EXPANDING STUDENTS' ROLE WHEN DOING PROOFS IN HIGH SCHOOL GEOMETRY

Patricio Herbst
University of Michigan
pgherbst@umich.edu

Mollee Shultz
University of Michigan
mollee@umich.edu

Inah Ko
University of Michigan
inahko@umich.edu

Nicolas Boileau
University of Michigan
nboilea@umich.edu

Ander Erickson
University of Washington, Tacoma
awerick@uw.edu

Using mixed-effects regression, we analyzed teachers' responses to a multimedia survey of instructional practices in posing proof problems in geometry. Teachers described and rated for appropriateness three different ways of involving students in deciding what to prove, including one in which the teacher chooses the givens and the conclusion to prove, and two others that expand the students' role in different degrees. While teachers recognized the former as normative, their ratings identified an alternative as more appropriate, having more positive value, and less negative value than the normative one. This alternative has students propose the givens or the conclusion to prove, and it allows the teacher to control the complexity of instruction by endorsing one proposal before students are to write the proof.

Keywords: Instructional activities and practices, Reasoning and proof, Research methods

Proof plays many roles in mathematics; among them is that of being a method for the discovery of new knowledge (de Villiers, 1990; Lakatos, 1976). Inasmuch as mathematical knowledge is usually represented in the form of conditional propositions (i.e., a conclusion is claimed as necessarily true if certain conditions or hypotheses are taken to be true), the creation of this mathematical knowledge involves both conjecturing a conclusion that can be asserted as true and hypothesizing the conditions that need to be assumed for that conclusion to be necessitated. A conclusion may be intuited as something that is only sometimes true, and one might search for what conditions might make the conclusion is necessarily true. At times, less conditions might also be sufficient to claim the same conclusion, while at other times less conditions may only allow one to make a weaker claim. In all of that, one exercises logical deduction as the process of making a valid inference based on two or more true premises, yet one does more than that: As Lakatos (1976) explained, one engages in a heuristic process of finding out what could be reasonably true. But, while this is part and parcel of mathematical work and quite relevant to making and critiquing mathematical arguments and modeling with mathematics as expected by the Standards for Mathematical Practice (Common Core State Standards Initiative, 2010), students rarely have opportunities to engage in such work in school (Stylianides, 2009). Why would those opportunities be scarce? How could such opportunities be created?

Our work examines those questions from the perspective of the teacher's management of the complex work of having students do proofs. Our goal in this to show that, from the teacher's perspective, there is a way in which the students' share of work could be expanded without compromising the teacher's capacity to manage the work. We document below how the literature has taken care of both curricular and learning perspectives on this matter. Yet, understanding whether opportunities like that are viable also requires consideration of the demands that such work places on the teacher.

Review of Literature

The mathematical work that students do in classrooms can be understood using frameworks associated with the instructional triangle (Cohen, Raudenbush, & Ball, 2003; Herbst & Chazan, 2012). Students' interaction with content happens especially in the context of problems and tasks. The choice of such problems is made by the teacher, whose work involves having students do work that puts them in interaction with target content.

Prior research has looked at the nature of the proof tasks that are afforded to students. Herbst's (2002) historical research showed that while a wide variety of opportunities for students to do original proofs emerged in the third quarter of the 19th century, some including the search for reasoned conjectures described above, the proof exercises in use by the early 1910's already consisted of separated sets of *given* and *prove* statements provided to students. Though efforts have increased to provide more opportunities to reason and prove in a variety of ways, such as with reform-based curricula, studying one such project Stylianides (2009) found that opportunities to find patterns and pose conjectures were low compared to opportunities to engage students in providing rationales. In an analysis of six secondary geometry textbooks, Otten, Gilbertson, Males, and Clark (2014) found that only between 5% and 20% of problems involved the construction of a proof, and most of those involved proving general or particular claims rather than constructing conjectures. In contrast, in a study of grade 8 Japanese geometry textbooks, Fujita and Jones (2014) found a significant portion of exercises providing opportunities for students to conjecture and discover properties. These exploratory problems typically came at the beginning of a lesson, so that by the time students prove or justify at the end of the lesson, they had already explored and investigated the facts on their own. Cirillo and Herbst (2012) have proposed some alternative problems that could be used to engage students in an expanded scope of work, where they could either produce the conclusion of a set of givens, or the givens needed to prove a particular conclusion. We surmise that the viability of these proof problems hinges on more than having such proof problems though. In particular, one might wonder whether students are able to do such work.

The second vertex in the instructional triangle is the student whose interactions with the content happen particularly in the context of their work on tasks. The literature on students thinking and learning reveals students' potential but also their difficulties with proof. Investigations on students' thinking have revealed that students, even at the elementary level, are capable of constructing arguments and proofs (Ball & Bass, 2003; Lampert, 1992; Reid, 2002). Teaching experiments (Norton, 2008) and classroom observations (Ellis, 2007) suggest students can be engaged in making reasoned conjectures. Yet, research has also shown that students sometimes take properties to be true on account of intuition and experience, not seeing the need for proof (Chazan, 1993). This might suggest that problems in which students have to figure out what might be true, might not so easily lend themselves to engaging them in proving.

The third vertex in the instructional triangle is the teacher. One way to inquire on the viability of engaging students in better proof problems might hinge on inspecting what teachers know about proof. Knuth (2002a) found that while secondary teachers acknowledged different important purposes of proof, they failed to recognize proof as a tool for learning mathematics. They also had difficulty knowing what constitutes a robust proof, failing to recognizing non-proofs and making judgements based on the form of an argument instead of the soundness of the reasoning. In a study of the mathematical knowledge for teaching (MKT) needed to teach proof, Steele & Rogers (2012) illustrated how secondary teachers' understanding of proof affected the way the teacher positioned students - namely as creators, "but only in the sense that they

Hodges, T.E., Roy, G. J., & Tyminski, A. M. (Eds.). (2018). *Proceedings of the 40th annual meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education*. Greenville, SC: University of South Carolina & Clemson University.

provided reasons for a set of predetermined steps" (p. 175). Teachers' attitudes towards students' ability to do proofs also play a role in what proving opportunities are presented to students. We aim to investigate ways to promote the creation of these opportunities for students in a way that teachers deem appropriate.

While the literature has been progressively relying on more classroom data, it has been common to frame issues of proving in the classroom in terms of having or not having resources, be those resources curricula, abilities, attitudes, or knowledge. In our work we have been interested in addressing the problem of students' share of work from a perspective centered on the complexity of the management of classroom instruction. While a teacher may or may not have resources with which to deal with such complexity, they are likely to have means to appraise that complexity and to relate to different practices that might differ amongst themselves by the amount of such complexity. We identify one complexity associated with expanding the work of students in proving here, then we describe how we studied it.

If we start from the nature of the task, we could wonder how teachers might relate to the possibility that students might come up with the proposition to be proved. One first complexity has to do with framing the problem. A teacher is likely to need to do more than identifying the goal of the task (to prove a proposition); some specifics of the thematic territory for the proposition to be proved may need to be identified. For this reason, Cirillo & Herbst (2012) proposed problems that expected students to provide the givens or the conclusion, but providing some of those resources. A second complexity draws on Doyle's (1986) characterization of classrooms as complex environments partly on account of the simultaneity of events. This complexity points to the number of possible responses that could ensue if the teacher asked students to propose what could be the givens or the conclusion to prove: If students had to come up with givens, many students could come up with many different sets of givens. A discussion of which set of givens makes the proposition stronger could be desirable; but managing such discussion might make the work of the teacher harder, especially if students invested themselves in proving different propositions that ended up not all being equally valuable. This work might be especially difficult to manage in classrooms where the norm may be one of accepting a variety of solutions to problems.

Based on prior exploratory work (e.g., Aaron & Herbst, 2017), we conjectured that teachers can appreciate having problems in which the students' scope of work on proofs extends beyond deductive reasoning, to conjecturing a conclusion or hypothesizing the givens. We also conjectured that teachers would perceive the need to manage the complexity of the multiple responses and would appreciate the opportunity to collect the students' thoughts and endorse the proposition whose proof will be written. While such work may still maintain something of a separation between conjecturing and proving (Aaron & Herbst, 2017), it may make it manageable for a teacher to engage students in doing work that is more authentic than current work on proof.

Method

We studied this question using three sets of scenario-based instruments, all related to the hypothesized norm that in proof problems the teacher is responsible for providing the givens and the prove statement, for which we had collected empirical evidence in a pilot study (Herbst, Aaron, Dimmel, & Erickson, 2013). Each instrument consisted of four item sets and each item set included a scenario of instruction, represented using a storyboard, and questions about the actions in the scenario: Participants were asked to describe what they saw happening, then to rate the appropriateness of the teaching they saw. One set of scenarios (DP-C) included only episodes of doing proofs in which the teacher took responsibility for providing the givens and prove statement (as expected by the hypothesized norm). A second set of scenarios (DP-GP) included only episodes where the teacher allowed the students to come up with the givens or the prove statement (a breach of the hypothesized norm). And a third set of scenarios (DP-TSGP) where the teacher again allowed students to propose the givens or the prove statement but also endorsed one of those proposals as the proposition for the whole class to prove.

Open responses to all scenarios were coded in two different ways. On the one hand, they were coded for whether or not they contained evidence that the respondent recognized the teacher's enactment of the norm (in the DP-C case) or its breach (in the DP-GP and DP-TSGP cases). On the other hand, each of those descriptions were coded for the presence of positive appraisals as well as negative appraisals of what the teacher was doing in the scenario (relying on Martin & White's, 2005, appraisal theory). All those coding operations had moderate interrater reliability. Three measures were derived from such codes, which we call *norm recognition* (INR), *positive appraisal* (PA), and *negative appraisal* (NA), all of them ranging from 0 to 4 in each instrument. Additionally, participants rated each scenario for appropriateness (AT) on a scale 1-6, with 1 being very inappropriate and 6 very appropriate.

Our conjectures included that (1) scores on INR(GP) and INR(TSGP) would both be larger than INR (C), indicating that participants noticed that both sets of scenarios breached the norm, but (2) AT(TSGP) would be larger than both AT(GP) and AT(C), which would be consistent with the conjecture that teachers preferred to expand the students' scope of work if the diversity of student proposals could be made more manageable. Appraisal scores were predicted to provide additional evidence: We conjectured that (3) PA(TSGP) would be larger than PA(GP) and PA(C) while (4) NA (TSGP) would be smaller than NA(GP) and NA(C). These conjectures would align with the interpretation that teachers would see value in expanding the students' share of labor in proof problems if the complexities that ensued from such expansion could be managed. We tested these conjectures running mixed effects regression models.

Data

Data comes from a nationally distributed sample of U.S. high school mathematics teachers. Instruments were administered in 2015-2016 through the LessonSketch (www.lessonsketch.org) online platform, where they could peruse scenarios and answer questions. There were 525 participants who completed at least one of the three instruments, and 347 participants who completed all three instruments. Most of the participants who completed all three instruments were white (86%) and female (61%), which is similar to the demographics of secondary high school teachers in the US. On average, participants had been teaching secondary mathematics for 14.7 years (SD = 8.69, min = 1, max = 40).

Results

Descriptives are shown in Table 1. Of the 525 participants who completed the DP-C instrument, 83.4% ($n = 438$) recognized the compliance of the norm at least in one scenario. Of those recognizers 40.4% ($n = 177$) provided at least one positive appraisal for this compliance, while 33.3% ($n = 146$) provided at least one negative appraisal of the compliance (both groups are not necessarily disjoint as any one recognition statement could be accompanied both by positive and negative appraisals). The mean positive appraisal scores was 1.24 (SD = 0.53) and the mean negative appraisal score was 1.17 (0.43), where both measures have a possible range 0-4. In comparison, of the 395 participants who completed the DP-GP instrument, 363 (91.9%) recognized a breach of the norm. Of those recognizers 53.2% ($n = 193$) positively appraised this

Hodges, T.E., Roy, G. J., & Tyminski, A. M. (Eds.). (2018). *Proceedings of the 40th annual meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education*. Greenville, SC: University of South Carolina & Clemson University.

breach, yielding a mean positive appraisal of the breach score of 1.51 (0.75); also 53.2 % (*n* = 193) of recognizers appraised the breach of the norm negatively, yielding a mean negative appraisal of the breach score of 1.83 (0.98). Finally, of the 495 participants who completed the DP-TSGP, 444 (89.7%) recognized a breach of the norm. Of those recognizers, 53.2% (*n* = 236) positively appraised the breach, yielding a mean positive appraisal of the breach score of 1.61 (0.82); of the recognizers, also 30.6% negatively appraised the breach, yielding a mean negative appraisal score of 1.53 (0.75). These descriptives suggest that people noticed the breach of the norm more saliently than its compliance (DP-C:83.4% < DP-GP:91.9%, DP-TSGP: 89.7%), they saw more positive as well as more negative issues with merely expanding the scope of work of the students. But when considering the possibility that the teacher might control that expansion by sanctioning the proposition to be proved, positive appraisals increased and negative appraisals decreased to a level comparable to the negative appraisals of complying with the norm.

A similar tendency could be observed with the appropriateness rating scores. We examined differences in how participants responded to a breach or compliance scenario with ratings toward the low or high end of the appropriateness scale with participants who completed all three instruments and recognized the DP-GP norm (*n* = 343). When asked to rate the appropriateness of the teaching showed in the DP-C scenarios, average scores were 4.42(SD = 0.65, *n* = 343), while those average appropriateness scores were 4.56 (SD = 0.94, *n* = 343) for the scenarios that breached the norm by asking students to provide the givens or the prove statement. The average appropriateness score went up to 4.83 (SD = 0.70, *n* = 343) in the case of the DP-TSGP scenarios where in addition to expanding the students' share of work, the teacher at some point sanctioned the proposition that students would prove.

Table 1. Descriptive statistics of recognizers and appraisers for each instrument

Instrument	Obs	Recognizers	Appraisers	Mean	SD	Min	Max	% Appraisers
DP-C	525	438	POS: 177	1.24	0.53	1	4	40.41%
			NEG: 146	1.17	0.43	1	3	33.33%
DP-GP	395	363	POS: 193	1.51	0.75	1	4	53.17%
			NEG: 193	1.83	0.98	1	4	53.17%
DP-TSGP	495	444	POS: 236	1.61	0.82	1	4	53.15%
			NEG: 136	1.53	0.75	1	4	30.63%

To ascertain the significance of the differences in AT, PA, and NA scores across different instruments, we calculated a set of mixed effect regression models. Given that each participant responded to multiple instruments, average appropriateness (AT), positive appraisal (PA), and negative appraisal (NA) that come from the same participant are not independent. Therefore, we conducted mixed effect linear regression models that could account for this non-independence among the scores. In the models, the categorical variable indicating a type of instrument was entered as a fixed effect and the variable indicating a participant's ID was entered as a random effect. The analyses were conducted using the STATA statistical software with the sample of participants who responded all three instruments and recognized the DP-GP norm (*n* = 343).

Table 2. Mixed effects linear regressions of scores on a type of instrument

	Appropriateness (AT) B(SE)	Positive Appraisal (PA) B(SE)	Negative Appraisal (NA) B(SE)
Fixed effects			
DP-C	-0.41 (0.05)***	-0.24 (0.07)**	0.17 (0.07)*
DP-GP	-0.27 (0.05)***	-0.38 (0.071)***	-0.03 (0.07)
Constant	4.83 (0.042)***	2.84 (0.06)***	2.04 (0.06)***
Random effects			
Constant	0.19 (0.03)	0.52 (0.065)	0.37 (0.06)
Residual	0.40 (0.02)	0.86 (0.05)	0.90 (0.05)
N	343	343	343

Standard errors in parentheses; $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$
*reference instrument group: DP-TSGP

As shown in Table 2 (the reference group is DP-TSGP), results generally support our conjectures. AT (appropriateness) for TSGP is higher than both AT(DP-GP) and AT(DP-C), AT is significantly lower for DP-C by about 0.41 and DP-GP by about 0.27 (in a scale 0 ~ 6). In addition, the variation associated with the participants explains about 32% of the total deviations from the predicted AT that are not due to a type of instrument. Similarly, PA is significantly lower for DP-C and DP-GP by about 0.24 and 0.38 (in a scale 0 ~ 4) than for TSGP, respectively. The participants random effect comprise about 38% of the total residual variance. NA for DP-TSGP yields significantly higher score than NA for DP-C, but it is not significantly different from NA for DP-GP. For the NA score, the participants random effect explains approximately 29% of the total residual variance.

Conclusion

The data suggests that teachers do recognize the norm that the teacher will provide the givens and the prove for proof problems, yet they do not appraise it as better than some of the alternatives presented. Instead, alternatives in which the teacher expands students' role by inviting them to propose the givens or the statement to prove are appraised as more highly positive. Apprehension for these kinds of proof problems is apparent in the fact that negative appraisals of these alternative kinds of problems are still higher than for the habitual proof problems and not significantly higher than the negative appraisals for instances of doing proofs in which students are free to prove whatever they decide. An important implication for practice of these results is that it suggests that inservice teacher education could focus on helping teachers anticipate what students could propose in response to problems such as those proposed by Cirillo and Herbst (2012) and in practicing how to bring the class to a consensus on what statement they all should be working on.

Acknowledgements

This work has been done with support of NSF grant DRL-0918425. All opinions are those of the authors and do not necessarily represent the views of the National Science Foundation.

References

Aaron, W. R. & Herbst, P. G. (2017, online first). The teacher's perspective on the separation between conjecturing and proving in high school geometry classrooms. Journal of Mathematics Teacher Education. doi 10.1007/s10857-017-9392-0

Ball, D. L. & Bass, H. (2003). Making mathematics reasonable in school. In J. Kilpatrick, W. G. Martin, & D. Schifter (Eds.), A research companion to principles and standards for school mathematics (pp. 27-44). Reston, VA: National Council of Teachers of Mathematics.

Ball, D. L., Thames, M. H., & Phelps, G. (2008). Content 513 knowledge for teaching what makes it special? *Journal of Teacher Education, 59*(5), 389-407.

Bieda, K. (2010). Enacting proof-related tasks in middle school mathematics: Challenges and opportunities. *Journal for Research in Mathematics Education, 41*(4), 351-382.

Chazan, D. (1993). High school geometry students' justification for their views of empirical evidence and mathematical proof. *Educational Studies in Mathematics, 24*(4), 359-387.

Cheng, Y. -H., & Lin, F.-L. (2007). The effectiveness and limitation of reading and coloring strategy in learning geometry proof. In J. H. Woo, H. C. Lew, K. S. Park, & D. Y. Seo (Eds.), *Proceedings of the 31st Conference of the International Group for the Psychology of Mathematics Education* (vol. 2, pp. 113-120). Seoul, Korea.

Cirillo, M. & Herbst, P. (2012). Moving toward more authentic proof practices in geometry. *The Mathematics Educator, 21*(2), 11-33.

Cohen, D. K., Raudenbush, S. W., & Ball, D. L. (2003). Resources, instruction, and research. *Educational Evaluation and Policy Analysis, 25*(2), 119-142.

Common Core State Standards Initiative. (2010). Common Core State Standards for Mathematics. Washington, DC: National Governors Association Center for Best Practices and the Council of Chief State School Officers.

de Villiers, M. D. (1990). The role and function of proof in mathematics. *Pythagoras, 24,* 17-24.

Doyle, W. (1986). Classroom organization and management. In M. Wittrock (Ed.), *Handbook of research on teaching* (*3rd edition*; pp. 392-431). New York: Macmillan.

Ellis, A. B. (2007). Connections between generalizing and justifying: Students' reasoning with linear relationships. *Journal for Research in Mathematics Education, 38*(3), 194–229.

Fujita, T. & Jones, K. (2014). Reasoning-and-proving in geometry in school mathematics textbooks in Japan. *International Journal of Education Research, 64,* 81-91.

Herbst, P. (2002). Establishing a custom of proving in American school geometry: Evolution of the two-column proof in the early twentieth century. *Educational Studies in Mathematics, 49*(3), 283-312.

Herbst, P., Aaron, W., Dimmel, J., and Erickson, A. (2013, April). Expanding students' involvement in proof problems: Are geometry teachers willing to depart from the norm? Paper presented at the 2013 meeting of the American Educational Research Association. http://hdl.handle.net/2027.42/97425

Herbst, P. & Chazan, D. (2012). On the instructional triangle and sources of justification for actions in mathematics teaching. *ZDM The International Journal of Mathematics Education, 44*(5), 601-612.

Hsu, H. Y. (2010). *The study of Taiwanese students' experiences with geometric calculation with number (GCN) and their performance on GCN and geometric proof (GP)* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3441210).

Knuth, E. (2002a). Secondary school mathematics teachers' conceptions of proof. *Journal for Research in Mathematics Education, 33*(5), 379-405.

Knuth, E. (2002b). Teachers' conceptions of proof in the context of secondary school mathematics. *Journal of Mathematics Teacher Education, 5,* 61-88.

Lakatos, I. (1976). *Proofs and refutations*. New York: NY: Cambridge University Press.

Lampert, M. (1992). Practices and problems in teaching authentic mathematics. In F. K. Oser, A. Dick, & J. Patry (Eds)., *Effective and responsible teaching* (pp. 295-314). San Francisco, CA: Jossey-Bass Publishers.

Martin, J. & White, P. (2005). *The language of evaluation: Appraisal in English*. New York, NY: Palgrave Macmillan.

Norton, A (2008). Josh's operational conjectures: Abductions of a splitting operation and the construction of new fractional schemes. *Journal for Research in Mathematics Education*, *39*(4), 401-430.

Otten, S., Gilbertson, N., Males, L. & Clark, D. (2014). The mathematical nature of reasoning-and-proving opportunities in geometry textbooks. *Mathematical Thinking and Learning*, *16*(1), 51-79.
Reid, D. A. (2002). Conjectures and refutations in grade 5 mathematics. *Journal for Research in Mathematics Education, 33,* 5-29.
Shaughnessy, J. M., & Burger, W. F. (1985). Spadework prior to deduction in geometry. *The Mathematics Teacher*, *78*(6), 419-428.
Steele, M. D. & Rogers, K. C. (2012). Relationships between mathematical knowledge for teaching and teaching practice: the case of proof. *Journal of Mathematics Teacher Education, 15*, 159-180.
Stylianides, G. J. (2009). Reasoning-and-proving in school mathematics textbooks. *Mathematical Thinking and Learning, 11*(4), 258-288.
Stylianides, G. & Stylianides, A. (2008). Proof in school mathematics: Insights from psychological research into students' ability for deductive reasoning. *Mathematical Thinking and Learning*, *10*(2), 103-133.