

CONSTRUCTING COMMUNAL CRITERIA FOR PROOF THROUGH CRITIQUING THE REASONING OF OTHERS

Sean P. Yee

University of South Carolina
yee@math.sc.edu

Yi-Yin (Winnie) Ko

Indiana State University
Winnie.Ko@indstate.edu

Justin D. Boyle

University of Alabama
jboyle@bamaed.ua.edu

Sarah K. Bleiler-Baxter

Middle Tennessee State University
Sarah.Bleiler@mtsu.edu

Fifty-seven students in mathematical content and secondary mathematics methods courses from four U.S. universities participated in an instructional sequence to generate communal criteria defining mathematical proof within their respective classrooms. Participants completed a proof-related task before class, worked together in small groups to evaluate instructor-selected arguments, communally agreed upon criteria for evaluating a proof based on their evaluations, and then revised their original argument to meet the communal criteria after class. Similar criteria were constructed across the four classrooms. Moreover, the four authors coded and compared students' initial and revised arguments with respect to proof schemes to identify specific shifts in students' work after the instructional sequence. Results indicated a majority of students' proof schemes changed in their revised argument with specific trends aligning with their class-based criteria for proof.

Keywords: Reasoning and Proof, Post-Secondary Education, Classroom Discourse, Teacher Education-Preservice

Current reforms have called for a stronger emphasis on mathematical reasoning and proving in school mathematics (National Governors Association Center for Best Practices & Council of Chief State School Officers [NGA & CCSSO], 2010; National Council of Teachers of Mathematics [NCTM], 2000). The challenge of meeting this recommendation is that students typically observe their teacher presenting a polished and completed proof (Stylianou, Blanton, & Knuth, 2009). Using this method of proof instruction suggests to students that their teacher is the primary authority to judge the validity of their proof (Harel & Sowder, 1998). Students then believe that the goal of learning proofs involves strategies that replicate similar problems (Bleiler, Ko, Yee, & Boyle, 2015). Such instructional methods complicate students' ability to judge the validity of mathematical arguments because students are not given the opportunity to "construct viable arguments and critique the reasoning of others" (NGA & CCSSO, 2010, p. 6), a necessary standard for mathematical practice. Therefore, it is not surprising that research has shown that students at all grade levels have considerable difficulty in judging the validity of a proof (Healy & Hoyles, 2000; Ko & Knuth, 2013; Segal, 1999; Weber, 2010).

Although students' difficulties with proof are well documented, to date, little research has explicitly articulated a classroom community's criteria for mathematical proof, including social dimensions necessary for proof validation. Bleiler, Thompson, and Krajčevski (2014) explain that "many researchers have called for the explicit instruction to proof validation to ameliorate common difficulties related to accepting inductive arguments or focusing on local rather than global elements of an argument" (p. 109). Concomitantly, Hanna and Jahnke (1993) remind us that "the social process of verification through which proof becomes accepted in the mathematical community should somehow be imitated in the school" (p. 433). To address this research gap, we designed an intervention where students: (1) constructed an initial argument justifying their generalization to the Sticky Gum Problem (SGP) depicted in Figure 1 before class; (2) critiqued instructor-selected

arguments in small groups and developed a list of five communal criteria necessary for writing proofs during class; and (3) revised their initial argument to the SGP to meet the classroom-based criteria for proof after class described in Table 1. In this paper, we investigated how students' written arguments differed before and after our instructional sequence. To answer this research question, we used Harel and Sowder's (2007) proof schemes to identify how the students' arguments shifted given their engagement in the instructional sequence.

The Sticky Gum Problem	
Ms. Hernandez came across a gumball machine one day when she was out with her twins. Of course, the twins each wanted a gumball. What's more, they insisted on being given gumballs of the same color. The gumballs were a penny each, and there would be no way to tell which color would come out next. Ms. Hernandez decides that she will keep putting in pennies until she gets two gumballs that are the same color. She can see that there are only red and white gumballs in the machine.	
<ol style="list-style-type: none"> 1) Why is three cents the most she will have to spend to satisfy her twins? 2) The next day, Ms. Hernandez passes a gumball machine with red, white, and blue gumballs. How could Ms. Hernandez satisfy her twins with their need for the same color this time? That is, what is the most Ms. Hernandez might have to spend that day? 3) Here comes Mr. Hodges with his triplets past the gumball machine in question 2. Of course, all three of his children want to have the same color gumball. What is the most he might have to spend? 4) Generalize this problem as much as you can. Vary the number of colors. What about different size families? Prove your generalization to show that it always works for any number of children and any number of gumball colors. 	

Figure 1. The Sticky Gum Problem originally created by Fendel, Resek, Alper, and Fraser (1996).

Table 1: Instructional Sequence to Create Communal Criteria to Understand Proof

Before-Class Activity	<ul style="list-style-type: none"> • Each student solves the Sticky Gum Problem (SGP) and submits the SGP electronically prior to the before-class activity. • The instructor chooses five distinct arguments submitted by their students.
During-Class Activity	<ul style="list-style-type: none"> • Students break into small groups of 2-4 students. Each group (1) looks at the same five instructor-selected classmate's arguments; (2) discusses and decides which of the five selected arguments are mathematical proofs; (3) determines how they decided whether each argument is a mathematical proof; and (4) creates a list of five proof criteria through small and whole group discussions. • All small groups rejoin the entire class. The entire class has discussions, compares the small group's criteria, and determines a class-wide communal list of five criteria for proof.
After-Class Activity	<ul style="list-style-type: none"> • Each student revises their original argument for SGP to satisfy the communal class-based criteria and submits their revised argument.

Theoretical Framework

To investigate how developing communal criteria for proof affected students' written arguments, this study builds on the frameworks about proof in the mathematics classroom (Stylianides, 2007), the process of proving (i.e. ascertaining and persuading, Harel & Sowder, 1998, 2007), and proof schemes (Harel & Sowder, 1998, 2007).

Proof in the Mathematics Classroom

A traditional perspective of proof is "a formal and logical line of reasoning that begins with a set of axioms and moves through logical steps to a conclusion" (Griffiths, 2000, p. 2). This definition describes a formal individual process of writing a proof, but does not acknowledge the negotiation aspects of constructing and validating proofs within a mathematics community. Stylianides (2007) suggested that proof is a connected sequence of assertions because it

- (1) uses statements accepted by the classroom community (set of accepted statements) that are true and available without further justification;
- (2) employs forms of reasoning (modes of argumentation) that are valid and known to, or within the conceptual reach of, the classroom

community; and (3) is communicated with forms of expression (modes of argument representation) that are appropriate and known to, or within the conceptual reach of, the classroom community. (p. 291)

Stylianides created this definition emphasizing proof that balances mathematics as a discipline and a learning tool. We use Stylianides' definition of proof as a frame of reference to interpret students' characteristics of proof and their classroom-based criteria for proof.

Ascertaining and Persuading

To understand how a mathematics community can support students' understanding of what counts as proof, it is also necessary to consider individuals' ideas of what constitutes a proof during the proving process. To incorporate both the individual and the communal aspects of proving, we use Harel and Sowder's (2007) action of proving as "the process employed by an individual (or a community) to remove doubts about the truth of an assertion" (p. 808). Furthermore, Harel and Sowder (1998, 2007) indicated proving as ascertaining (convincing oneself) and persuading (convincing one's community). Weber's (2010) research found that undergraduate mathematics majors would judge a deductive argument as proof, even if they were not convinced by its argument or the argument was logically flawed. Both Weber (2010) and Segal (1999) emphasize the need to push individual's convictions into public discussion, which was an aim of our instructional sequence. When students have to judge the validity of an argument, they are forced to negotiate between ascertaining and persuading because the judgement is based on how the argument would convince themselves and others. To this end, we designed our instructional sequence around students judging the validity of instructor-selected arguments, which allowed them to use their understanding of what counts as proof to create a discursive space and reach consensus on their judgements.

Proof Schemes

As the purpose of this study is to examine the effect that our instructional sequence would have on students' arguments, we use Harel and Sowder's (2007) proof schemes to analyze students' work because a proof scheme consists of "what constitutes ascertaining and persuading for that person (or community)" (p. 809). Harel and Sowder divided proof schemes into three main categories, which are external conviction proof schemes, empirical proof schemes, and deductive proof schemes. External conviction proof schemes mean that conviction is made by the form of an argument, such as a teacher, a textbook, or symbolic representations. In the classroom external convictions often take the form of an authoritarian proof scheme (Harel & Sowder, 2007). For empirical proof schemes, the validity of a conjecture is made by specific cases or individuals' mental images. Regarding deductive proof schemes, a conjecture's veracity is determined by logical deduction. Harel and Sowder (1998) make a strong point "that these schemes are not mutually exclusive; people can simultaneously hold more than one kind of scheme" (p. 244). For that reason, it is important for us to look for overlaps within the three proof schemes when analyzing our data.

Harel and Sowder's (1998) original proof-scheme taxonomy aligns with the focus of this study:

Our definitions of the process and proving and proof scheme are deliberately psychological and student-centered... Thus, this classification is not of proof content or proof method... In contrast, it is the individual's scheme of doubts, truths, and convictions, in a given social context, that underlies our characterization of proof schemes. (p. 244)

In the same ilk, the focus of this research emphasizes the "social context" of the classroom. Hence, we use proof scheme categorizations as comparative tools to look for changes in students' initial and revised arguments. We are not trying to evaluate the correctness of students' arguments but rather assign a qualitative descriptor of students' justification. This allows us to gain evidence of students'

relative convictions (Weber & Mejia-Ramos, 2015) of what constitutes mathematical proof before and after the instructional sequence.

Method

Participants

This study was conducted across four separate courses across four institutions where the authors were the instructors. Two of the courses (Authors 1 and 4) were mathematical content courses, and the other two (Authors 2 and 3) were secondary mathematics methods courses. Concomitantly, three authors were within mathematics departments, and one author was in an education department. There were a total of 57 undergraduate and graduate mathematics or secondary mathematics education majors participating in the study. All of the participants either had completed calculus courses and an introductory proof course or were presently in an introductory proof course.

Design and Implementation of the Instructional Sequence

The instructional design was organized around the SGP (see Figure 1) into a before-class activity, during-class activity, after-class activity (see Table 1). The participants submitted their solutions to the SGP prior to the during-class activity. The four instructors reviewed the arguments submitted by the students and selected five arguments from each class to have the students evaluate with the during-class activity. Arguments were chosen for their diversity in mode of argumentation, mode of argument representation, and sets of accepted statements (Stylianides, 2007). Simultaneously, the four authors selected arguments that would illustrate differing modes of argumentation to promote discourse around what counts as proof.

Students individually evaluated the five selected arguments in class and then worked in small groups to negotiate and come to a consensus on which arguments they believed to be valid proofs. This negotiation provoked important conversation about what defines a proof because negotiations highlighted if the argument persuaded the students, aligned with how the students had ascertained justification, or both. The students were then instructed to articulate three to seven criteria for determining arguments to be proofs. After each small group created their list, the whole class discussed and agreed upon five appropriate criteria for proof in their classroom. Each of the four classes had some differing criteria, but their criteria all showed that a proof needs to be generalizable and have logical structure (see Boyle, Bleiler, Yee, & Ko, 2015 for a more detailed discussion of the communal criteria). An important note is that these two categories, generalizability and logical structure, are similar to what Harel and Sowder (2007) labeled as necessary characteristics of the transformational deductive proof scheme, generality and logical inference.

Data Collection and Analysis

All four authors collected their students' work and class artifacts throughout the instructional sequence. Each student's initial argument and revised argument were coded according to Harel and Sowder's (1998, 2007) proof schemes to identify their changes. As discussed in the theoretical framework section, Harel and Sowder suggested that overlaps of the categories of external conviction, empirical, and deductive proof schemes were to be expected. Thus, three more categories, "empirical/external conviction, external conviction/deductive, and empirical/deductive," were developed and included in our data analysis. We provide Gabriel's work depicted in Figure 2 as an example of the need for the overlap in coding.

Figure 2 is Gabriel's initial argument coded as Empirical/Deductive. It could be easy to classify Gabriel's argument with an empirical proof scheme because of the justification using one example with three children and three colors. However, there are elements of the deductive proof scheme within his work as well. Gabriel generalized the argument with statements, such as "worst case

scenario” and the concluding remark, “if all but one child have all the colors of gum, any color will ensure they all have the same.” These statements reference the restrictive characteristics of the transformational proof scheme (Harel & Sowder, 1998), a subset of the deductive proof scheme. Gabriel presumed to understand the restrictions that transform the situation by assuming the worst case scenario and discussing what transformation would be necessary to satisfy the last child receiving the same color. Thus, Gabriel’s argument was coded as Empirical/Deductive. It is important that the proof schemes are not evaluating the correctness of students’ arguments but rather than describing what characteristics students communicate with their justifications.

The formula for maximum cents spent is as follows:
 $k = \text{number of children}$
 $c = \text{number of colors} \rightarrow c(k-1) + 1 = \text{max } \$ \text{ spent}$
 I came to this by writing out the worst case scenarios. For example 3 children 3 colors looks like
 red white blue
 red white blue
 any color
 This shows that if all but one child have all the colors of gum, any color will ensure they all have the same. Thus $\text{colors}(\text{children}-1) + 1$

Figure 2. Gabriel’s original argument classified via empirical and deductive proof schemes.

All four authors met and agreed upon the structure and purpose of the six categorizations, and evaluated 15 randomly-selected arguments to make sure that there was strong interrater reliability. Authors then all double-coded the remaining arguments. There were a total of 54 original and 46 revised arguments, and each of them was coded by at least two authors. Any varying proof schemes were discussed amongst the authors who coded the proof scheme to find consensus. Once all arguments had been coded and consensus had been reached amongst the codes, the authors compared original and revised proof schemes amongst all students.

Results

Table 2 illustrates the coding of each student’s before-class argument (rows) and after-class argument (columns). Colors have been added to illustrate blends. For example, empirical is red and external conviction is yellow, so empirical/external conviction is orange. Each cell in Table 2 illustrates the number of students that transitioned from one proof scheme to another. A row or column identified as “None” means that the students offered no argument.

Table 2: Proof Schemes of all original and revised arguments

Before-Class Proof Schemes (rows) Revised Proof Schemes (columns) TOTAL	Revised None	Revised Empirical	Revised Empirical/External Conviction	Revised External Conviction	Revised External Conviction/Deductive	Revised Deductive	Revised Empirical/Deductive	Original Argument Total
Original None	0	2	0	0	0	1	0	3
Original Empirical	5	5	3	6	0	6	2	27
Original Empirical/External Conviction	1	0	0	0	0	3	0	4
Original External Conviction	2	1	1	3	0	1	0	8
Original External Conviction/Deductive	0	0	0	0	0	2	0	2
Original Deductive	1	0	0	1	0	4	0	6
Original Empirical/Deductive	2	0	0	0	1	3	1	7
Revised Argument Total	11	8	4	10	1	20	3	57

The main diagonal of Table 2 shows that five empirical proof schemes remained empirical (row2, column2), three external proof schemes remained external convictions (row4, column4), four deductive proof schemes remained deductive (row6, column6), and one empirical/deductive remained empirical/deductive. Altogether, 13 of the 57 participants kept with the same proof scheme after the instructional sequence, demonstrating that the majority of students constructed arguments that provided evidence of altered proof schemes after the instructional sequence. Looking along the diagonal also reveals that no arguments that was coded initially with an overlapping code (e.g. external/deductive), kept the same proof scheme after being revised except for one student whose original argument and revised argument were coded as empirical/deductive.

Table 2 also shows that the number of strictly empirical proof schemes declined from 27 to 8 while the number of strictly deductive proof schemes grew from 6 to 20 after our instructional sequence. For instance, Kimmie’s original and revised argument depicted in Figure 3 is an example of this transition. Kimmie’s original argument bases its justification around the given information and justifies the general rule using examples. For her revised argument, Kimmie uses generality and logical inference, which are two aspects of Harel and Sowder’s (2007) deductive proof schemes. Kimmie’s work illustrates how her argument shifted from an empirical proof scheme to a deductive proof scheme after participating in the instructional sequence.

Finally, the external conviction proof schemes increased slightly from 8 to 10 after the instructional sequence. When looking closely at arguments demonstrating external conviction, the students seemed to tend to replicate a method that had been used in one of the five instructor-selected arguments (e.g., mathematical induction on both the number of children and the colors of bubble gum) to construct their revised arguments. More specifically, those students might mimic one of the instructor-selected arguments without being able to fully justify why their generalization to the SGP is always true.

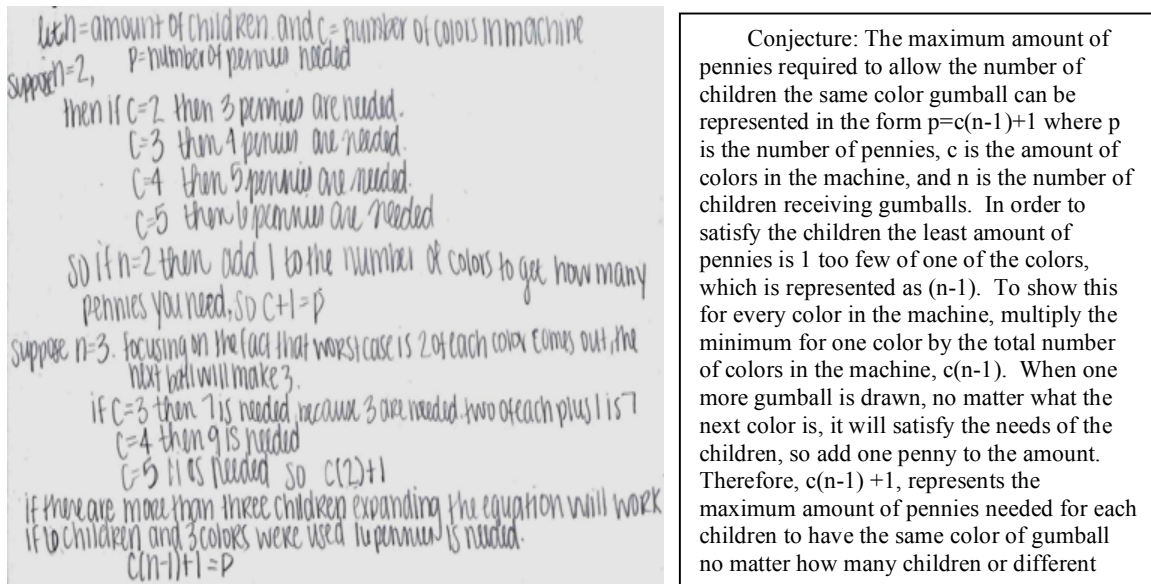


Figure 3: Kimmie's initial (hand written) and revised (typed) arguments coded as empirical and deductive, respectively.

Discussion

This study challenges the current borders surrounding students' classroom engagement with constructing and critiquing arguments. Our instructional sequence gives students opportunities to develop criteria for proof through evaluating, discussing, and negotiating what counts as proof. In addition, these results illustrate that the instructional sequence allowed students to judge the validity of arguments, empowering students to generate criteria from those judgements to influence what constitutes as ascertaining and persuading (i.e., Harel & Sowder's proof scheme, 2007) in their classroom community. Thus, this instructional sequence can aid teachers in developing what constitutes a proof through evaluating arguments, as well as discussing and negotiating their classroom-based criteria for proof.

By using Harel and Sowder's (1998) lens of proof schemes, our results show that a majority of students changed their argument after the instructional sequence. Of those changed, the number of empirical proof schemes decreased while the number of deductive proof schemes increased. Also, the number of external conviction proof schemes remained about the same. The decrease in empirical proof schemes and increase in deductive proof schemes might appear to align with the communal criteria for proof students developed, because the two of the criteria shared by all four classes—logical structure and generalizability—are the characteristics of Harel and Sowder's (2007) transformational deductive proof schemes, logical inference and generality.

When evaluating the validity of arguments, Weber and Mejia-Ramos (2015) indicate that students' convictions are often not absolute but relative, meaning that writing an argument is more than justifying one's beliefs. The research described in this paper did not measure the mathematical validity of students' initial and revised arguments. However, we identified characteristics of student justifications with respect to Harel and Sowder's (2007) proof schemes, because our study desired to determine how engaging in our instructional sequence would affect students' revised arguments. When implementing our instructional sequence, one should not see the communal criteria as an absolute final product, but they serve as a means to improve students' understanding of proof and their abilities to write proofs. Thus, we see the generation of classroom-based criteria is not limited to assessment of learning or assessment for learning, but it also serves an assessment *as* learning (Bleiler, Ko, Yee, & Boyle, 2015). A valuable future research study could include multiple iterations

of this instructional sequence with various the proof-related problems to lead to a better understanding of what students have learned about proof and proving over a full semester and across different mathematical content areas.

In sum, this study offers an alternative method of proof instruction to help students take ownership of their proof construction instead of viewing their teacher as the primary authority in the classroom. Moreover, our results show a means in which to have students develop a common understanding of what counts as proof through critiquing the reasoning of others. Thus, this instructional sequence offers a social space in which teachers and students can discuss and negotiate the meaning of what constitutes a proof.

References

- Bleiler, S. K., Ko, Y. Y., Yee, S. P., & Boyle, J. D. (2015). Community development and evolution of a rubric for proof writing. In C. Suurtamm & A. R. McDuffie (Eds.), *Annual Perspectives in Math Ed 2015: Assessment to Enhance Learning and Teaching* (pp. 97-108). Reston, VA: NCTM.
- Bleiler, S. K., Thompson, D. R., & Krajčevski, M. (2014). Providing written feedback on students' mathematical arguments: proof validations of prospective secondary mathematics teachers. *Journal of Mathematics Teacher Education*, 17(2), 105-127.
- Boyle, J. D., Bleiler, S. K., Yee, S. P., & Ko, Y. Y. (2015). Transforming perceptions of proof: A four-part instructional sequence. *Mathematics Teacher Educator*, 4(1), 32-70.
- Fendel, D., Resek, D., Alper, L., & Fraser, S. (1996). *Interactive mathematics program year 1 - unit 2: The game of pig* (p. 99). Emeryville, CA: Key Curriculum Press.
- Griffiths, P. A. (2000). Mathematics at the turn of the millennium. *American Mathematical Monthly*, 1-14.
- Hanna, G., & Jahnke, H. N. (1993). Proof and application. *Educational Studies in Mathematics*, 24(4), 421-438.
- 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). Washington DC: Mathematical Association of America.
- Harel, G., & Sowder, L. (2007). Toward comprehensive perspectives on the learning and teaching of proof. In F. K. Lester (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 805-842). Charlotte, NC: Information Age Pub.
- Healy, L., & Hoyles, C. (2000). A study of proof conceptions in algebra. *Journal for Research in Mathematics Education*, 31(4), 396-428.
- Ko, Y. Y., & Knuth, E. J. (2013). Validating proofs and counter examples across content domains: Practices of importance for mathematics majors. *Journal of Mathematical Behavior*, 32(1), 20-35.
- National Council of Teacher of Mathematics (2000). *Principles and standards for school mathematics*. Reston, VA: Author.
- National Governors Association Center for Best Practices & Council of Chief State School Officers (2010). *Common Core State Standards Initiative*. Washington, DC: Author.
- Segal, J. (1999). Learning about mathematical proof: Conviction and validity. *Journal of Mathematical Behavior*, 18(2), 191-210.
- Stylianides, A. J. (2007). Proof and proving in school mathematics. *Journal for Research in Mathematics Education*, 38(3), 289-321.
- Stylianou, D. A., Blanton, M. L., & Knuth, E. J. (2009). *Teaching and learning proof across the grades: A K-16 perspective*. New York: Routledge.
- Weber, K. (2010). Mathematics majors' perceptions of conviction, validity, and proof. *Mathematical Thinking and Learning*, 12(4), 306-336.
- Weber, K., & Mejia-Ramos, J. P. (2015). On relative and absolute conviction in mathematics. *For the Learning of Mathematics*, 35(2), 15-21.