

Running Head: Toulmin Lesson Study Debriefing

**Challenges and Benefits of Using Toulmin's Argumentation Model to Assess Lesson Study
Debriefing Sessions**

Randall E. Groth

Salisbury University

D. Jake Follmer

West Virginia University

Abstract

As lesson study becomes more prevalent, there is a need to continuously develop theoretical and methodological infrastructure to support and refine this means of professional development. In this article, we present a critical methodological analysis of the challenges and benefits of using Toulmin's argumentation model to assess the debriefing phase of lesson study. During debriefing sessions, teachers offer arguments about how to improve teaching that are grounded in observations of students' learning. Toulmin's model provides a means to analyze the structure of such arguments. Using an empirical example, we illustrate challenges of using the model, such as determining appropriate grain sizes for data and claims, identifying and evaluating qualifiers, recognizing multiple categories of backing, identifying implicit warrants, and deciding between the individual or the group as a unit of analysis. We also discuss benefits such as being able to systematically compare pedagogical arguments against one another, assess attainment of debriefing session goals, and characterize group discursive dynamics. Despite the challenges of using the Toulmin model, we conclude that it provides a useful framework for systematic analysis of lesson study debriefing sessions. The present article can help researchers anticipate and address challenges of conducting Toulmin-based qualitative analyses of debriefing session discourse.

Keywords: Lesson study, argumentation, discourse, qualitative research, professional development

Challenges and Benefits of Using Toulmin's Argumentation Model to Assess Lesson Study Debriefing Sessions

Research on lesson study has grown steadily during the past two decades (Barlow et al., 2021; Takashi & McDougal, 2016). With this growth, several variations of lesson study have emerged, but these variations tend to have a shared set of core components. Commonly shared components include teachers' collaborative planning of a lesson around an identified student learning goal, implementation and group observation of the lesson, and group discussion and analysis of the lesson after it has been implemented (Lewis et al., 2009). Research related to lesson study requires infrastructure to support assessment of teachers' activities within and across these core components. Although the development of such infrastructure is underway, substantial work remains to be done (Rasmussen, 2016; Widjaja et al., 2017).

In this article, we explore a theory and method for assessing the teacher conversations that occur after lesson implementation. The Japanese word for this component of lesson study is *hanseikai* (Peterson, 2005). English words and phrases that have been used for it include "post-lesson conversation" (Seino & Foster, in press), "post-lesson discussion" (Bieda & Huhn, 2017; Wijada et al., 2017), "post-lesson reflection" (Bieda & Huhn, 2017; Rasmussen, 2016), "lesson study analysis meeting" (Amador & Carter, 2018), and "debriefing" (Groth, 2011; Ricks, 2011; Murata et al., 2012; Suh & Seshaiyer, 2015). We use "debriefing" in the present manuscript to concisely denote a lesson study group session devoted to recounting the group's experiences implementing and observing their collaboratively planned lesson to discuss what went well and what could be improved. This type of discourse about enacted lessons is often a powerful catalyst for professional growth during lesson study (Ricks, 2011; Suh & Seshaiyer, 2015; Widjaja et al., 2017).

Amador and Carter (2018) noted that lesson study debriefing “provides an opportunity for teachers to notice professionally as they focus on students’ thinking in reflective discussions” (p. 6). They defined professional noticing as consisting of “(a) attending to student thinking, (b) interpreting student thinking, and (c) deciding how to respond based on the analysis of student thinking or with a connection to broader principles of teaching and learning” (p. 7). Others characterize the central purpose of lesson study debriefings in a similar manner. For example, Wijada et al. (2017) spoke of debriefings as opportunities for participants to “examine and discuss their ‘evidence’ of student learning and share ideas to improve the teaching and learning process” (p. 359). Seino and Foster (in press) noted that, during these sessions, it is “necessary to interpret and discuss students’ thinking, difficulties, and transformation of thinking in the lesson, based on detailed information about what the students said and did” (n.p.). Likewise, Lewis et al. (2009) emphasized the importance of sharing and discussing student learning data from the lesson and drawing implications for instruction. Given past research of this nature, we start from the premise that lesson study debriefings are to elicit teachers’ arguments about how to improve instruction that are grounded in their observations of students’ learning.

Purpose

The Toulmin (1958/2003) model has been used to assess teachers’ pedagogical argumentation in several studies (e.g., Chazan et al., 2012; González & Eli, 2017; Metaxas et al., 2016; Potari & Psycharis, 2018; Skultety et al., 2017; Steele, 2005). The purpose of this article is to present a critical methodological analysis of the potential of the Toulmin model to support assessment of lesson study debriefing sessions. We begin with a general overview of Toulmin’s model and how it can be used to assess pedagogical arguments. Then, we introduce an empirical example to illustrate the model’s application to the specific context of assessing outcomes from a

lesson study debriefing session. Drawing upon this empirical example, we identify several methodological complexities researchers must confront when conducting Toulmin-guided qualitative analyses of debriefing session discourse. We then describe specific benefits of employing Toulmin-based analyses of pedagogical actions and arguments. We conclude by suggesting directions for future research to continue to advance theoretical and methodological tools for the effective assessment of lesson study.

Toulmin Model Overview

The Toulmin model has six primary components: data, warrant, claim, backing, qualifiers, and rebuttals (Chazan et al., 2012; González & Eli, 2017). The nature of each component can be explained by considering a hypothetical, generic lesson study debriefing session argument. As noted earlier, lesson study debriefings are to contain evidence-based arguments about what students have learned during a lesson. Suppose a teacher observed that students successfully solved mathematical task A. This observation could be considered the *data* for an argument. From the data, the teacher might make the *claim* that the students have attained desired learning objective B. The (usually implicit) *warrant* for the claim would be that if students are able to successfully solve task A, then they have attained learning objective B. Warrants, in general, link data and claims with if-then statements of this nature (Warren, 2010). The warrant could be supported with various types of *backing*. For example, the teacher might point out that task A is carefully aligned to learning objective B and/or has helped reveal student understanding in the past. *Qualifiers* may enter the argument if there is room for doubt about the extent to which a successful solution to task A indicates students have obtained learning objective B. For instance, a qualifier such as *probably* might be used in a statement such as “students’ solutions to task A indicate that they probably have attained learning objective B.”

Rebuttals expand on potential limitations. In the present example, perhaps task A occasionally allows students to obtain the “right answer for the wrong reason.” If so, a well-developed argument would include such information. The elements of such arguments are traditionally summarized in Toulmin diagrams like the one shown in Figure 1.

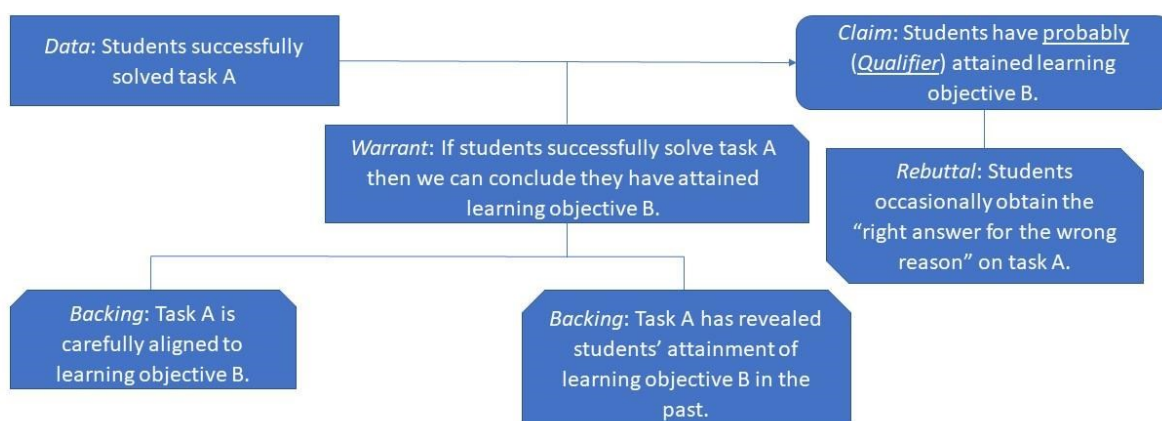


Figure 1. Toulmin diagram for hypothetical debriefing session argument about student learning

Ideally, lesson study debriefing arguments should link claims about student learning to improvements in teaching practice. For example, the earlier claim that students have obtained a desired learning objective could become data for a new claim that teaching strategies during the lesson were successful and should be used again in future lessons. Claims that students have not obtained a desired learning objective could become data for a new claim that teaching strategies should be altered during the next iteration. Warrants and backing, which are sometimes referred to collectively as “justification” in mathematics teacher education research (Chazan et al., 2012;

Gonzalez & Eli, 2017), may support pedagogical decisions about either retaining an existing strategy, modifying it, or using a new one during the next lesson iteration. Such justifications might draw upon mathematical and pedagogical considerations as well as experiences, values, and beliefs (Nardi et al., 2012). Qualifiers and rebuttals could enter arguments to characterize the likelihood that the chosen strategy would help students learn and to acknowledge its potential limitations.

Although Toulmin's model provides a means for initial analysis of the structure of an argument, it generally cannot provide the final word on issues of argument quality (Verheij, 2005). Toulmin (1958/2003) asserted that field-dependent standards must also come into play in assessing arguments. Accordingly, the quality of provided justifications often must be assessed using field-dependent standards; justifications may at times be irrelevant, invalid, not acceptable in the field, or incorrect (Simpson, 2015). As an illustration, consider a debriefing session about a lesson on fraction division. Mathematics education research indicates the importance of helping students establish a conceptual basis for fraction division rather than relying on solely procedural teaching of the invert-and-multiply algorithm (Empson & Levi, 2011; Ma, 1999). Suppose that, during a debriefing session, teachers notice that students produced incorrect answers to fraction division problems. On the basis of the data, some might claim that the invert-and-multiply algorithm should be explicitly taught at the outset. Teachers might justify such a claim by explaining that this strategy has helped their students produce many correct answers in the past. From the Toulmin perspective alone, there may be little to critique about such an argument. Bringing a mathematics education perspective to bear, however, would subject many aspects of the warrant, claim, and backing to scrutiny. In this particular instance, one might question whether the short-term gain of correct answer production would have a long-term

consequence of comparatively shallow learning. Taken together, Toulmin's model can help identify whether important argument elements are present, and it may also afford analysis of whether elements of the warrant and backing support or contradict one another. However, on its own, it cannot always speak to the veracity of field-specific justifications for claims and their limitations. In this work, we discuss an approach to assessing the quality of lesson study debriefing discourse in mathematics education and illustrate the application of this approach.

Background for the Empirical Illustration

Participants

The lesson study group providing the focus for our illustration consisted of three pre-service teachers and an in-service teacher. They were working together as part of a larger lesson study project (Groth, Bergner, Weaver, & Welsh, 2020). We refer to those in the group using the pseudonyms Ms. Jackson, Emily, Brian, and Hannah. Ms. Jackson was an in-service middle school mathematics teacher and the lesson to be discussed was implemented with one of her algebra classes. Emily and Brian were pre-service secondary mathematics teachers. Hannah was a pre-service secondary science teacher in the same scholarship program. Her inclusion in this group was intended to support one of the larger project's goals of encouraging interdisciplinary work between mathematics and science teachers during lesson study (Groth, Bergner, Weaver, & Follmer, in press).

The Research Lesson

The lesson discussed during the debriefing session to be considered was taught to a group of 36 eighth-grade students. Because the lesson took place during the COVID-19 pandemic, it was taught online, via Zoom. The primary goal of the lesson was to help students understand connections between the graph of a parabola and its corresponding vertex form. A modified

version of a Desmos lesson entitled, “Will it Hit the Hoop?” provided the basis for their lesson (Desmos, n.d.). The initial slides in the lesson introduced the idea of using a parabola rather than a line in some circumstances to fit functions to data. Students then viewed a series of basketball shots and predicted if each shot would go through the hoop. Students were then prompted to use graphs of parabolas to model shot trajectories to refine their predictions. After checking their predictions by watching videos of the completion of each shot, students were asked to look back on one of the videos with a coordinate grid superimposed. Using the grid, they were to find details such as height of the shooter and the coordinates of the hoop. They were then shown the vertex form for the parabola to model the trajectory and asked to explain what the vertex form revealed about the trajectory of the shot. The Desmos interface had a dashboard that allowed the lesson study group to see students’ work as it was being produced and also retain the work for later discussion and analysis.

Debriefing Session Structure

The debriefing session was also conducted via Zoom and took place two days after lesson implementation. To start the debriefing session, the first author of this essay asked participants to respond to the following discussion prompts:

- What do you believe students learned as a result of the lesson? How do you know?
- What remains for students to learn, in regard to the lesson objectives? How do you know?
- What changes would you make to the lesson before implementing it again? Why?

These prompts were designed to elicit and promote professional noticing, in the sense of attending to and interpreting students’ mathematical learning as a basis for making teaching decisions (Amador & Carter, 2018). The questions “how do you know” and “why” were posed to

prompt participants to justify their claims. The debriefing session began with participants' responses to the first prompt. The pre-service teachers each responded to the first prompt before hearing the in-service teacher's response. The same turn-taking pattern (i.e., pre-service before in-service) was followed for the remaining two prompts, with a different pre-service teacher starting the conversation each time.

Data Analysis

We applied Toulmin's (1958/2003) model to the analysis of the debriefing session video and the accompanying transcript. First, we viewed the video independently. This brought to bear the perspective of a debriefing session attendee (the first author) as well as that of an outside observer (the second author). Then, we reviewed the first pedagogical argument from the debriefing session transcript together to develop shared conventions for marking the rest of the transcript. We then independently annotated the transcript, looking for data, claims, warrants, backing, qualifiers, and rebuttals in participants' arguments about learning and teaching in response to the three main discussion prompts listed above. In addition, we independently wrote analytic memos (Saldaña, 2013) about our assumptions related to applying the Toulmin model, interesting patterns we noticed in debriefing session discourse, and perceived strengths and weaknesses of the arguments offered during the debriefing session. After working independently, we met to compare our annotations and memos. Comparing our work allowed us to identify several methodological challenges and decisions that must be navigated when using the Toulmin model to analyze debriefing session discourse. We also identified benefits to be derived from working through the challenges of applying the Toulmin model.

Challenges of Using Toulmin's Model to Assess Debriefing Sessions

Although the Toulmin model provides a well-established frame of reference for analysis of pedagogical arguments, it does not automatically resolve all ambiguities inherent in making sense of qualitative data from debriefing session conversations. Toulmin's model has been interpreted and applied in slightly different ways across multiple studies (Knipping & Reid, 2015; Krummheuer, 2015). The annotations and memos from our own independent analyses revealed points of convergence and divergence that emerged when using Toulmin's model to make sense of arguments offered during debriefing sessions. We discuss our similarities and differences in interpretation next in order to draw attention to key decision points one can expect to encounter when using the Toulmin model to assess debriefing sessions. The resolution of such decision points during qualitative analysis is often not given adequate attention in the literature (Elliot, 2018; Thompson et al., 2004). In applying the Toulmin model to the task of debriefing session assessment, key challenges we encountered were determining appropriate grain sizes for data and claims, identifying and evaluating qualifiers, recognizing multiple categories of backing, identifying implicit warrants, and deciding between using the individual or the group as a unit of analysis. As we illustrate and discuss each of these challenges, we also offer thoughts on how to approach and address each one.

Determining Appropriate Grain Sizes for Data and Claims

In some cases, we found it challenging to discern the number of data-claim pairs present in a given conversational turn. One of Emily's conversational turns (Figure 2) is illustrative. Emily offered the comments shown in Figure 2 in response to the prompt, "What do you believe students learned as a result of the lesson? How do you know?" One of us initially marked the turn as containing one data-claim pair, and the other marked three data-claim pairs. Taking the

approach of marking one data-claim pair, one could consider the entire lesson to be the “data” portion of Toulmin’s structure, with the accompanying claim that students understood how linear and quadratic functions can be fit to data. Alternatively, three data portions could be discerned: (i) initial parts of the lesson that introduced the idea of best-fit line, (ii) subsequent student/teacher interactions about linear and quadratic functions, and (iii) students’ work related to the axis of symmetry. Three accompanying claims would be: (i) students learned the idea of line of best fit (lines 2-3, Figure 2), (ii) students understood the difference between linear and quadratic functions (line 4), and (iii) students demonstrated an understanding of the axis of symmetry (lines 13-14). Importantly, the one-pair approach and the three-pair approach to discerning data and claims were both justifiable in light of the evidence available to us.

1	I wrote down a few notes, while the lesson was taking place and I noticed that one idea that the
2	students didn't really know before coming into this was the line of best fit. So, I thought that they
3	learned with the line of best fit was.
4	And they also understood the difference between a linear function and also one that is quadratic.
5	Because, with our example we did the basketball hoops and they were basically just upside down Us or
6	quadratic functions. And so, then, I think I don't know if it was Miss Thompson or one of us, we asked of
7	an example of a linear function and then one student came up with the idea of just like age against time.
8	So, like as someone is older grows older times going on as well, obviously. And then I think Miss
9	Thompson came up with the price of goods so like (Miss Thompson interjects: “hourly wage”), yeah
10	hourly wage. Like the more you work, the more money you're going to get, which was a constant
11	function that kept going up at a constant rate. So this straight line relationship was different than a
12	current relationship, which was what was happening with the basketball.
13	And then also another new term that the students learned or I think they – I don't think they knew it
14	before – but it was the axis of symmetry. And that was basically the line that was drawn straight down
15	from the vertex of the parabola, and that was what one student came up with.
16	And also another concept that was brought up was the y-intercept, which I think they knew about, but
17	they needed a refresher. It was basically the height of the shooter.
18	But, those are a few ideas I had about when the students were like, answering questions, and that's
19	what I thought that they had learned.

Figure 2. Emily’s conversational turn in response to the first debriefing session prompt

Each data-claim pair grain size choice has distinct advantages. Choosing a larger grain size as the focus of the argument analysis increases the likelihood of agreement (i.e., consistency) in coding and may, in turn, increase precision in the claims that are identified and linked. In the Figure 2 excerpt, we both agreed it was safe to assume that Emily was referring to her observations from the lesson study group's research lesson as a basis for her claims. Researchers working in teams may wish to adopt the largest grain size possible if agreement among team members is a high priority. Of course, such an approach runs the risk of leaving out smaller, important details and nuances. If detailed micro-analyses of data-claim linkages are of interest, researchers may opt for smaller grain sizes. In the Figure 2 example, it seemed reasonable to infer that Emily was referring to specific portions of the lesson in making claims related to what students learned about differentiating between linear and quadratic functions and about lines of symmetry. Although Emily did not explicitly refer to specific lesson segments, she did reference events related to each claim that occurred at different points in time during the lesson (e.g., Ms. Jackson's question in line 9 and a student's sketch of an axis of symmetry in line 15). For this reason, making justifiable inferences about the data implicitly used as the basis for claims may more effectively identify the portions of lessons teachers believe supported specific student learning outcomes. Hence, the purpose of a study and researchers' judgements about justifiable levels of inference must play prominent roles in choosing grain sizes for data-claim pairs.

Identifying and Evaluating Qualifiers

At times, it can also be challenging to decide if an utterance functions as a qualifier or not. We encountered this particular challenge when analyzing debriefing session arguments about what students had learned during the lesson. In some cases, examples of student thinking

were offered as backing for claims about students' learning, and in other cases, they served the function of helping qualify such claims. For example, in line 7 of Figure 2, Emily made reference to a student-generated example of age vs. time as backing for the idea that students learned about linear functions. In contrast, later in the debriefing session, Ms. Jackson referred to students' work in the process of qualifying a claim students had learned that the vertex represented the position of the basketball at its highest point. She said,

A couple of the kids actually were also able to pick up on the vertex being the height of the basketball - the basketball at its highest point, I should say. Not everyone did, but going back through their responses, there were several kids, I would say, probably 25%.

Ms. Jackson's reference to students' work, in this case, qualified the claim that the class learned this connection (with "probably") and went an extra step of estimating the percentage who likely did so ("25%"). Emily's reference, in contrast, seemed to use one student's correct response as evidence that the entire class had learned (though we cannot claim so decisively without questioning Emily further about her argument). Students' work ideally will be discussed many times during a debriefing session, so it is useful for researchers using the Toulmin model to be alert to the challenge of distinguishing among the different reasons for which it may be cited.

Judging the veracity of qualifiers is another challenge associated with using the Toulmin model to assess debriefing sessions. In some cases, lesson data were readily available to assess the appropriateness of the qualifiers used. For example, Ms. Jackson's statement that "probably 25%" of students made the intended connection could readily be checked against lesson data gathered via the online Desmos teacher dashboard. In contrast, Emily's qualified statement about students' learning of the axis of symmetry (lines 13-14, Figure 2) was not as easy to verify. She said, "I *think* (they learned)," and noted that her uncertainty about students' learning stemmed

from not knowing if they already learned the idea in a previous lesson. Similarly, later, Emily stated “I think they (the students) knew about it (the y -intercept, prior to the lesson)” (lines 16-17, Figure 2). Again, assessing the accuracy of this statement was beyond the scope of the data we had gathered. Where feasible, researchers aiming to assess the accuracy of debriefing session arguments and claims should attempt to pair Toulmin-based analyses with domain-specific theory and methods, in part by gathering and analyzing supplemental data, especially when claims may be based upon data outside the scope of the lesson study experience. For instance, statements about what students knew or did not know prior to the lesson could be checked against supplemental data from pre-assessments designed to capture theoretically important aspects of students’ thinking in relation to the lesson’s objectives.

Recognizing Multiple Categories of Backing

Backing is another aspect of the Toulmin model that can be challenging to discern from debriefing session data. During initial transcript analyses, our opinions about when and where backing was offered differed at times. One of Emily’s statements is again illustrative. At one point, Emily stated, “One idea that the students didn’t really know before coming into this was the line of best fit” (lines 1-2 in Figure 2). One researcher took the statement as backing for an argument that students learned about the line of best fit during the lesson, and the other did not. This difference in opinion may have stemmed from the types of backing we expected to see before analyzing the transcript. As noted earlier, a primary purpose of debriefing sessions is to support one’s arguments using classroom data from the lesson. Emily’s statement does not contain that sort of backing. Instead, it supported the argument that students attained the intended learning objective during the lesson by conjecturing that they had not attained the objective

beforehand. Given the brevity of the conjecture and its lack of connection to student learning data, it was not easily and readily identifiable as backing in a first pass through the data.

Multiple types of backing were present throughout the remainder of the debriefing transcript. Although, as expected, classroom data from the lesson were used as backing for some arguments, conjectures about what students knew before the lesson were offered as backing by others as well. Not all backing was of equal quality. For example, Emily at times provided stronger backing for her claims. At one point, Emily claimed that the lesson did not contain enough opportunities for students to recognize important mathematical connections. In her backing, she noted the mathematical importance of having students explore connections between graphs and their corresponding equations. She went on to support this backing by explaining how changes in an equation's coefficients correspond to changes in its graph (Emily's argument is shown in its entirety in Figure 3, which we use later to re-analyze the argument from a different perspective). Her comments represented not only another type of backing present in the conversation (strictly mathematical backing), but seemingly another level of quality as well (backing with sub-backing). So, in analyzing the backing for debriefing session arguments, researchers may find it useful to attend to two categorical variables: argument type and argument quality. The first of these two seems akin to a nominal categorical variable, with types of backing such as empirical observations of students' work, connections to students' prior curricular experiences, and mathematical considerations. The second of the two categorical variables is more ordinal in nature, as backing can have varying levels of sophistication and complexity.

Mathematics education domain-specific theory can be useful for evaluating backing along both nominal and ordinal dimensions. For example, van Es (2011) offered a framework to characterize the quality of teachers' noticing of students' mathematical thinking. Nominally, the

framework contains the categories of *what* teachers notice and *how* they notice. Ordinarily, lower levels of *what* teachers notice involve attending to non-mathematical aspects of the classroom environment and teacher pedagogy. Higher levels involve attending to how students' mathematical thinking is linked to specific teaching strategies. The construction of such connections is ideally at the core of lesson study debriefing sessions. Lower levels of *how* teachers notice are characterized by offering general impressions of what occurred during a lesson with little or no evidence to support analysis. Higher levels are marked by the presence of interpretive comments that draw upon specific classroom events and interactions. Researchers using this particular domain-specific framework to complement Toulmin analyses, then, should direct attention to both the type and quality of backing evidenced in debriefing session arguments.

Working directly from a Toulmin perspective, Nardi et al. (2012) offered categories researchers may find useful in characterizing nominal aspects of backing. Their framework describes types of support mathematics teachers may offer for pedagogical arguments, such as mathematical theorems and definitions, general pedagogical principles, accepted curricular practices, and past teaching and learning experiences. Such a categorization scheme is an important part of assessing lesson study debriefing sessions, which ideally contain arguments backed by experiences from teaching the lesson. Other types of backing, such as mathematical theorems and definitions, certainly are appropriate, but one can conclude a lesson study debriefing session has not met its essential purpose if backing based on teaching and student learning during the lesson is entirely absent. Technically, the Nardi et al. categories are intended to characterize warrants, but we discuss them here because some research teams combine warrants and backing into the single category of "justification" when doing Toulmin analyses in

mathematics education (Chazan et al., 2012; Gonzalez & Eli, 2017). Next, we turn to the issue of deciding between identifying warrants and backing separately or using the broader category of “justification” to capture both.

Identifying Implicit Warrants

Backing for teachers’ arguments tends to be explicitly stated in debriefing session transcripts, but warrants are usually implicit. The implicit nature of the warrant is not unique to the case of analyzing lesson study debriefing, but pervades Toulmin analyses across professional fields (Warren, 2010). This phenomenon can be particularly vexing for teams of researchers trying to reach agreement on the precise warrant for any given data-claim pair. Combining warrant and backing into the single category of justification removes a great deal of ambiguity that can come into play in discerning warrants. Hence, combining the two is a reasonable approach for teams of researchers who are primarily interested in evaluating how teachers support their pedagogical arguments. Those using the Toulmin model to analyze debriefing sessions could also then choose to use the broader category of justification rather than separating backing from warrant.

Before completely setting aside the idea of discerning the warrant, however, it is useful to consider potential benefits of deciphering warrants from data-claim linkages. One situation in which identifying the warrant may be beneficial is when no explicit backing is offered for an argument. For example, when Brian responded to the question of what remained for students to learn, he stated,

So, what remains to be learned in the lesson I think is like how the equations, how different equations, would change, (and) how the graph looks. Like adding stuff to the -

which part of the equation - at the end - the last slide, like, how would that change the graph and visualizing it.

No explicit backing was discernible in this comment. Possible backing could have included student work indicating difficulties with relating graphs and equations during the lesson. It might also be productive to cite experiences or research/theory suggesting that the types of tasks suggested for future lessons were likely to be successful. In absence of explicit backing, we could plausibly assume that he may have been operating under the simplistic warrant that adding a specific (in this case, under-specified) problem in a future lesson would ensure student learning. From a mathematics education perspective, this type of warrant might be grounded in a naive pedagogical theory that knowledge can be transmitted directly from teacher to student simply by including content in a lesson. From an assessment standpoint, there is value in understanding the pervasiveness of such warrants in pedagogical arguments. Having such knowledge allows researchers to assess the quality of argumentation during debriefing sessions, and it also alerts lesson study facilitators to be conscious of challenging teachers to question and deepen any surface-level pedagogical theories that may be offered during debriefing.

When backing is included in a debriefing session argument, it can help remove some of the ambiguity involved in discerning implicit warrants. For example, when talking about how the lesson could be improved before implementing it again, Ms. Jackson commented on possibilities that would be available in a face-to-face, rather than online, setting. She stated,

I mean, in a perfect world, from my perspective, I feel like it'd be great to actually have them, you know tossing a basketball, and you know, sketching, what does this basketball (trajectory) look like? And, you know, then we can move into this lesson where it's, you know, even taking it sort of a step back - what's the shape of a parabola? Where does that

data come from that we would use to graph the data? You know, what point on the graph is the maximum or is the vertex, at what point does it hit the ground, where would that be represented on our graph, that kind of thing. You know, obviously in this virtual setting, we have to be digital, but if we can incorporate more hands-on and digital that would be the dream, from my perspective.

Ms. Jackson's explicit backing for these proposed changes centered on the idea that physical experience would provide enhanced opportunities for students to understand how parabolas can model basketball shot trajectories. Given this theme in the backing, one might summarize the warrant as, "If students did not have opportunities to connect physical experience to mathematical representations, then including such opportunities will enhance their learning."

In each example above (one without and one with explicit backing), the warrant concisely summarizes the implicit pedagogical theory undergirding the argument. Researchers who decide to parse warrant and backing may find it useful to characterize the backing of an argument first (if backing was stated), and be sure to capture any overarching theme in the backing when stating a warrant. Identifying warrants seems essential to studies that have a goal of concisely summarizing the underlying theories that guide teachers' pedagogical reasoning. Of course, not all research on lesson study debriefing will have such a focus, so, in many cases, using the broader category of "justification" to encompass backing and warrant is also viable. In some cases, describing the justification will also provide insight into teachers' underlying pedagogical theories, though perhaps not as directly as would be achieved by identifying warrants.

Deciding between Individual and Group Arguments as Units of Analysis

Toulmin's (1958/2003) model was originally designed to analyze the structure of arguments offered by individuals. However, mathematics education researchers have at times

adapted it in order to model the structure of group-constructed arguments (Knipping & Reid, 2015). As we analyzed the data for our empirical example, portions of the debriefing session transcript stood out as candidates for group rather than individual argument analysis. One such transcript portion is shown in Figure 3, and an accompanying Toulmin diagram to illustrate the group argumentation structure is depicted in Figure 4. The conversation shown in Figures 3 and 4 occurred in response to the prompt asking what remained for students to learn in regard to the lesson objectives. We have discussed how some portions of this particular conversation can be modeled as individual arguments earlier in this manuscript; here, we show how one might consider these elements to comprise a group-constructed argument instead.

Brian: So, what remains to be learned in the lesson I think is like how the equations, how different equations, would change, (and) how the graph looks. Like adding stuff to the - which part of the equation - at the end - the last slide, like, how would that change the graph and visualizing it.

Hannah: Yeah that's kind of going off of what I was gonna say, because I feel like they had a pretty good understanding of what we were talking about at least in the beginning – and then that, that last slide. They had seen that equation, and that they'd never been presented with before or the form of it, at least. So, um, I guess just maybe like increased exposure to that after the fact, after it was introduced.

Emily: You know, I was going to talk about the equation, too. Because, I feel like, how graphs – you know, charts and equations - they're also interrelated. That like, you know, they have to be like, explained, and it doesn't matter the order. But as long as they show their connections and I also feel like – what was I gonna say so, I lost my train of thought – but something like, I feel like, the equations are important to show because, like a quadratic it has a particular shape, so it has a u-shape. And one thing that you can look if the equation has other than negative in front, you can tell it's either flipped over the y-axis or over the x-axis, or over the y axis. And also like, that y-intercept like if it's plus six it's going to be up to six, or you know, I feel like that would be important to show.

Figure 3. Portion of debriefing transcript lending itself to the modeling of a group-constructed argument

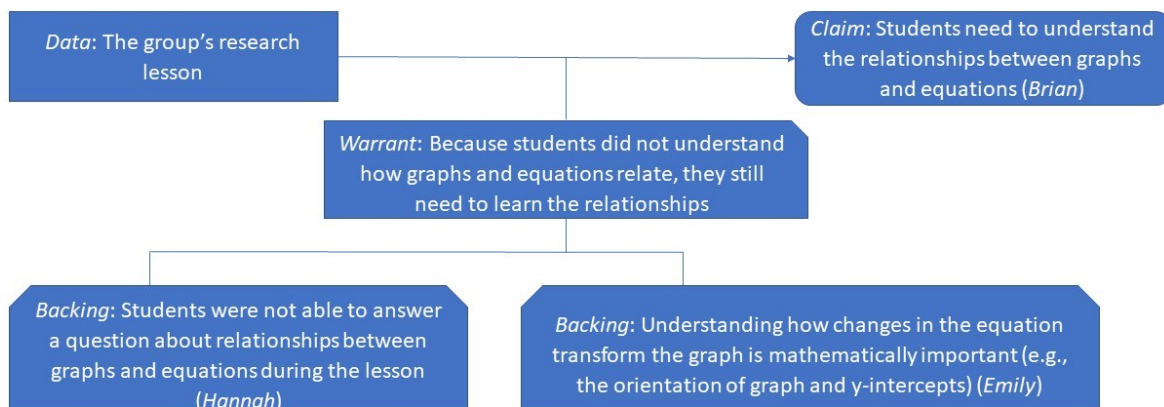


Figure 4. Toulmin diagram to show the structure of the group-constructed argument in Figure 3.

Earlier in this manuscript, we discussed how Brian's argument, shown in the first conversational turn (Figure 3), lacked backing. We said that in absence of explicit backing, we might assume a simplistic warrant that adding a specific, under-specified problem in a future lesson would ensure student learning. Analysis of his individual argument would then come to a rapid conclusion. If we put the claim in the context of group conversation, however, another perspective emerges. Hannah and Emily both agreed with Brian's claim, and they offered backing for it to help construct a more robust, collaborative argument. Hannah made reference to students' previous work as indicative of their need to better understand connections between graphs and equations. Emily then offered a degree of mathematical backing for the claim, as shown in Figure 3, and as we discussed earlier. Of course, both types of backing could have been stronger, as Hannah could have offered specific examples of student work to support her

observations, and Emily could have been clearer and more precise in discussing the mathematical rationale. Nonetheless, the backing offered by Hannah and Emily added a dimension to the group discourse that Brian's initial comment did not. Brian's contribution, though, might still be framed in a positive light, as it established a starting point for what could be considered a group-constructed argument.

Deciding between the individual and the group as the unit of analysis is ultimately a question of one's theoretical orientations toward knowledge construction, discourse, argumentation, and the Toulmin model. It is beyond the scope of the present manuscript to delve into all of the potential theoretical considerations in regard to such issues. Here, we conclude our discussion of this challenge by observing that each perspective has potential advantages. Individual analyses allow us to compare the quality of arguments against one another. Such comparison could be important to lesson study debriefing facilitators who would like to anticipate the need to press certain group members for deeper justification of their arguments in future debriefings. Group analyses provide insight by setting individual arguments in the context of group conversation. Knowing that Emily's backing, for example, was motivated by Brian's claim reveals potentially important group dynamics that took shape during the debriefing session and may engender critical, socially-shared understandings. Moreover, some researchers will likely find it theoretically tenable to merge individual and group analyses to capture benefits to be derived from taking – and combining – each perspective.

Benefits of using Toulmin's Model to Assess Debriefing Sessions

Any framework for analyzing qualitative data will have interpretive and methodological challenges, and Toulmin's model is no exception. Our discussion in the previous section is intended to help others anticipate such challenges rather than to provide reasons against

implementing Toulmin analyses of debriefing session arguments. Embracing and addressing such challenges can lead to deeper analyses of data. We believe Toulmin has a great deal to offer in providing a lens to assess whether or not debriefing sessions meet their primary goal of catalyzing claims about teaching and learning that are grounded in classroom data.

In particular, Toulmin analyses provide a systematic means to compare pedagogical arguments against one another and against the goal of backing one's claims with student learning data. To illustrate, consider the comments Ms. Jackson offered immediately after the pre-service teacher comments shown in Figure 3. In response to these comments, Ms. Jackson stated,

I was writing down what Emily just said, so that I can remind myself to jump into that, yeah, so the students had seen the quadratic equation in standard form. But this last slide that everyone has referenced, the extension number four, was the first time that they had seen the vertex form of the equation. And so really what I – when we designed this lesson – what I was trying to get them to see was that the vertex is (h,k) in the vertex form. By looking at their responses, only a few picked up from that equation that the vertex was $(10.76, 14.8)$. But only a few of them picked up from the equation that that was actually the vertex. Most of them were able to see on the graph, OK, that's where the ball was the highest. And on the graph, $(10.76, 14.8)$ is where that ball was located, but not many connected that to this part of the equation is where that came from.

Ms. Jackson went on to describe how she used this information to design subsequent lessons for her students.

From the perspective of Toulmin and domain-specific theory of mathematics teachers' professional noticing, we can see that Ms. Jackson's backing in the comment immediately above

was notably different from the pre-service teachers' (Figure 3). In particular, her backing included specific qualitative observations about the work students had done during the lesson when constructing an argument about what remained for students to learn. Summatively, it is useful to note such differences in argument structure to characterize the degree of success of a debriefing session and qualitative differences among participants' contributions. Formatively, it may be useful for facilitators of lesson study to have group members compare comments like Ms. Jackson's to others like the ones in Figure 3 to help establish a discursive norm of backing one's arguments with specific observations about students' learning.

Along with supporting formative and summative assessment related to individual pedagogical arguments, Toulmin analyses can provide insight about a lesson study group's discursive dynamics. We made this point earlier in regard to deciding between the group or the individual as the unit of analysis. Although the unit of analysis decision largely depends on one's theoretical orientation, multiple theoretical perspectives can make use of Toulmin analyses to identify significant group dynamics. From perspectives that embrace theories of collaborative knowledge construction, diagrams like the one shown in Figure 4 can provide at-a-glance summaries of large segments of debriefing session conversation. Those operating under theories that emphasize individual construction of knowledge might opt for individual Toulmin diagrams, but these can also be marked to indicate backing, data, claims, and other elements in common with others. In either case, examining the extent to which lesson study group members build upon one another's arguments and/or share common understandings with one another provides an assessment of the cohesiveness of the group.

Conclusion

Toulmin analyses of debriefing sessions open several directions for future research beyond those we have discussed in detail in this manuscript. For instance, our empirical example was typical of other research in which rebuttals were largely absent in pedagogical arguments (Knipping & Reid, 2015; Nardi et al., 2012). Because there is value in having teachers grapple with the limitations of proposed practices (Sawyer et al., 2020; Milewski et al., 2021), others may wish to design debriefing session prompts to try to elicit rebuttals, test them during debriefing, and refine as necessary in light of Toulmin analyses of the resultant data. A related line of research would be to investigate the discussion prompts and circumstances associated with more and less structurally sophisticated debriefing session arguments. Toulmin again provides a sound framework for such investigations. Relevant theory from mathematics education can play a complementary role in allowing one to assess argument quality alongside structural sophistication.

Our work suggests it would also be useful to view debriefing session comments from a Toulmin perspective not only after the session has concluded, but also while it is taking place. This, of course, is a non-trivial task. In this manuscript, we described time-consuming challenges associated with using the Toulmin model to analyze debriefing session arguments. A challenge for future research would be to distill such decision-making processes into tools and protocols that would help lesson study facilitators make fairly quick, reasonable judgments about the sophistication and veracity of the arguments being offered during debriefing sessions they lead. With such rapid analyses, they would be in position to press for additional detail and reflection in real-time. Research has explored the comments lesson study facilitators offer during debriefing

sessions (Seino & Foster, in press); research focused on their in-situ formative assessment practices and how they can be developed would be beneficial as well.

Lesson study has become more prevalent in many countries over the course of the past two decades (e.g., Barlow et al., 2021). As it continues to grow as a mode of professional development, there is a need to continuously develop the theoretical infrastructure needed to support it. We believe the Toulmin model can provide some of the vital theoretical infrastructure for the debriefing portion of lesson study. It is a viable means to assess the extent to which the central purpose of constructing arguments about teaching and learning from classroom data is met. Persevering through the challenges of Toulmin analysis can enhance summative and formative assessment of debriefing sessions and open new avenues for research related to this phase of lesson study.

Acknowledgement

This research was supported by the National Science Foundation (NSF) under award DUE-1852139. The views expressed herein are those of the authors and do not necessarily reflect those of the NSF.

References

- Amador, J. M., & Carter, I. S. (2018). Audible conversational affordances and constraints of verbalizing professional noticing during prospective teacher lesson study. *Journal of Mathematics Teacher Education*, 21, 5-34. <https://doi.org/10.1007/s10857-016-9347-x>
- Barlow, A. T., Pair, J. D., Hartland, K., Schmidt, T. A., Kassaei, A. M., & Woodard, C. A. (2021). Supporting the development of mathematics teacher educators through lesson study. *Investigations in Mathematics Learning*, 13(2), 125-140. <https://doi.org/10.1080/19477503.2020.1865029>

- Bieda, K. N., & Huhn, C. (2017). Investigating problem-solving perseverance using lesson study. *Mathematics Teacher*, 111, 207-212. <https://doi.org/10.5951/mathteacher.111.3.0207>
- Chazan, D., Sela, H., & Herbst, P. (2012). Is the role of equations in the doing of word problems in school algebra changing? Initial indications from teacher study groups. *Cognition and Instruction*, 30(1), 1-38. <https://doi.org/10.1080/07370008.2011.636593>
- Desmos. (n.d.). *Will it hit the hoop?*
<https://teacher.desmos.com/activitybuilder/custom/56e0b6af0133822106a0bed1>
- Elliot, V. (2018). Thinking about the coding process in qualitative data analysis. *The Qualitative Report*, 23(11), 2850-2861.
- Empson, S. B., & Levi, L. (2011). *Extending children's mathematics: Fractions and decimals*. Portsmouth, NH: Heinemann.
- González, G., & Eli, J. A. (2017). Prospective and in-service teachers' perspectives about launching a problem. *Journal of Mathematics Teacher Education*, 20, 159-201.
<https://doi.org/10.1007/s10857-015-9303-1>
- Groth, R. E. (2011). Improving teaching through lesson study debriefing. *Mathematics Teacher*, 104, 447-451. <https://doi.org/10.5951/MT.104.6.0446>
- Groth, R.E., Bergner, J.A., Weaver, S.D., & Follmer, D.J. (in press). Virtual tools and protocols to support collaborative reflection during lesson study. *Innovations in Science Teacher Education*
- Groth, R. E., Bergner, J. A., Weaver, S. D., & Welsh, G. S. (2020). Using Japanese Lesson Study to merge inservice professional development and preservice clinical experiences. *The Clearing House*, 93(2), 93-99. <https://doi.org/10.1080/00098655.2020.1729082>

- Knipping, C., & Reid, D. (2015). Reconstructing argumentation structures: A perspective on proving processes in secondary mathematics classroom interactions. In A. Bikner-Ahsbahr, C. Knipping, & N. Presmeg (Eds.), *Approaches to qualitative research in mathematics education: Examples of methodology and methods* (pp. 75-101). Dordrecht: Springer.
https://doi.org/10.1007/978-94-017-9181-6_4
- Krummheuer, G. (2015). Methods for reconstructing processes of argumentation and participation in primary mathematics classroom interaction. In A. Bikner-Ahsbahr, C. Knipping, & N. Presmeg (Eds.), *Approaches to qualitative research in mathematics education: Examples of methodology and methods* (pp. 51-74). Dordrecht: Springer.
https://doi.org/10.1007/978-94-017-9181-6_3
- Lewis, C. C., Perry, R. R., & Hurd, J. (2009). Improving mathematics instruction through lesson study: A theoretical model and a North American case. *Journal of Mathematics Teacher Education*, 12, 285-304. <https://doi.org/10.1007/s10857-009-9102-7>
- Ma, L. (1999). *Knowing and teaching elementary mathematics: Teachers' understanding of fundamental mathematics in China and the United States*. Mahwah, NJ: Erlbaum.
- Metaxas, N., Potari, D., & Zachariades, T. (2016). Analysis of a teacher's pedagogical arguments using Toulmin's model and argumentation schemes. *Educational Studies in Mathematics*, 93, 383-397. <https://doi.org/10.1007/s10649-016-9701-z>
- Murata, A., Bofferding, L. Pothen, B. E., Taylor, M. W., & Wischnia, S. (2012). Making connections among student learning, content, and teaching: Teacher talk paths in elementary mathematics lesson study. *Journal for Research in Mathematics Education*, 43(5), 616-650.
<https://doi.org/10.5951/jresmetheduc.43.5.0616>

- Nardi, E., Biza, I., & Zachariades, T. (2012). 'Warrant' revisited: Integrating mathematics teachers' pedagogical and epistemological considerations into Toulmin's model for argumentation. *Educational Studies in Mathematics*, 79, 157-173.
<https://doi.org/10.1007/s10649-011-9345-y>
- Petersen, B. E. (2005). Student teaching in Japan: The lesson. *Journal of Mathematics Teacher Education*, 8, 61-74. <https://doi.org/10.1007/s10857-015-9341-8>
- Potari, D., & Psycharis, G. (2018). Prospective mathematics teacher argumentation while interpreting classroom incidents. In M. E. Strutchens, R. Huang, D. Potari, & L. Losano (Eds.), *Educating Prospective Secondary Mathematics Teachers. ICME-13 Monographs* (pp. 169-187). Cham: Springer. https://doi.org/10.1007/978-3-319-91059-8_10.
- Rasmussen, K. (2016). Lesson study in prospective mathematics teacher education: Didactic and paradidactic technology in the post-lesson reflection. *Journal of Mathematics Teacher Education*, 19, 301-324. <https://doi.org/10.1007/s10857-015-9299-6>
- Ricks, T. E. (2011). Process reflection during Japanese Lesson Study experiences by prospective secondary mathematics teachers. *Journal of Mathematics Teacher Education*, 14, 251-267.
<https://doi.org/10.1007/s10857-010-9155-7>
- Saldaña, J. (2013). *The coding manual for qualitative researchers* (2nd ed.). London: Sage.
- Sawyer, A. G., Dredger, K., Myers, J., Barnes, S., Wilson, R., Sullivan, J., & Sawyer, D. (2020). Developing teachers as critical curators: Investigating elementary preservice teachers' inspirations for lesson planning. *Journal of Teacher Education*, 71(5), 518-536.
<https://doi.org/10.1177/0022487119879894>

- Seino, T., & Foster, C. (in press). Analysis of the final comments provided by a knowledgeable other in lesson study. *Journal of Mathematics Teacher Education*
<https://doi.org/10.1007/s10857-020-09468-y>
- Simpson, A. (2015). The anatomy of a mathematical proof: Implications for analyses with Toulmin's scheme. *Educational Studies in Mathematics*, 90, 1-17.
<https://doi.org/10.1007/s10649-015-9616-0>
- Skultety, L., González, G., & Vargas, G. (2017). Using technology to support teachers' lesson modifications during lesson study. *Journal of Technology and Teacher Education*, 25(2), 185-213. <http://www.learntechlib.org/p/172139/>
- Steele, M. D. (2005). Comparing knowledge bases and reasoning structures in discussions of mathematics and pedagogy. *Journal of Mathematics Teacher Education*, 8, 291-328.
<https://doi.org/10.1007/s10857-005-0854-4>
- Suh, J., & Seshaiyer, P. (2015). Examining teachers' understanding of the mathematical learning progression through vertical articulation during lesson study. *Journal of Mathematics Teacher Education*, 18, 207-229. <https://doi.org/10.1007/s10857-014-9282-7>
- Takashi, A., & McDougal, T. (2016). Collaborative lesson research: Maximizing the impact of lesson study. *ZDM Mathematics Education*, 48, 513-526. <https://doi.org/10.1007/s11858-015-0752-x>
- Thompson, C., McCaughan, D., Cullum, N., Sheldon, T.A., & Raynor, P. (2004). Increasing the visibility of coding decisions in team-based qualitative research in nursing. *International Journal of Nursing Studies*, 41(1), 15-20. <https://doi.org/10.1016/j.ijnurstu.2003.03.001>
- Toulmin, S. (1958). *The uses of argument*. Cambridge: Cambridge University Press.

- Toulmin, S. (2003). *The uses of argument (updated edition)*. Cambridge: Cambridge University Press.
- van Es, E. A. (2011). A framework for learning to notice student thinking. In M. G. Sherin, V. R. Jacobs, & R. A. Philipp (Eds.), *Mathematics teacher noticing: Seeing through teachers' eyes* (pp. 134-151). New York: Routledge.
- Verheij, B. (2005). Evaluating arguments based on Toulmin's scheme. *Argumentation*, 19, 347-371. <https://doi.org/10.1007/s10503-005-4421-z>
- Warren, J. E. (2010). Taming the warrant in Toulmin's model of argument. *The English Journal*, 99(6), 41-46. <https://www.jstor.org/stable/20787665>
- Widjaja, W., Vale, C., Groves, S., & Doig, B. (2017). Teachers' professional growth through engagement with lesson study. *Journal of Mathematics Teacher Education*, 20, 357-383. <https://doi.org/10.1007/s10857-015-9341-8>