Mathematicians' grading of proofs with gaps

David Miller — West Virginia University — millerd@math.wvu.edu
Nicole Infante — West Virginia University — nengelke@math.wvu.edu
Keith Weber — Rutgers University — keith.weber@gse.rutgers.edu

In this study, we presented nine mathematics professors with three proofs containing gaps and asked the professors to assign the proofs a grade in the context of a transition-to-proof course. We found that the participants frequently deducted points from proofs that were correct and assigned grades based on their perceptions of how well students understood the proofs. The professors also indicated that they expected lecture proofs in the transition-to-proof course to have the same rigor as those demanded of students, but lecture proofs could be less rigorous than the rigor demanded of students in advanced mathematics courses. This presentation will focus on participants' rationales for these beliefs.

Key words: Mathematical Proofs, Assessment of Proofs, Transition to Proofs, Gaps in Proofs

In the United States, many mathematics majors are required to take a transition-to-proof course prior to taking proof-oriented courses such as real analysis and abstract algebra. A central goal of the transition-to-proof course is to help mathematics majors master the mechanics of proving so that they can produce acceptable proofs in their future advanced courses. There are, of course, a variety of strategies that a mathematics professor may use to achieve these goals, including being explicit about what types of inferences are valid (e.g., Alcock, 2010) and modeling good proving behavior (e.g., Fukawa-Connelly, 2012). There is currently a modest but growing body of research on how mathematics professors introduce students to proof-oriented mathematics in their lectures and their motivations for doing so (e.g., Alcock, 2010; Hemmi, 2006; Lai & Weber, 2014; Moore, 1994; Nardi, 2008; Weber, 2012). Recently, however, Moore (2014, submitted) identified an important facet of teaching that has received little attention from mathematics education researchers: professors' grading.

Moore (2014, submitted) found that mathematics professors viewed their grading, including both the marks they assigned and the commentary they provided, as essential parts of their teaching. As Moore observed, this raises important research questions. What student-written proofs in a transition-to-proof course constitute an acceptable product? What criteria do professors use for assigning grades? Do professors use the same criteria and assign similar grades to the same proofs? Or is there variance in the criteria that they use and the marks they assign? The purpose of this contributed paper is to further explore these questions. In particular, we presented nine mathematicians with proofs that contained gaps and asked them to grade these proofs. We were interested in how mathematicians would evaluate these gaps in their grading.

Related Literature

Mathematicians grading

As we noted in the introduction, there is little research on how mathematicians assign grades to students' proofs in a transition-to-proof course (or in any other course). Here we summarize the main findings from Moore's (2014, submitted) exploratory study on this topic. Moore asked four mathematics professors to assign grades to seven authentic student proofs with the aim of investigating the consistency, or lack thereof, in the marks that professors assigned. The main findings from Moore's study were that there was substantial variance in the scores they assigned to some proofs that could not be attributed to performance error (i.e., a participant overlooking a flaw in the proof). These disparities were based in part on the seriousness of the errors that stemmed from disagreement over whether an error was due to a mere oversight on the part of the student (which would receive only a small deduction in score or no deduction at all) or a significant misconception held by the student (which would receive a larger deduction). The professors all remarked that grading was an important part of their pedagogical practice.

Moore (submitted) called for expanding research in three directions: (i) conducting studies with more mathematicians to gain confidence in the generality of the findings, (ii) exploring how mathematics professors score different types of proofs, and (iii) looking into more depth about what criteria professors use to grade proofs. In this contributed report, we follow the recommendations of Moore. We presented nine mathematicians with a different type of proof grading tasks – looking at proofs with gaps – that allowed us to both replicate Moore's central findings and to explore them in more depth.

Proofs contain gaps in mathematicians' practice

A proof is sometimes defined as a deductive argument where each new statement in a proof is a permissible assumption (e.g., an axiom, a definition, a previously proven statement) or a necessary logical consequence of previous statements. In mathematical practice, it is commonplace for proofs to contain have *gaps* (Fallis, 2003). That is, the proof may not explicitly state exactly how a new assumption follows from previous assumptions, instead leaving this task up to the reader (Weber & Alcock, 2005). In many cases, this is because the author of the proof believed that the gap could easily be filled in by a knowledgeable reader. However, in other cases, the gap might be quite large and require the construction of a non-trivial sub-proof (Fallis, 2003; Selden & Selden, 2003). Mathematicians and philosophers have argued that gaps are both necessary and desirable in mathematical practice. Proofs would be impossibly long if every logical detail were included (Davis & Hersh, 1981) and supplying excessive logical detail would mask the main methods and ideas of the proof, which is a primary reason why mathematicians read published proofs in the first place (Rav, 1999; Thurston, 1994; Weber & Mejia-Ramos, 2011).

The pedagogical proofs that mathematics professors present to their students also contain gaps. Like published mathematical proofs, this is not only necessary for the sake of time and brevity, but also potentially beneficial, as students may learn mathematics from filling in some of the details of the proofs themselves (e.g., Alcock et al, 2015; Lai, Weber, & Mejia-Ramos, 2012). Prior research found a difference in how mathematics majors and mathematics professors regarded gaps in proofs presented in mathematics lectures: most mathematics majors believed that all the logical details should be specified in a well-written proof. In contrast, the majority of mathematics professors believed that even with a well-written proof, students would still be expected to justify some of the inferences within the proof (Weber & Mejia-Ramos, 2014). In this contributed report, we further explore the differences that mathematicians believe are needed with respect to rigor and gaps in lecture-based proofs and the proofs that students hand in for credit.

Mathematicians' evaluations of proofs with gaps

Prior research has explored how the existence of gaps affect the validity of proof in mathematicians' practice. In particular, mathematicians have not always agreed on the validity of specific proofs that contained gaps (e.g., Inglis, Mejia-Ramos, Alcock, & Weber, 2013; Weber, 2008; see also Inglis & Alcock, 2012). Mathematicians claimed that the permissibility of a gap was dependent upon the author of the proof, with some mathematicians claiming that they would be inclined to give an expert the benefit of the doubt when reading a proof with a large gap (Weber, 2008; Weber & Mejia-Ramos, 2013). In this paper, we illustrate how the mathematicians' estimation of the competence of the student plays a role in their grading when grading a students' proof with gaps.

Methods

Participants

Nine mathematicians in a mathematics department at a large state university in the United States agreed to participate in this study. These mathematicians represented a variety of mathematical subfields, including combinatorics, graph theory, number theory, partial differential equations, topology, and approximation theory. All were tenure-track or tenured at the time of the study with three professors representing each level of Assistant, Associate, and Full Professor. Six of the participants' had significant teaching and grading experience of over 10 years (three of which had over thirty years of experience) and the other three had four or more years of experience. We anonymized the data by referring to the first mathematician that we interviewed as M1, the second as M2, and so on.

Methods

In this study, we examined proofs of three theorems from number theory that might be proven in a transition-to-proof course. For each theorem, we generated two proofs. The *Gap Proof* is a proof that we designed to employ a logically correct line of reasoning but leaving some of the steps in the proof without a justification. The *Gapless Proof* is a modification of the Gap Proof such that the justifications for each of the steps were filled in. We refer to the three theorems as Theorem 1, Theorem 2, and Theorem 3. We refer to the two proofs of Theorem 1 as Gap Proof 1 and Gapless Proof 1, the two proofs of Theorem 2 as Gap Proof 2 and Gapless Proof 2, and the two proofs of Theorem 3 as Gap Proof 3 and Gapless Proof 3.

Procedure

Each participant met individually with the first two authors and was videotaped during a task-based interview. The interviewers made sure that each interviewee understood that the given proofs were from a transition-to-proof class. Each interview contained three phases. In the **Lecture Proof Evaluation** phase, the participant was told that a professor presented Gap Proof 1 in lecture, asked if they thought the proof was valid, and asked to comment on the pedagogical quality and appropriateness of the proof. This process was repeated for Gap Proof 2 and Gap Proof 3. In the **Student Proof Evaluation** phase, the participant was presented with Gap Proof 1 and told that a student submitted that proof for credit. They were asked to evaluate whether Gap Proof 1 was correct, assign a grade on a ten-point scale to that proof, and explain why they assigned that grade. They were then shown Gapless Proof 1 and asked to do the same thing. This process was repeated for the two proofs of Theorem 2 and Theorem 3. In the **Open-Ended Interview** phase, each participant was asked general questions about their pedagogical practice with respect to proof, with an emphasis on the grading of proof. One particular question was, "Do you expect the proofs that students hand in to have the same level of rigor as the proofs that the professors present in their lectures?"

Analysis

All interviews were transcribed. The research team engaged in thematic analysis as follows. First, each member of the research team individually read each transcript, flagging and commenting on passages that might be of theoretical interest. The research team met to discuss and compare their findings and identify themes that might be interesting to analyze in more detail. Next, each member of the research team individually read the transcripts again, searching for occasions in which a participant made a comment related to one of the themes in question and put this excerpt into a file related to that theme. The research team met again; for each theme, they used an open coding scheme in the style of Glaser and Strauss (1990) to create categories of participants' comments related to each theme.

Results

Summary of Evaluations

Although some participants were critical of the pedagogical quality of some of the Gap Proofs, they usually evaluated them to be correct. In the Lecture Proof Evaluation phase of the study, in all but one instance, the participants judged the Gap Proofs that they read to be correct. The one exception was when M8 could not decide if Gap Proof 1 was correct. In the Student Proof Evaluation phase of the study, none of the participants changed their evaluations of the correctness of the proof. Hence, there was only one instance (M8 evaluating Gap Proof 1) in which a participant evaluated the student proof as incorrect.

A summary of the grades that the professor assigned to the proofs that they evaluated in the Student Proof Evaluation phase is presented in Table 1. As can be seen from Table 1, there was substantial variance in the grades that the participants assigned from Gap Proof 1, with scores ranging from 6 through 10, thus replicating the findings of Moore (2014, submitted). There were 13 instances in which a Gap Proof received a score of less than 10; in 12 of those instances, the participant had judged the proof to be correct, with one score being as low as 6 out of 10. This illustrates how a correct proof is not guaranteed to receive full credit.

Table 1: Mathematicians' assessment of proofs with gaps authored by students

Proof	M1	M2	M3	M4	M5	M6	M7	M8	M9	Average
1 – Gap	6	9	8	9	9	7	10	6	8	8.00
1 – Gapless	9	10	10	10	10	9	10	9	9	9.56
2 – Gap	10	10	8	10	10	8	10	8	10	9.33
2 – Gapless	10	10	10	10	10	10	10	10	10	10
3 – Gap	10	9	10	10	10	10	10	10	8	9.67
3 – Gapless	10	10	10	10	10	10	10	10	10	10

Students' Proofs as a Model of Their Understanding

When discussing how they graded proofs, eight of the nine participants said that their grade was based on how well the student understood the proof that they handed in. For instance, M9 said[1],

M9: I think the way that I grade things is you're trying to see if the student understands and you believe he understands. Not so much that they have every period or word that you are looking for, but did they understand the concept … and if you could question them, then they could fill in the gaps, but they may have left them out.

[1] To increase the readability of the transcript, we lightly edited them by removing stutters, repeated words or phrases, and short fragments of text that did not carry meaning. We indicate where we have done so with an ellipsis (…). At no point did we add or alter words that participants said or change the meaning of the participants' utterances.

Similarly, in describing what he was looking for when he graded, M6 said, "I think it is… to see if they really understand what is going on. Careful proof and the steps, probably show that they understand the ideas and they can think – they can relate to one sentence to the other sentence and see the flow and how things put together".

The way that the participants modeled students' understanding influenced their grading in a number of ways. First, five participants indicated that if there was a gap in the proof but the inference could only be produced if the student had an adequate understanding of why it was true, they would not take off for it. For instance, in explaining why he didn't require a justification for the assertion that *n* is even in Gap Proof 1, M5 remarked:

M5: I guess I would say that if somebody was going to write a textbook and this was going to be a sample in the textbook, I would want them to say a little bit more. But anyone who says what is on the paper here, wouldn't say it without understanding it. So jumping from the fact that *n* squared is even to *n* is even, that does not bother me at all.

Similarly, five participants remarked that they would require more justification than was necessary for a proof to be correct to ensure that the student fully understood the proof and did not just copy or recall a proof that he or she saw elsewhere. For instance, consider the exchange between M4 and the interviewer when evaluating Gap Proof 1.

I: Do you think the proof is correct?

M4: Yeah.

I: If you had to, as we usually do as instructors, we grade things on some type of scale. Say you graded this on a zero to ten scale, what grade would you give this student?

M4: Well I would probably have to take off a little for not saying those [referring to an unjustified statement about why numbers are composite]. These two are obviously composite, composite numbers.

I: When you take off, a little off, what do you mean? How much is a little?

M4: Probably nine out of ten.

I: Would you make any comments on the students' papers?

M4: I would say why? Or explain why? And then I would think, did the student copy this somewhere? [The interviewer and M4 both laugh] Because it is sort of written in a mature style, leaving things out which are yes indeed. As I said before, compact proof, nicely done. Maybe too nicely.

Further, three participants indicated that they would take into consideration the past performance of the student when deciding whether to penalize a gap in the proof, with better students being more likely to get the benefit of the doubt. In discussing his strong students, M7 described:

M7: It is okay to leave some steps out because I might get to a point where I respect their mathematical minds enough so that I give them the benefit of the doubt that they understood what was going on without writing it down. So that is actually nice, a lot of people do that – in our own research we do that.

Level of Rigor in Lectures and Student Proofs

We asked the participants whether students' proofs should contain the same level of rigor as professor's proofs in lectures. Participants' responses frequently indicated that it depended on the course. For transition-to-proof courses, the answers were mixed, but on average was that the level of rigor in lecture proofs and student-generated proofs should be about the same. (Three participants remarked that they would expect *less* rigor for a student proof than a lecture proof based on their assumptions of what students would be capable at that point of their mathematical development). However, for upper-level courses like real analysis, six participants would expect students' proofs to contain *more* rigor. We summarize the participants' responses in Table 2.

Table 2: Rigor demanded in proofs that students submit for credit

Context	Less than lecture proof	Same as lecture proof	More than lecture proof
Transition-to-proof	M2, M3, M5	M1, M6, M7, M9	M4, M8
Advanced course	M5	M6, M9	M1, M2, M3, M4, M7, M8

M1 explained his justifications as follows. In the context of a transition-to-proof course, M1 exclaimed, "Yeah, at this level, especially if you are going to take off points, then I think the instructor owes it to the student to, you know, to write those things down carefully". For more advanced courses, M1 said:

M1: I think that is less important as the higher up you go and there is a difference in the venue of a lecture presented in class where there is just a, you know, for example for a graduate class, there well may be details or things that are clear from context that you would not want to spend time writing or take up class time for. You would like to focus on the mathematics itself.

Discussion

In this study, we replicated several of Moore's (2014, submitted) preliminary findings about proof grading. We found substantial variance in mathematicians' grading of Gap Proof 1 and we found that grading involved the professor's building models of how well they thought their students understood the proofs. Having these themes independently emerge in a study with a larger sample than Moore provides more confidence that his results would generalize to a larger number of mathematicians. Still, our sample size of mathematicians is rather small. As conducting qualitative analysis of interviews with a substantially larger number of mathematicians is impractical, we suggest further research might make use of the recent survey methodology that has been used to probe mathematicians' beliefs (e.g., Mejia-Ramos & Weber, 2014).

Our analysis builds on Moore's work by delving deeper into how and why participants use their models of students' understanding in assigning grades. We found that some participants would penalize students for gaps that would ordinarily be permissible, especially in proofs generated for a transition-to-proof courses, because they could not assume that students knew how to bridge these gaps. This helps explain why professors would assign scores of less than 10 to proofs that they judged to be correct. On the other hand, some participants would not penalize students for leaving a gap in the proof if they felt the student had a full understanding of the proof. Further, some gaps would be permissible for students who had previously earned the mathematical respect of the professor. Staples, Bartlo, and Thanheiser (2012) claimed that classroom proofs and mathematical proofs satisfy different needs and should be judged by different standards; in particular, in K-12 classrooms, the request for a proof is often used by a teacher as a lens to evaluate their understanding, something not ordinarily done in mathematicians' practice. The results of this study suggest that Staples, Bartlo, and Thanheiser's insight is relevant for transition-to-proof courses as well.

We also found that most participants believed an instructor of a transition-to-proof course should not expect students' proofs to be more rigorous than the proofs that he or she presents in lecture. However, for more advanced courses, most participants felt that the lecturer could be less rigorous in his or her lecture proofs than what he would demand of his or her students. This is significant for two reasons. First, this corroborates a finding reported in Lai and Weber (2014) that to mathematicians, proofs in advanced mathematics lectures have the purpose of communicating content while student-generated proofs are used for demonstrating students' capacities to write proofs. Second, previous research has shown that mathematics majors did not get this message (Weber, in press; Weber & Mejia-Ramos, 2014). Mathematics majors do not appear to read proofs as a tool to understand mathematical content or methods (c.f., Weber, in press) and they do not expect a good mathematical proof to have gaps (Weber & Mejia-Ramos, 2014). This has the context that they may ignore or misinterpret the most important ideas that a professor attempts to convey when delivering a lecture proof (Lew, Fukawa-Connelly, Mejia-Ramos, & Weber, in press). An implication of finding out how professors' standards of rigor in lectures and grading shifts from transition-to-proof courses to more advanced courses is that this information should be better conveyed to mathematics majors.

References

Alcock, L. (2010). Mathematicians' perspectives on the teaching and learning of proof. In F. Hitt, D. Holton, & P. Thompson (Eds.), *Research in Collegiate Mathematics Education VII* (pp. 63-92). Providence, RI: American Mathematical Society.

Alcock, L., Hodds, M., Roy, S., & Inglis, M. (2015). Investigating and improving undergraduate proof comprehension. *Notices of the American Mathematical Society, 62*, 742-752.

Davis, P., & Hersh, R. (1981). *The mathematical experience*. New York: Viking Penguin Inc.

Fallis, D. (2003). Intentional gaps in mathematical proofs. *Synthèse*, 134, 45–69.

Fukawa-Connelly, T. (2012). A case study of one instructor's lecture-based teaching of proof in abstract algebra: making sense of her pedagogical moves. *Educational Studies in Mathematics*, *81*(3), 325-345.

Hemmi, K. (2006). *Approaching proof in a community of mathematical practice*. Doctoral dissertation. Stockholm University.

Inglis, M., & Alcock, L. (2012). Expert and novice approaches to reading mathematical proofs. *Journal for Research in Mathematics Education*, *43*(4), 358-390.

Inglis, M., Mejia-Ramos, J. P., Weber, K., & Alcock, L. (2013). On mathematicians' different standards when evaluating elementary proofs. *Topics in Cognitive Science, 5*, 270-282.

Lai, Y. & Weber, K. (2014). Factors mathematicians profess to consider when presenting pedagogical proofs. *Educational Studies in Mathematics*, 85, 93-108.

Lai, Y., Weber, K., & Mejía-Ramos, J. P. (2012). Mathematicians' perspectives on features of a good pedagogical proof. *Cognition and Instruction*, *30*(2), 146-169.

Lew, K. ,Fukawa-Connelly, T., Mejia-Ramos, J.P., & Weber, K. (in press). Lectures in advanced mathematics: Why students might not understand what the professor is trying to convey. To appear in *Journal for Research in Mathematics Education*.

Mejia-Ramos, J. P., & Weber, K. (2014). Why and how mathematicians read proofs: further evidence from a survey study. *Educational Studies in Mathematics*, *85*(2), 161-173.

Moore, R. C. (1994). Making the transition to formal proof. *Educational Studies in mathematics*, *27*(3), 249-266.

Moore, R. C. (2014). What constitutes a well-written proof? In T. Fukawa-Connelly, G. Karakok, K. Keene, and M. Zandieh (Eds.), *Proceedings of the 17th Annual Conference on Research in Undergraduate Mathematics Education.* Denver, Colorado.

Moore, R.C. (submitted). Mathematics professors' evaluations of students' proofs: A complex teaching process. Submitted for publication.

Nardi, E. (2008). *Amongst mathematicians: Teaching and learning mathematics at university level*. USA: Springer.

Rav, Y. (1999). Why do we prove theorems? *Philosophia Mathematica, 7,* 5-41.

Selden, A., & Selden, J. (2003). Validations of proofs written as texts: Can undergraduates tell whether an argument proves a theorem? *Journal for Research in Mathematics Education, 36*(1), 4-36.

Staples, M., Bartlo, J., & Thanheiser, E. (2012). Justification as a teaching and learning practice: Its multifacted (potential) role in middle school classrooms. *Journal of Mathematical Behavior, 31,* 447-462.

Thurston, W.P. (1994). On proof and progress in mathematics. *Bulletin of the American*

Mathematical Society, 30(2), 161-177.
Weber, K. (2008). How mathematicians determine if an argument is a valid proof. *Journal for Research in Mathematics Education*, *39*, 431-459.
Weber, K. (2012). Mathematicians' perspectives on their pedagogical practice with respect to proof. *International Journal of Mathematics Education in Science and Technology*, 43, 463-482.
Weber, K. (in press). Effective proof reading strategies to foster comprehension of mathematical proofs. To appear in *International Journal for Research in Undergraduate Mathematics Education*.
Weber, K. & Alcock, L. (2005). Using warranted implications to understand and validate proofs. *For the Learning of Mathematics*, *25*(1), 34-38.
Weber, K. & Mejía-Ramos, J.P. (2011). Why and how mathematicians read proofs: An exploratory study. *Educational Studies in Mathematics, 76*(3), 329-344.
Weber, K., & Mejia-Ramos, J. P. (2013). On the influence of sources in the reading of mathematical text. *Journal of Literacy Research*, *45*, 87-96
Weber, K., & Mejia-Ramos, J.P. (2014). Mathematics majors' beliefs about proof reading. *International Journal of Mathematics Education in Science and Technology*, *45,* 89-103