate more debate among the community of mathematicians as to whether or not it suffices as a proof. However, because one of our goals is to help

translate the characteristics presented in Weber's (2014) cluster concept of proof for a middle-grade context, we included revised descriptions for each of the six characteristics at each level of proof.

Our first pass through the student work included a search for (1) empiricism, which tests only a subset of cases to which the claim applies, (2) conceptual insights, and (3) searches for conceptual insights. This pass helped us understand the nature of students' proving processes and prepared us to apply the rubric. We asked ourselves questions such as:

1. What type of claim is being made, and what is the claim's domain?
2. What prior results and knowledge does the student leverage to support their claim?
3. Are these results and knowledge leveraged in a manner consistent with what was assumed to be taught or with what is canonical to mathematics?
4. Are the inferences, implicit or explicit, logical necessities?

Ultimately, we used the rubric more as a guide than a checklist, as it was difficult to assess every characteristic in any piece of student work because the students were often vague and perhaps unfamiliar with ways of expressing their mathematical thinking. Although a score of 2 required that most or all of the characteristics were met and the four questions above were answered in the affirmative, a score of 1 was awarded if there was compelling evidence that the student searched for a conceptual insight, as noted in the overall summary of a score of 1 in the rubric. To us, all six characteristics are implicit in the expression of a conceptual insight explaining why the two approaches must produce the same outcome.

Each author reviewed the student work separately and scored it using the rubric. The scores were compared and discussed until the authors reached agreement on the scores.

We also reviewed the materials associated with the lesson from which the student work was generated. This analysis also involved our rubric and included transcripts of the teaching episode and other materials provided to the students, including a description of "what makes a good justification." Our review of the transcript was performed for two purposes. One purpose was to triangulate our analysis of the student work with any discussion of the student work, particularly by the student. Here, we looked for any comments about the students' work that might shed light on meaning in the student work that might have been overlooked. The other purpose was to better understand the context in which the student work was constructed and presented, including attending to Mr. MC's prompts and contributions that related to the six elements of proof as articulated in our rubric. This later analysis provided us with an opportunity to consider consistencies and possible mismatches between the teacher's justification and argument goals and our analysis scheme based on the definition of mathematical proof we were provided.

Findings

Classroom Episode

Based on Mr. MC's activities during the lesson and the materials provided to the students, we found that the teacher valued and at times sanctioned student activities that showed promise for developing a proof, as defined by our working definition of mathematical proof and as described in our rubric. We found numerous occasions throughout the transcript where Mr. MC asked students to consider whether or not the students' examples or reasoning demonstrated that Jessie's Number Trick works for all cases or every case. Such teacher moves are likely to encourage *convincing arguments* and *arguments that do not admit rebuttals*. These requests were given in both the finite and infinite-domain contexts, when students were addressing Jessie's Number Trick for natural numbers 1 through 10, and again when students were addressing Jessie's Number Trick for any number, which we assumed to be any natural number.

We also found numerous occasions throughout the transcript where Mr. MC asked students to explain their thinking, explain why their claim about Jessie's Number Trick is true, and explain how they knew it worked for all cases. These requests were also made in both contexts, when students were addressing the finite-domain prompt and when students were addressing the infinite-domain prompt. We found it interesting that the terms "explain" and "explanations" were used to encourage students to be more explicit about their reasoning in both these contexts. In lines 134–135, when students are addressing the finite-domain prompt, Mr. MC tells students, "You're going to have to explain how you think… why you think it true," and in lines 239–240, Mr. MC refocuses students with the prompt, "Explain your reasoning." Both of these statements were made during exchanges in which Mr. MC was also encouraging students to test every case from 1 to 10. This observation is not to criticize Mr. MC's uses of these phrases but to point out that the use of the expressions "show why" and "explain why" in the classroom could lead to a mismatch between students' notions of showing why and notions of "a perspicuous argument…[offering] an understanding of why a theorem is true," (Weber, 2014, p. 537), as well as the notion of conceptual insights as described in our rubric. Exhaustive arguments may be included in the classroom community's concept image of arguments that explain or show why a theorem/claim is true, while our use of the term refers to arguments that express conceptual insights, which typically do not include exhaustive arguments. We wonder if "showing [or explaining] why" was a catch-all phrase for being explicit about your thinking and argument approaches or why you believe your argument is viable as opposed to its intended meaning in the working definition of mathematical proof and in our rubric. This lack of clarity in the terms "explain why" and "show why" could lead to confusion among students as to what constitutes viable arguments and proofs in mathematics class.

Having acknowledged this possible mismatch between our use of these phrases and the classroom community's use of these phrases, we did find evidence that Mr.

MC valued and encouraged arguments that expressed conceptual insights often by revoicing students' arguments. The exchange below was representative of these teacher activities when encouraging searches for conceptual insights:

S4: When you add these two numbers together, it's going to be higher than when you… first and then you add it: since eight is more than four, you need to have a higher number.

S6: They come in doubles… The second number that you add needs to be higher, so they need to be the same [inaudible].

T: So, you're saying that here you're multiplying the number by something before you add something to it. Up here, you're adding something and then multiplying. So what you are multiplying is going to be bigger here… So, that's some great thinking in terms of the general case….

In exchanges like this, we see the teacher focus students on structural aspects of the situation (here, the order in which multiplication and addition occur and how that affects the size of the outcome) that could lead students to understand why 8 must be added instead of 4 when the input number is doubled first. We interpreted this type of teacher move as encouraging and valuing conceptual insights, as well as encouraging perspicuous arguments that show why.

Yet, we found little evidence that Mr. MC encouraged or explicitly described the other properties of a proof as described in our rubric and working definition. Most of the episode involved discussion of student approaches both in small-group and whole-class settings. There was little emphasis on how students represented their ideas in their written work and no explicit instruction on what it meant to provide a deductive argument. Mr. MC sanctioned several students' approaches with phrases such as "perfectly valid explanation" and "good logic," but there was no explicit standard for validity and good logic. Finally, the convincingness of an argument was implicit in the teacher's emphasis on why the two approaches would equate for any real number; yet, the standard for convincing others, such as knowledgeable mathematicians, was not mentioned.

The lack of evidence of Mr. MC explicitly discussing these properties of proof raises the question of whether the definition of proof we applied was a match for his goals for this lesson. Our analysis of the transcribed episode suggests that Mr. MC's primary goals were to introduce students to the distributive property and encourage good justifications for a general claim. Further, this speculation aligns with the two descriptions of a good justification that Mr. MC had provided to the students and referenced during the classroom episode we analyzed. One of these was a sheet of paper titled "R.A.C.E. What makes a good mathematical justification?" This document suggested: (1) reword[ing], restate the question; (2) answering, include your answer and make it reasonable; (3) cite[ing], use information from the problem and what you previously learned; and (4) explain[ing], draw pictures, show work, explain thinking in words, and give specific details. The second document was titled, "What makes a good justification," which restated the above suggestions without referencing the R.A.C.E acronym. Early during the teaching episode, the teacher reminded students of the R.A.C.E. sheet and directed students to "follow your steps in R.A.C.E." Because these documents do not completely align with the

properties of proof on our definition, it is not surprising that students' arguments would not attend to all of these features. However, it is reasonable to anticipate that the students' work would score at Level 1 in our rubric, because our overarching indicator for this score is that the student at least searched for a conceptual insight.

Students' Written Work

Findings from our analysis of the students' written work were similar to our findings from our analysis of the teaching episode and the associated materials. When addressing generalizations with infinite domains, more than half of the students (six of nine) described structural reasons why the two approaches must produce the same result. Yet, these reasons were generally too vague to be viewed as proof. These students appeared to be searching for conceptual insights, which suggested that the students were at least aware of what they needed to do to prove their general claims. While the Number Trick task asked students to argue for a claim with the finite domain of natural numbers between 1 and 10 as well as for a claim with the infinite domain of all real numbers, most students did not specify which of these domains they were addressing in their written work. Thus we include in our analysis below our own inferences about which claim or claims each student was attempting to prove.

In total, we analyzed the nine student work samples provided to us by the editors. Due to space limitations, we discuss several representative cases and then summarize our findings from the sample.

Shawn, Finite-Domain Argument Score, N/A; Infinite-Domain Argument Score, 0

Shawn explicitly asserted that the two approaches will equate for any input numbers, including large natural numbers outside the originally proposed domain. He offered several examples that conformed to his claim. No conceptual insight was found in his response, and we find no evidence that he searched for a structural link between the two approaches and the results. His argument was purely empirical.

Hope, Finite-Domain Argument Score, 2; Infinite-Domain Argument Score, N/A

Hope tested every natural number case from 1 to 10 but made no claim about all real numbers. We judged her argument to be a proof of a claim with a finite domain.

Jared, Finite-Domain Argument Score, 2; Infinite-Domain Argument Score, 0

Jared claimed the equivalence "works with all" numbers. Jared compared the two approaches for all natural numbers 1 to 9 and placed check marks beside the work, as if noting that every case had been tested. If Jared's claim was restricted to this domain, then his support was an exhaustive proof. (The case of $n = 10$ was not included; perhaps he interpreted "between 1 and 10" as not including 10.) Because Jared included no structural link equating the approaches and his work does not suggest any search for structure, his work is not proof of the more general claim.

Emma, Finite-Domain Argument Score, 1; Infinite-Domain Argument Score, 1

We coded Emma's argument in regard to both the finite-domain and the infinite-domain claims because at first she claimed that the Number Trick works for

numbers 1–10 (assumed to be natural numbers) but scratched that out. In both cases, Emma's argument, and our analysis of it, focused on structure that she leveraged to equate the two approaches. Emma noted that the 4 was to be doubled regardless of whether it was added to the 5 prior to doubling or after doubling.

In particular, Emma wrote, "In the first equation when she added the 5 + 4 and doubled it, but you must realize that 4 is still part of the equation even though it was smushed [sic] in with the 5, you did double the 4 but when it was part of the 5 [sic]." Emma clearly searched for a conceptual insight and found one, but we had to make *significant assumptions* about Emma's approach. We assumed that Emma noted that both the 5 and the 4 were ultimately doubled in both approaches (perhaps she implicitly invoked the commutative and associative properties), but our assumptions required considerable *gap-filling* given that Emma's response was vague and clumsily worded.

Jenna, Finite-Domain Argument Score, 2 or 1, Infinite-Domain Argument Score, 2 or 1

Jenna wrote, "Jessie's trick will work for any number between 1-10." We assumed Jenna referred to natural numbers in this range, a finite domain, but her argument could also be interpreted as addressing an infinite domain: all real numbers between 1 and 10. Jenna wrote, "$(e + 4) \cdot 2 = e \cdot 2 + 4 \cdot 2$, or $e \cdot 2 + 8$." as her key representation of the structure she observed. These equations illustrate Jenna's thinking about how to transform one of the general expressions into the other using the distributive property. Jenna's argument could be viewed as proof, if we assume that the distributive property was a prior result to her. If this were so, the distributive property served as a tool for linking the two structures, a conceptual insight explaining why the two approaches produce the same outcome for numbers in the specified domain.

However, this interpretation assumes that the distributive property was a prior result for Jenna, which brings out a dilemma. As noted earlier in chapter "Overview of Middle Grades Data" (this volume), the teacher used the Number Trick task as a way of introducing the distributive property. Consequently, if Jenna's argument was viewed as communication between her and her classroom community, then the distributive property could not be taken as shared knowledge. From this perspective, Jenna's response did not have backing in a prior result. Instead, Jenna's key equations could be viewed as merely stating a generalized version of what Jessie found in her specific instance, $e = 5$, as given in the task. In that case, Jenna's score would be a 1 because she expressed awareness that *a general argument is based in mathematical structures of the conditions and the conclusion* and awareness that *inferential links between these structures must be illustrated.* However, Jenna failed to provide reasoning based in knowledge assumed to be prior knowledge. Thus, Jenna's score depended on whether the argument was a communication with herself, with her teacher, or with her peers, because students in a single classroom may have very different mathematics backgrounds.

Summary

Although only one student in the data set presented an argument that could be taken as proof of the general claim with infinite domain—under certain assumptions—six of nine students presented evidence that they searched for conceptual insights. We found few characteristics of "proof" as described in the definition provided to us among the students' work, but we found evidence that students engaged in a practice critical to proof construction: finding a structural link between the two approaches that allows one to equate the two approaches. This finding was also supported by exchanges found in the transcripts, where students discussed with peers and teachers the reasons why the two approaches would equate for any number. Practices such as searching for structure linking two approaches/expressions can be groundwork for learning about the other characteristics of proof and proving such as being explicit about the prior results leveraged in the reasoning and explicit about the deductive inferences made in a progression toward writing proofs in the canonical sense.

Is a Proof Accessible to This Classroom Community?

After our analysis, we wondered, "Can we envision an argument that does not rely on the distributive property and could be taken as proof?" We also wondered, "Would such an argument be accessible?" We cannot answer these questions definitively, as we lack knowledge of these students' mathematical backgrounds and past experiences. Yet, based in our knowledge of many US states' mathematical content standards, we developed an argument that we believe could be accepted as proof and is within the conceptual reach of these students.

The argument we developed relies upon the interpretation of multiplication as repeated addition. This would serve a prior result, an informal definition. In notation, $2(n + 4) = (n + 4) + (n + 4)$. Applying the associate and commutative properties, $(n + 4) + (n + 4) = 2n + 2 \cdot 4 = 2n + 8$. We could also envision prose or a generic example argument that accomplishes a similar proof. To label the argument as proof, we would not require the students to explicitly name the prior results used.

The argument above could be leveraged as a generic example proof of the distributive property if the domain was restricted to natural numbers. Extending the argument to other multipliers (e.g., non-integer rational numbers and irrational multipliers) would be difficult or impossible for this classroom community. After all, the distributive property is generally taken as an axiom in advanced mathematical classes where the real number system is developed.

Data from Another Project

We were concerned that the data provided to us came from students who had not received explicit instruction on "the rules of the game" for proof and proving as described in the working definition provided to us and in the rubric we developed from this definition. Yopp (2015) pointed out that students benefit from explicit instruction on types of claims in mathematics and how these claims are written. Yopp also noted that students benefit from explicit instruction on how to present their arguments and what modes of argumentation are acceptable in mathematics.

Below, we include data on a different but related task from our own project, Longitudinal Learning of Argument Methods for Adolescents (LLAMA) (see acknowledgements). We do this only to illustrate the possibilities for proof and proving in classrooms where the "rules of the game" for proof and proving are explicitly taught. Students in the LLAMA project were taught by teachers who learned about our models and methods for viable argumentation and proving, and the teachers and students had access to our project-developed lessons. These lessons developed Common Core Grade 8 content (NGA & CCSSO, 2010) through viable argument activities. A complete description of LLAMA is beyond the scope of this chapter, but for our current purposes, we summarize key features of LLAMA in terms of what students were taught:

1. Viable arguments/proofs include well-worded general claims or existence claims using language such as "for all," "if-then," or "there exist."
2. Viable arguments/proofs for general claims with large or infinite domains use representations (referents) that illustrate a general case(s) and the logical steps/ transformations pertinent to showing the claim is true. Explicitly, viable arguments/proofs *eliminate the possibility of counterexamples* (Yopp, 2015) to the claim.
3. Viable arguments/proofs include a narrative that links the representations/referents in the argument to the claim, notes the prior results used, the method/mode of argumentation used, and how the argument's steps are logical and demonstrate the truth of the claim.
4. Viable arguments/proofs of general claims use established methods/modes of argumentation such as exhaustion, direct (e.g., a sequence of modus ponens transformations/steps), contrapositive, and contradiction.

We used the term viable argumentation in place of proof and proving to acknowledge that axiomatic systems are not necessarily in place in the middle grades. We also wished to emphasize the phrasing in Common Core Mathematical Practice 3, *Construct viable arguments and critique the reasoning of others* (NGA & CCSSO, 2010).

Figures 2 and 3 illustrate sample work from two LLAMA students, Students A and B. This work is not representative of all students who received the intervention but illustrates the possibilities. Our point is that if we make explicit to students the

2. Maria claims she has two different computation approaches that produce the same answer when using a secret number. Approach 1: Maria takes her secret number, adds 5 to it, and then multiplies the entire result by 3. Approach 2: Maria takes her secret number, multiplies it by 3, and then adds 5.

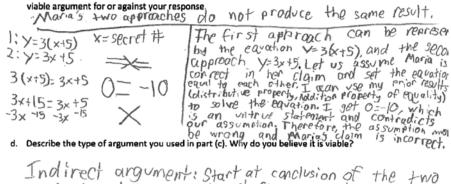a. Develop a variable expression for each of Maria's approaches.

Approach 1 = Approach 2   in the sense that they make
$3(x+5) = 3x+5$ the same answer

|| = "=" but rotated 90° to show two things above & below each other being equal

b. Set the two expressions equal and solve.

$3(x+5) = 3x+5$ $3x+15 = 3x+5$

$3x+15 = 3x+5$ $\dfrac{3x+15 = 3x+5}{3}$ $5 = 5$
$-3x \quad -3x$ " ① "
                                      $x+5 = x+1.\overline{6}$ $5 = 5$
$\Rightarrow \Leftarrow \quad 15 \neq 5$ No matter what $3(x+5)=3x+5$
                                      you do, x will NEVER $3x+15=3x+5$
                                      be equal to x $-3x \quad -3x$
                                                                         $15 = 5$

c. Do Maria's two approaches with the secret number produce the same result? Develop a viable argument for or against your response.

Let's assume that Maria's two approaches are equal. Due to prior results we know that we can use the properties of equali on a true equation to preserve their solution. However when using the properties of equality on Maria's 2 approaches (which SHOu be equal) u

d. Describe the type of argument you used in part (c). Why do you believe it is viable?

Indirect, because instead of   find a contra
proving logically why my   diction. Thus
claim can only be true. I'm   Maria's two
proving why my claim can't NOT   approaches mus
be true. (cuz claims only have 2   be equal.
states, true & un-true)

$DO + N't = Don't$ | $MUS + N't = MUSn't$ | $Yes + N't = YESN'^2$
$+ \cdot - = -$ |

Fig. 2 Response from Student A, a US eighth-grade student who participated in the LLAMA project

"rules of the game" for proof and proving, we are more able to assign the label of proof to their arguments. For example, we call Student A's argument "proof" of the claim "For all real numbers, none solve $3(x + 5) = 3x + 5$" because we can identify:

1. A general claim that Maria's two approaches are unequal no matter the choice of x—although the wording and labeling of the claim could be improved

c. Do Maria's two approaches with the secret number produce the same result? Develop a viable argument for or against your response.

·Maria's two approaches do not produce the same result.

$1: y=3(x+5)$ $x=$ secret #

$2: y=3x+5$

$3(x+5)=3x+5$ $0=-10$

$3x+15=3x+5$

$-3x-15=-3x-15$ ✗

The first approach can be represer by the equation $y=3(x+5)$, and the second approach $y=3x+5$. Let us assume Maria is correct in her claim and set the equation equal to each other. I can use my prior results (distributive property, addition property of equality) to solve the equation. I get $0=-10$, which is an untrue statement and contradicts our assumption. Therefore, the assumption must be wrong and Maria's claim is incorrect.

d. Describe the type of argument you used in part (c). Why do you believe it is viable?

Indirect argument: Start at conclusion of the two approaches being equal and find a contradiction

I stated my claim, showed evidence, and explained my reasoning.

Fig. 3 Response from Student B, a US eighth-grade student who participated in the LLAMA project

2. A referent, the equation-solving steps, with the appropriate generality and a narrative discussing the logic expressed in the referent: that the use of "solution-preserving steps" (the key prior result) leads to a contradiction, rendering the assumption of a solution false

3. A clear, unambiguous expression of the method/mode of argumentation, proof by contradiction, and an explicit discussion of this method/mode as proof

Moreover, when we applied the rubric developed for this chapter, we still arrived at a Level 2 rating. The argument convinces us that the general claim is true (Characteristic 1), uses deductive-like inferences from prior results (Characteristic 2), and needs little gap-filling (Characteristic 3). The referent provides the structural links (Characteristic 4), the referent (equation-solving approaches) is canonical (Characteristic 5), and we all sanctioned the argument as proof (Characteristic 6). We also labeled Student B's argument as proof for similar reasons. We included Student B's argument as a more succinct proof that explicitly names the prior results used in the logical inference: the distributive property.

Conclusion

Applying a definition of proof to middle-grade students' arguments is challenging when the definition is presented as a list of characteristics derived from social-cultural norms in mathematics and formal logical constructions like deduction. Middle-grade students may have access to very different guidelines on what

characteristics should present in proofs, arguments, and justification. Our focus was on the "rules of the game" for proof and proving, but the students in Mr. MC's classroom were focused on rules for justifications as expressed in the R.A.C.E sheet. As we noted in the beginning, our theoretical framework did not align well with the goals of instruction in this particular episode, and our focus may have misconstrued, and even overlooked, learning opportunities presented to students in this class. Consequently, our chapter can serve as a cautionary tale for researchers about applying a definition of proof or proving to a classroom where the students have different goals and notions about what it means to argue that a claim is true and justify their thinking.

Having acknowledged this possible mismatch, we did find that the definition provided to us and the rubric we generated, which valued conceptual insights, proved useful in finding proof-like features among students' work. Most students in Mr. MC's class at least searched for conceptual insights that explained why the two approaches must produce the same result. Perhaps searches for conceptual insights can be encouraged in classrooms that use very different guidelines for developing acceptable justifications, arguments, and proofs.

Our rubric was also useful in determining what was missing from an argument that prevented it from being sanctioned as proof. Scores of 0 and 1 were most easily determined by assessing whether or not a student searched for a conceptual insight. Category 1 ended up being a broad category containing responses that expressed a conceptual insight that could be leveraged toward proofs and responses that made clear the students at least searched for conceptual insights. Category 0 contained responses that included empirical support or no evidence that the student searched for a conceptual insight. Perhaps, for practitioners, the rubric can serve as a formative tool for giving students feedback as they learn the "rules of the game" for mathematical proof and proving.

Our assessment did however rely heavily on our abilities to recognize insights among the "noise" found in middle-grade students' writing. This type of assessment poses risks. Researchers might overlook potentially fruitful student reasoning that is novel and expressed ambiguously, or researchers might inadvertently impose sophisticated reasoning onto an argument that the student who wrote it did not possess.

In closing, we argue that giving students access to the "rules of the game" for proof and proving in mathematics, including the language of mathematics and the accepted methods/modes of proving, can help students understand what they are required to do when proving as well as help them to clearly communicate their reasoning to teachers, researchers, and peers. The rules of the game can also empower students with knowledge of the mathematical obligations for proving a general claim, such as representing the general case and demonstrating through logical applications of prior results that every case satisfying the conditions also satisfies the conclusion. And yet, we acknowledge that the "rules of the game" as we describe them may not be appropriate for all classrooms and may be inconsistent with some teachers' and researchers' goals.

References

Cheng, P. W., & Holyoak, K. J. (1985). Pragmatic reasoning schemas. *Cognitive Psychology, 17*(4), 391–416. https://doi.org/10.1016/0010-0285(85)90014-3

Czocher, J. A., & Weber, K. (2020). Proof as a cluster category. *Journal of Research in Mathematics Education, 51*(1), 50–74.

Hanna, G. (1990). Some pedagogical aspects of proof. *Interchange, 21*(1), 6–13.

Harel, G., & Sowder, L. (1998). Students' proof schemes: Results from exploratory studies. In A. Schoenfeld, J. Kaput, & E. Dubinsky (Eds.), *Research in collegiate mathematics education III* (pp. 234–283). American Mathematical Society.

Johnson-Laird, P. N. (1983). *Mental models: Towards a cognitive science of language, inference, and consciousness*. Cambridge University Press.

National Governors Association Center for Best Practices & Council of Chief State School Officers. (2010). *Common core state standards for mathematics*. Authors.

Sandefur, J., Mason, J., Stylianides, G. J., & Watson, A. (2013). Generating and using examples in the proving process. *Educational Studies in Mathematics, 83*, 323–340. https://doi.org/10.1007/s10649-012-9459-x

Stylianides, A. J. (2007). Proof and proving in school mathematics. *Journal for Research in Mathematics Education, 38*(3), 289–321.

Stylianides, G. J., Sandefur, J., & Watson, A. (2016). Conditions for proving by mathematical induction to be explanatory. *The Journal of Mathematical Behavior, 43*, 20–34. https://doi.org/10.1016/j.jmathb.2016.04.002

Stylianides, G. J., & Stylianides, A. J. (2008). Proof in school mathematics: Insights from psychological research into students' ability for deductive reasoning. *Mathematical Thinking and Learning, 10*(2), 103–133.

Weber, K. (2014). What is a proof? A linguistic answer to an educational question. Paper presented at the 17th Annual conference Special Interest Group of the Mathematical Association of America on Research in Undergraduate Mathematics Education. Denver, CO. Abstract retrieved from http://sigmaa.maa.org/rume/crume2014/Schedule/Papers.htm.

Yopp, D. A. (2015). Prospective elementary teachers' arguing and claiming in responses to false generalizations. *The Journal of Mathematical Behavior, 39*, 79–99. https://doi.org/10.1016/j.jmathb.2015.06.003

Yopp, D. A., & Ely, R. (2016). When does an argument use a generic example? *Educational Studies in Mathematics, 91*, 37–53. https://doi.org/10.1007/s10649-015-9633-z

Dr. David A. Yopp's research focuses on teaching and learning mathematics with and through proof, proving, and viable argumentation. He has served as the principal investigator for the NSF-funded DRK-12 projects Examining Mathematics Coaching (EMC), Learning Algebra and Methods of Proof (LAMP), and Longitudinal Learning of Viable Argument in Mathematics for Adolescents (LLAMA). His work is published in journals such as *Educational Studies in Mathematics*, *Journal of Mathematical Behavior*, and *ZDM Mathematics Education*.

Dr. Rob Ely's research is focused in two main areas: one is how middle-grade students learn the conceptual elements of mathematical argumentation. The other area is student reasoning in

calculus, particularly about infinitesimal quantities. He is currently a co-PI on the NSF-funded Longitudinal Learning of Viable Argument in Mathematics for Adolescents (LLAMA) project. His work has been published in a variety of journals, including *Journal for Research in Mathematics Education* and *Educational Studies in Mathematics*.

Dr. Anne E. Adams' research focuses on the development of student mathematical reasoning and argument and on developing teachers' ability to support student reasoning. She has been awarded funding for numerous research projects and is currently a co-principal investigator for the NSF-funded project Longitudinal Learning of Viable Argument in Mathematics for Adolescents (LLAMA). A secondary research interest is informal mentoring and supporting new faculty. Her work has been presented at regional, national, and international conferences and published in national and international journals, including *Journal of Mathematical Behavior*, *Mathematics Education Research Journal*, *Mentoring and Tutoring*, and *The Mathematics Enthusiast*.

Annelise W. Nielsen is a doctoral student at the University of Idaho working under the advisement of Dr. Yopp. She served as a research assistant on the NSF-funded project Longitudinal Learning of Viable Argument in Mathematics for Adolescents (LLAMA).