PROOF STRUCTURE IN THE CONTEXT OF INQUIRY BASED LEARNING

Alyssa Eubank
Shawn Garrity
Todd A. Grundmeier

California Polytechnic State University, San Luis Obispo

Data was collected from three sections of an introductory proofs course that was taught from an inquiry-based perspective. Inquiry-based learning (IBL) gives authority to students and allows them to present to their peers, rather than having the instructor be the focus of the class and authority on proof. Data from the final exams of 68 students was analyzed with a focus on proof structure. Proofs chosen to analyze included concepts considered "prior knowledge", as well as problems that required new concepts from class. This research utilizes an adaptation of Toulmin's method for argumentation analysis. Our goal was to compare the proof structures generated by these students to previous research also applying some form of Toulmin's scheme to mathematical proof. There was significant variety of proof structures, which could be a result of the IBL atmosphere.

Keywords: Proof, Inquiry-Based learning, Undergraduates, Structure, Toulmin

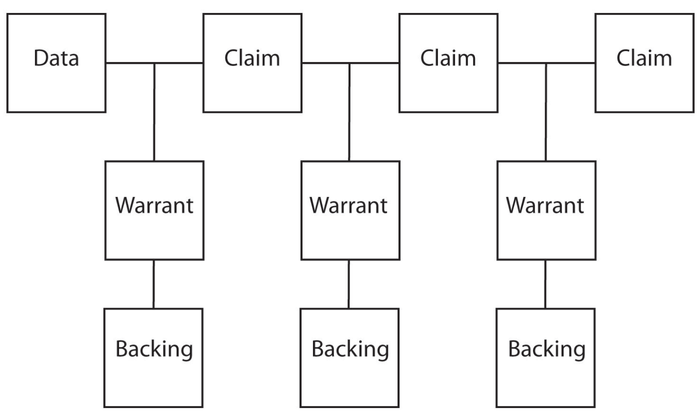
Stephen Toulmin analyzed argumentation and developed a new approach to formal logic. He classified six interrelated components he believed essential in constructing a sound and convincing argument: claim, grounds (data), warrants, backing, rebuttal, and modal qualifiers. In Toulmin's (1979) work, an argument was defined as "the sequence of interlinked claims and reasons that, between them, establish the content and force of the position for which a particular speaker is arguing" (p.13). This definition differs from that of mathematical proof. Argumentation relies on making claims and then justifying, while mathematical proof relies on making inferences of previous results to come to a claim (Barrier, Mathe, & Durand-Guerrier, 2009). In other words, argumentation relies on the content of each claim and proof relies on the function of each claim. (In the context of this paper, we use mathematical argument and proof to mean the same thing). Though Toulmin did not devise his scheme from the perspective of mathematical proof, it is valuable in this context because of the many parallels between argumentation and proof.

Many mathematics education researchers agree that a formal proof contains elements identified by Toulmin, as evidenced by the various proof analysis schemes adapted from Toulmin's work. Though much merit is given to Toulmin's scheme, it is often the case that a *restricted* version is applied in the context of proof. However, Inglis, Mejia-Ramos, and Simpson (2007) argue that you need Toulmin's *complete* model to analyze proof (i.e. consider all six elements). In their research, an interviewer interacted with a student in order to understand the process by which they came to their conclusion. This interaction enabled researchers to witness corrections of mistakes and possible uncertainty of each student. In this case, the *complete* scheme can be considered necessary. On the contrary, a researcher analyzing written proof only witnesses the final stage of the student's thought process; students have effectively already qualified their statements and considered plausible rebuttals and do not present any uncertainty. Hence qualifiers are no longer being stated, a *restricted* scheme is sufficient.

Toulmin's criteria for the structure of argumentation have been used in various contexts of mathematics education research: traditional lecture-based classrooms (Fukakawa-Connely, in progress), interviews (Inglis, Mejia-Ramos, & Simpson, 2007), and classroom discussions (Krummheuer, 2007). Toulmin's scheme has also been applied to everyday argumentation, such as that found in the workplace (Simosi, 2003). One common finding in this line of research is a lack of warrants and backing within an argument or proof. While the structure of mathematical proof and argumentation has been explored in varied contexts, little or no research exists about the structure of student proof in the context of an IBL mathematics course. This work attempts to fill this gap in the literature.

After careful consideration of Toulmin's (1979) definitions and those of past research on mathematical proof, we agreed on definitions to classify statements in student proof and scheme for coding proofs. The only major shift from past work is related to *qualifiers*. In this paper we propose that, by Toulmin's (1979) definition, a *qualifier* is not directly associated with proof and hence suggest a new definition to use within the context of proof. According to Toulmin, *qualifiers* determine the strength of an argument in that they restrict the situation in which the final claim is true. In a mathematical proof, the final claim should be true in every situation, but its validity can rely on sub-cases within the proof. Hence, we chose to identify each sub-case within a proof as a *qualifier* because the arguments that follow are limited to that specific case. After developing our scheme, we applied it to 16 proofs to see if any adjustments needed to be made before finalization. When coding the remaining proofs, we found that new situations arose and the coding scheme evolved. We took note of every evolution, and had to go back to previously coded work to apply the most recent coding scheme. The end result was a scheme that exhausted the coding of the 136 proofs.

The proofs of the statements in the coded problems required several steps. Thus proofs consisted of a string of claims, each with its own warrants and backing. For every student's proof, the arguments were mapped using a similar schematic to one often seen in research of this nature.



Two proofs from each final exam were coded. One proof was related to "past knowledge" as it pertained to a statement about divisibility (referred to as the Integer problem). The second proof was related to functions being 1-1 and onto both ideas that were new to students in this course (referred to as the Functions problem). After mapping each proof the following characteristics were recorded: length, existence of warrants, existence of backing, floaters, qualifiers, incorrect statements and incorrect implications. This presentation will focus on the structure codes of length, warrants, backings, floater and qualifiers in these proofs. We will also compare this coding to instructor grading of the coded proofs.

Length refers not to the number of words used in each proof, but rather to the number of steps taken. Thus the total number of claims was counted. When recording length, a designation of *short* (S – one or two claims), *average* (A – three or four claims), or *long* (L – five or more claims) was assigned to each proof. If a student did not write any claims, then the proof was called "other." With regards to warrants each proof was identified as *complete* (*c* – warrants given for 100% of the claims), *most* (*m* – more than 50%, but not all, of the claims are warranted), *limited* (*l* – at most 50% of the claims are warranted), or *none* (*n* – 0% of the claims are warranted). With regards to backing each proof was given a designation of *complete* (*c* –backings supplied for every warrant), *limited* (*l* – some backings provided, but not for every warrant), or *none* (*n* – no attempted backing). This coding code resulted in a three-letter designation for each proof. For example, a proof with *average* length, *complete* warrants, and *limited* backing would be given a designation of Acl.

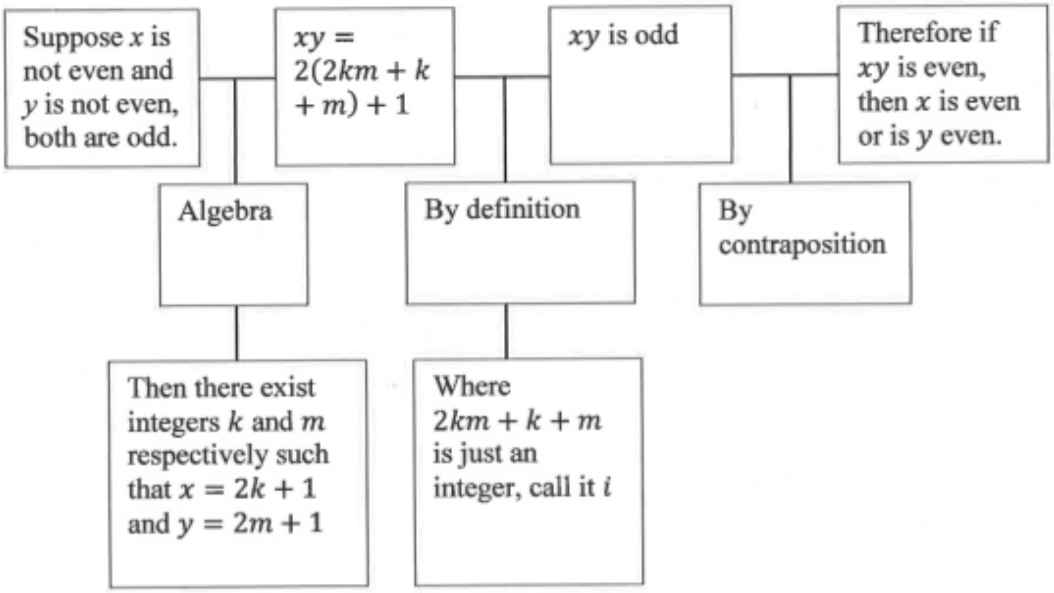
Since the coding scheme constructed for this project was based on proof length, warrants, and backings, there were 28 potential codes. 20 of these arose in the coding of this particular data set of proofs. When looking across the five most frequently used codes for Integer and Function problems, there were three common codes: Acl, Aml, and Lml.

Integer Problem						
Structure Code	Acl	Acc	Amc	Aml	Lml	
Frequency	21	10	9	9	5	

Function Problem						
Structure Code	Scc	Acl	Lml	Lcl	Aml	Lcc
Frequency	11	9	7	6	5	5

All Problems						
Structure Code	Acl	Aml	Scc	Acc	Lml	
Frequency	30	14	13	12	12	

The most common code was Acl. This was often a proof that was correct, but the very last step was not given any backing. An example of the coding of such a proof follows.



To prove the proposition, many students chose to use contraposition. For the final claim that the proposition is true, many of the students whose proofs were coded as Acl stated that they used contraposition as their warrant but did not give backing. In this case, the backing would have been writing out that the contrapositive is true.

With regards to length using three or four claims in the proof was deemed *average* (A). Here the term *average* is used somewhat loosely, for it refers to the most common length and not to the actual average number of claims. A proof with more than four claims was deemed *long* (L), and a proof with less than three claims was deemed *short* (S). It was discovered that the Function proofs have much more variety in length than the Integer proofs. The vast majority (76.5%) of the proofs in the Integer category were *average* length. On the other hand, even though the Function category had more *average* proofs than any other length, the percentages of each length are much closer. In both categories, *long* proofs were more common than *short* proofs. The *long* proofs had the highest overall average score, with 8.46 out of 10. The *short* proofs had the lowest average score (4.68), and the *average* proofs had an average score of 7.55.

Providing warrants for claims is one of the most important parts of a proof. It was thus encouraging to see that every coded proof had some kind of warrant. used, since every proof had at least one of its claims supported. Even further, in both Integer and Function categories, the majority of the proofs were *completely* warranted. While the numbers of proofs with *complete*(c) warrants were similar between the two categories (37 and 36, respectively), there were large differences in the distributions of the *most*(m) and *limited* (l) proofs. The Integer category has 23 more *m* proofs than *l* proofs, whereas in the Function category there is a difference of only 8. As expected, the average score of proofs with *complete* warrants was higher than that of the other proofs, however the difference is very small and there does not appear to be any relationship between score and proportion of warrants

As with warrants, all Integer proofs had some kind of backing, which gives validation to the warrants. The Function proofs, however, had 10 proofs with no backing present. In contrast to the results found concerning warrants, *complete* backing was not the most common occurrence. In both Integer and Function proofs there were more *limited* proofs than *complete* proofs. This most common code (*l*) also had the highest average score overall (7.64). The *completely* backed proofs, on the other hand, had a slightly lower average score (7.43), which was somewhat unexpected. However, the proofs with no backing still had the lowest average score (5.3).

Of the 136 coded proofs, only 15 proofs contained a *floater*. Recall that *floaters* are unnecessary pieces of information that do not follow from or directly connect to the logical sequence of the main proof. Twelve of these proofs were found in the function category. The distribution of codes associated with proofs containing *floaters* was somewhat varied. The most common structure code containing *floaters* was Acl (4) and the only other category with more that one floater was Lmc (2).

Recall that *qualifiers* identify sub-cases. Ten student proofs contained *qualifiers*, all of which were in the Integer category. All but two of these proofs were *long*; the other two were *average*. The proofs that contain *qualifier*s were widely distributed among the structure codes, but the most common was Lml (3) and two other categories had more than one proof that contained qualifiers, Lcl (2), Lmc (2).

A large part of this IBL classroom was student presentation and peer collaboration. It could be expected that this level of collaboration would influence students to approach and prove statements with similar structure, but the data from this project seems to provide evidence to the contrary. It was found that there was not a consistent proof structure across all analyzed proofs. The most common structure code appeared only 30 times (22.1% of all proofs) and the next most common code only appeared 14 times (10.3% of all proofs). This lack of consistency may be due to the fact that the class did not have one consistent person modeling formal mathematical proof but instead an entire classroom of presenters.

In Fukakawa-Connely's (in progress) work, it was found that in a traditional, lecture-based classroom, students modeled their proofs after an authoritative figure (i.e. the instructor). It may be the case that the IBL students did not model their proofs after their figures of authority because they did not fully trust each student presenter's mathematical competency. This lack of trust would likely force students to critically think and assess the validity of each proof instead of trusting that the professor is correct. Also, as students are being exposed to the many different proof structures provided by various presenters, they are able to judge which proof structure makes the most sense and works best for them. The wide variety of proof structures identified in this research analysis shows that IBL classrooms facilitate a flexible environment that encourages student creativity within formal proof. The fact that average scores were not drastically different between proof structures also supports this.

There was a wider variety of proof structures among student proofs in this research than we anticipated. This may suggest that students in the IBL class learned to take responsibility for their proof style and not solely rely on the authority of the instructor. Since every student provided some level of warrant, students seemed to understand the importance of justification in mathematical proof. Backing, on the other hand, was less common. It was found that lack of backing did not significantly affect score, which may mean that implicit backing is acceptable in mathematical proof and not simply in oral argumentation. Would applying our coding scheme to problems from lecture-based classes or different IBL classes yield similar results?

References:

Barrier, T., Mathe, A. C., & Durand-Guerrier, V. (2009). Argumentation and Proof: A Dsicussion About Toulmin's and Duval's Models. *CERME 6* (pp. 191-200). Lyon France: INRP.

Fukakawa-Connely, T. (in progress). Using Toulmin Analysis to Link an Instructor's Proof-Presentation and Student's Subsequent Proof-Writing Practices.

Inglis, M., Mejia-Ramos, J., & Simpson, A. (2007). *Modelling Mathematical Argumentation: The Importance of Qualification.* New York: Springer.

Krummheuer, G. (2007). *Argumentation and Participation in the Primary Mathematics Classroom.* Orlando, Florida: Elsevier.

Simosi, M. (2003). *Using Toulmin's Framework for the Analysis of Everday Argumentation: Some Methodological Considerations.* Netherlands: Kluwer Academic Publishers.

Toulmin, S., Rieke, R. D., & Janik, A. (1979). *An Introduction to Reasoning.* New York: Macmillian.