Formative assessment of proof comprehension in undergraduate mathematics

(article begins on next page)

WEIZMANN INSTITUTE OF SCIENCE

Formative assessment of proof comprehension in undergraduate mathematics

Formative assessment of proof comprehension in undergraduate mathematics: Affordances of iterative lecturer feedback

Alon Pinto and Jason Cooper

Weizmann Institute of Science, Israel; alon.pinto@weizmann.ac.il; jasonc2107@gmail.com

Research shows that lecturer feedback on students' proofs is crucial for developing proof-comprehension in advanced mathematics courses, yet students often fail to comprehend lecturer feedback, and only rarely receive further feedback on their revisions. In this study we investigate the affordance of a novel formative-assessment scheme, designed and enacted by a mathematics professor, which employed multiple rounds of lecturer-feedback / student-revision. We analyze one such round, focusing on various facets of proof comprehension that underlie the lecturer's feedback and the student's subsequent revisions. On the basis of this analysis we discuss various ways in which lecturers can leverage feedback/revision cycles, not only for gaining insight into students' comprehension, but also for fostering meta-level ideas, and affording opportunities for students to develop agency and holistic proof comprehension.

Keywords: Undergraduate mathematics education, proof comprehension, formative assessment.

Introduction

One of the most prominent learning activities in advanced mathematical courses is proof reading. Students listen to their professors as they present proofs in lectures, and are expected to continue studying these proofs extensively after class, using their lecture notes or the course textbook (Weber, 2012). However, reading, validating and comprehending mathematical proofs are not easy tasks. It involves not only strategic knowledge in specific content areas, but also knowledge and norms specific to proof and reasoning (Knapp, 2005). In a study of students' proof validation practices, Alcock and Weber (2005) found that students focused on superficial features of proofs, while failing to develop a holistic view of the proof and neglecting arguments' explicit or implicit warrants. In spite of these difficulties, instructors rarely attend directly to proof reading in their teaching, presumably because knowledge related to proof reading strategies is largely tacit (Weber, 2012). Weber (2012) concludes that the mathematicians he interviewed provided "little guidance to students on how to engage in the complicated process of reading and comprehension of proofs and, by their own admission, lacked adequate methods for assessing students' understanding of a proof" (p. 478).

Lecturer feedback is an important aspect of assessment practice. Moore (2016) has argued that lecturer feedback is instrumental not only for proof comprehension, but also for developing the notion of proof and the ability to write proofs. Yet, relatively little empirical research has been conducted on how proofs are assessed in undergraduate mathematics courses, particularly in relation to lecturer feedback. Moore, Byrne, Hanusch, and Fukawa-Connelly (2016) have investigated students' comprehension of written lecturer feedback, and found that students "were generally quite capable of writing revised proofs that remediated the issues indicated by the professor's marks and comments, even when they could not fully explain the rationale for the comments" (p. 320). Consequently, students' written proofs were found to be insufficient for distinguishing between different levels of comprehension. These findings highlight that when students do not resubmit their revised proofs, neither the students nor the lecturers have a way of knowing whether students have interpreted the feedback and respond to it in ways that promote proof comprehension.

One way to address this limitation is by having students' work iteratively critiqued and resubmitted until deemed acceptable by the lecturer. Our research goal is to investigate how multiple rounds of lecturer feedback and student revision can be used effectively for formative assessment of proof comprehension. We use the term *formative* for assessment that is integrated with teaching to contribute, through feedback, to student learning. Mejia-Ramos, Fuller, Weber, Rhoads, and Samkoff (2012) have proposed a conceptual framework for proof comprehension based on an extensive literature review and on interviews with mathematicians. Based on this framework, they have proposed an assessment model that comprises questions for probing various facets of proof comprehension. They suggest that this model can be used by researchers "to examine how proof comprehension develops in students and to evaluate different means of improving it" (p. 4), and by lecturers to "inform what specific aspects of a given proof students understand and what aspects they do not understand" (p. 4). In this study we draw on Mejia-Ramos et al.'s work in order to scrutinize the affordances for formative assessment of proof comprehension of a novel assessment scheme that includes cycles of feedback/revision, which was designed and enacted by a mathematics professor.

Setting

The assessment scheme examined in this paper was used in a proof-oriented Real-Analysis course at a large public research university in the United States. The key aspect of the assessment scheme was the instructor's decision to replace the "traditional" problem sets (weekly homework assignments; exams) with a term-paper assignment: Students were required to produce and submit weekly 'lecture notes' that would present selected proofs taught in the lecture in the students' own words. The instructor, whom we will call Mike, is a research mathematician who had been teaching proof-oriented courses for more than two decades. In the term under investigation he invested a great deal of time and effort in this novel assessment scheme, planning his lectures accordingly, redesigning the assignments for the course, and providing extensive written feedback on student submissions. Assessment of the students' proofs proceeded in cycles of feedback and revision until the instructor was satisfied. These submissions were a passing requirement for the course.

Prior research of Mike's goals and expectations for the term-paper assignment revealed that it was meant to promote students' proof comprehension by scaffolding their independent proof reading (Pinto & Karsenty, 2018). Thus, the term paper assignment can be seen as part of a formative assessment scheme for proof comprehension. Though this assessment scheme was not designed by educational researchers, it nevertheless seemed to address some of the key limitations for formative assessment of traditional assessment practices (e.g., Moore et al., 2016). Many of the students' initial submissions contained numerous flaws. Mike's assessment scheme gave him more flexibility when facing feedback related challenges, for example:

- Prioritizing comments: Indicating all of the deficiencies in a proof can be overwhelming for students, while selective feedback may be misconstrued as endorsement of unmarked errors.
- Diagnosing student (mis)comprehension: Deficiencies in a proof can be linked to different kinds of miscomprehension. Has the instructor correctly diagnosed them?

- Providing effective feedback: What will students learn or mis-learn from the instructor's comments? Will they be able to leverage the feedback to produce a satisfactory proof?

In this paper we describe how Mike addressed these challenges in one feedback/revision cycle and present an in-depth analysis of its affordances for formative assessment, thus shedding light on ways lecturer and students can leverage feedback/revision cycles to promote proof comprehension.

Methodology

The main source of data for our analysis is iterations of students' proofs, along with Mike's comments on each iteration. Additional background data include: video-documentation of lectures; field notes; informal discussions; and two 90-minute interviews with Mike, one conducted at the beginning of the course and the other after its conclusion. The interviews focused on the course design and aimed at eliciting Mike's goals and expectations (for further detail, see Pinto & Karsenty, 2017).

For the study reported herein we chose to analyze a particularly long cycle: six submissions of a student, whom we will call Ben, for a proof of the theorem, *a non-empty subset of the real numbers bounded above has a least upper bound.* The 6th submission was accepted without comment. Mike defined the real numbers as the set of all decimals and proved the theorem in class by constructing the least upper bound, digit after digit in an infinite iterative process. In his first submission, Ben tried a different approach for proving the theorem. This submission had numerous deficiencies, including a structural flaw, as Ben relied on a corollary of the theorem he was attempting to prove. The number of iterations and the variety of flaws in Ben's first submission, which intensify the dilemmas discussed in the previous section, made this cycle particularly suitable for our investigation.

We view students' submission of a proof "in their own words" as an opportunity for assessing their comprehension of the proof presented during the lecture. Such assessment is formative when the instructor, in his feedback, invites students to reflect on and develop their comprehension of the proof. We characterize opportunities for formative assessment through analysis of the instructor's feedback (i.e., which aspects of proof-comprehension he stresses) and of the students' subsequent revisions (i.e., how they attend to the feedback). For this analysis we draw on the seven facets of proof comprehension proposed by Mejia-Ramos et al. (2012, see Figure 1), which were grouped into *local* – discerned by studying a few related statements within the proof – and *holistic* – ascertained by inferring the ideas or methods that motivate a major part of the proof or the proof in its entirety.

These seven facets were operationalized in terms of 19 assessment items (pp. 5–6). In coding Mike's feedback, we asked ourselves: "to which of the 19 assessment items is Mike implicitly asking students to attend?" For example, the feedback "one can only prove a statement, not a bound" (Figure 1) was coded as *meaning of terms and statements*, because Mike appears to want Ben to notice that the statement should have been "to prove *the existence of* an upper bound," and therefore his feedback could have been rephrased as "explain the meaning of the term *prove a bound.*" In coding Ben's submissions, we listed all the deficiencies that we found in his proof, and for each we asked ourselves: "Which of the 19 assessment items can we conclude that Ben has responded to inadequately?" For example, if a statement was correct but did not serve a logical purpose in the proof sequence, we coded for *logical status*, because Ben would apparently not have had an adequate response to the implicit question "identify the purpose of the sentence."

Local Facets (*example*)	Holistic Facets (*example*)
• Meaning of terms and statements (*state the definition of a term*) • Logical status of statements and proof framework (*identify the purpose of a sentence*) • Justification of claims (*identify the specific data supporting a given claim*)	• (Summary of) high level ideas (*provide a good summary of the proof*) • Transferring general ideas or methods to another context (*appreciate the scope of the method*) • Modular structure (*identify the logical relation between modules of a proof*) • Illustrate with examples (*illustrate a sequence of inferences with a specific example*)

Figure 1: Seven facets of local and holistic proof comprehension (Mejia-Ramos et al., 2012)

The coding was conducted in three stages. In the first stage, each author individually coded local facets of each of Ben's submissions, as a backdrop against which to consider the instructor's feedback. We refrained from coding holistic facets, which would have been highly speculative. In the second stage, each author individually coded Mike's feedback on Ben's submissions. Some feedback addressing normative mathematical writing (e.g., inadequate notation) did not align with the framework, and was excluded from the analysis. Disagreements in coding were discussed and resolved. In the third stage, we located Ben's responses to Mike's comments, and coded them according to the deficiencies that they did or did not address. Our analysis of the coding had three foci: 1. Ben's initial submission, reflecting the range of issues from which Mike could choose what to attend to; 2. Subsequent interplay of feedback and revision; 3. Ben's final submission and overall appraisal of the assessment process.

Analysis

Theorem 4.1: A non-empty set $E \subset \mathbb{R}$ bounded above has the least upper bound (denoted $\sup E$).

Why? It might be clear to you, but not to me.

"To prove an upper bound? One can only prove a statement, not a bound.

Proof 4.1: Let A be the set of real numbers as follows: $\alpha \in A$ if and only if there is some $x \in E$ such that $\alpha < x$. We denote B as the set of all real numbers not in A. It is clear that no member of A is an upper bound of E, and it is also clear that every member of B is an upper bound of E. Thus, to prove a least upper bound of E, it must be proved that B contains a smallest number. It must be verified that A and B satisfy the following hypotheses:

(a) Every real number is either in A or in B
(b) No real number is in A and in B
(c) Neither A nor B is empty:
(d) If $\alpha \in A$ and $\beta \in B$, then $\alpha < \beta$:

I've already voiced my objections to this argument last time. If you don't understand my comments, you should ask. The reason why it cannot be correct is because you never use the definition of real numbers, and so the argument applies verbatum for rational numbers. But in this case the statement is false: in Q, the set E of rational numbers x with x^2<2 do not have the least upper bound.

Given these hypotheses are true, then there is one and only one real number γ such that $\alpha \leq \gamma$ for all $\alpha \in A$ and $\gamma \leq \beta$ for all $\beta \in B$.

(a) and (b) hold by contradiction. Suppose there are two numbers: γ_1 and γ_2, and $\gamma_1 < \gamma_2$. Choose γ_3 such that $\gamma_1 < \gamma_3 < \gamma_2$. Then, $\gamma_3 < \gamma_2$ means γ_3 must lie in A, but $\gamma_1 < \gamma_3$ implied γ_3 must lie in B. Thus, there is only one such γ that lies in either A or B. (c) must hold because since E is not empty, there is some $\alpha < x$, so A is not empty. Since E is bounded above by B, there is a y such that $x \leq y$ for $x \in E$, thus there is a $y \in B$ so B is not empty. (d) must hold since it has been shown that $\alpha < x$. If there is some $\beta \in B$, $x \leq \beta$. Then, $\alpha < \beta$.

Counterexample in Q:
Let A and B be rationals respectively smaller and greater than \sqrt{2}.

There exists a corollary such that if the above 4 hypotheses are satisfied, then either A contains some largest number or B contains some smallest number. But, A cannot contain some largest number: let us choose an α' such that $\alpha < \alpha' < x$. Then, $\alpha' \in x$ and α cannot be the largest number of A.

Figure 2: Ben's 2nd submission; revisions of the 1st submission are highlighted; Mike's feedback appears in red

1. Ben's first submission: In Ben's first submission (see Figure 2, only the highlighted text was added in the second submission) we recognized substantial flaws: Ben relied on a corollary of the theorem he was attempting to prove; this is a case of cyclic reasoning, which we coded as *logical status of statements and proof framework*, because we assume he might have responded inadequately to the question "identify the purpose of the sentence within the proof framework." Furthermore, the proof of this corollary that was presented in the lecture relied on properties of the construction of $\mathbb{R}$,

a crucial point that Ben did not mention in the proof. We coded this oversight as *justification of claims*, because Ben would presumably have failed to answer the question "make explicit an implicit warrant in the proof." We coded Ben's use of the phrase "it is clear that" as another case of *justification of claims*, because in our judgment what followed required elaboration, suggesting that Ben had not appreciated how subtle the justification of this claim might be. The phrase "it must be proved that" was coded as *logical status of statements*, because in fact it is not *necessary* to prove the claim that followed, but rather it is *sufficient*.

2. <u>Interplay of feedback/revisions</u>: Mike's feedback on the 1st submission referred only to the holistic structure of the proof, without explicitly drawing attention to the issue of cyclic reasoning: "Your argument cannot be correct because you are not using the definition of real numbers and therefore your argument equally applies to rational numbers [where] the statement […] is false." We note that the course definition of the real numbers had a central role in the lecture proof of the existence of least upper bounds in the domain of real numbers. Thus, Mike's feedback underlined a key idea that was absent in Ben's 1st submission, and was therefore coded *Holistic*: *High-level ideas*. In response to this feedback, in his 2nd submission, Ben made two changes (Figure 2, highlighted text): he added the full statement of the corollary that he was relying on (*Local: Meaning of statements*) and he added justifications as to why the corollary's conditions are satisfied (*Local: Justification of statements*). The remainder of the proof was unchanged. Therefore, there is a discrepancy between Mike's holistic feedback and Ben's local revisions. Seeing that the structural flaw was not resolved, Mike reiterated his criticism in his feedback on Ben's 2nd submission: "I've already voiced my objection to this argument last time…" (*Holistic*: *high-level ideas*) and added an example of how the corollary does not hold for rational numbers (*Holistic: illustrate with examples*). Mike also commented on local issues that were already present in Ben's 1st submission, criticizing Ben's use of the phrase "it is clear that…" (*Local: Justification of claims)* and drawing attention to his misuse of the word "prove" (*Local: Meaning of terms).* There were several additional indicators of possible comprehension issues that we identified in Ben's 2nd submission that Mike did not comment on. For example, it is *sufficient* and not *necessary* to verify the conditions of the corollary (*Local: Logical status of statements*).

<u>Synopsis of submissions 3-6</u>: In his 3rd submission, Ben reverted to the proof that had been presented in a lecture, though it is not clear to what extent he appreciated the structural problem in his initial approach. In his 3rd feedback, Mike implicitly endorsed Ben's new approach, and his feedback addressed issues of local comprehension as well as providing structural hints (e.g., "Now prove that b^* is an upper bound, and is smaller than any other upper bound," *Local: proof framework*). The 4th feedback included a mix of local and holistic comments; the 5th and final feedback was strictly local. In both the 4th and 5th feedbacks, Mike focused on notational issues. Our analysis–comparing successive submissions–highlighted improvements in Ben's proofs, yet Mike did not comment on this, limiting his feedback to errors and deficiencies. Our analysis also highlighted that in many cases, Ben's revisions, while attending to Mike's feedback, introduced new local issues, some of which were commented on and subsequently revised.

3. <u>Ben's final submission</u>: While most of Mike's comments were resolved by the 6th submission, our analysis reveals some unresolved issues. Nevertheless, it was implicitly endorsed when accepted without comment.

In spite of several apparent *local* issues in Ben's 1st submission, Mike chose to focus his first feedback on big ideas, attending to them in a *holistic* manner, stressing that as long as there is no use of properties of real numbers the proof cannot be correct, because it would apply to rational numbers where the claim does not hold. He illustrated this through a counter-example: The set of rational numbers less than $\sqrt{2}$ does not have a rational least upper bound. Ben's revisions generally attended to Mike's comments, yet he did not always attend to them immediately. For example, in his 4th feedback, Mike criticized local aspects of Ben's definition of a sequence of nested sets: "here there needs to be a condition on a." Ben's 5th submission did not address this problem, so Mike commented again in his 5th feedback, and Ben eventually addressed it in his 6th and final submission.

Discussion

Our analysis highlights potential affordances of multiple cycles of feedback and revision for formative assessment of proof comprehension, which we now turn to discuss while keeping in mind the three feedback-related challenges discussed in the Setting section: prioritizing comments, diagnosing student (mis)comprehension, and providing effective feedback.

Our analysis of Ben's 1st submission highlighted various deficiencies. In a traditional assessment model, providing only one opportunity to provide feedback, Mike might have responded to Ben's 1st submission by commenting on all deficiencies, which could have been overwhelming for Ben. Alternatively, Mike could have picked one or two issues that he considers most crucial. He would need to be quite explicit in his feedback, because he would not have the opportunity to clarify subtle points that Ben could not make sense of on his own. Instead, in the cycle we have analyzed, Mike's feedback began with holistic structural issues, later moving to local issues of notation, terms, and logical status and justification of claims. Mike did not need to be exhaustive in his feedback, knowing he would have opportunities to return to unattended issues later on. This transition suggests lecturers can leverage multiple cycles of feedback for prioritizing feedback according to a didactic agenda. Postponing feedback related to local issues until resolving issues related to structure and big ideas of a proof does not only support students' development of proof comprehension, but also signals which aspects of comprehension are most valued. For example, in his feedback to Ben's first two submissions, Mike highlighted a meta-level idea related to proof and proving–does the proof make use of all the assumed conditions, and how would it fail without them? This is a central theme in university mathematics, where the domain of a result is of interest, and reflects a "mathematical habits of mind" of *tinkering* (Cuoco, Goldenberg, & Mark, 1997).

In the cycle we examined, there were deficiencies that Mike highlighted in his feedback that were not resolved in Ben's next submission, and deficiencies that emerged during the revision process. While an analysis of Ben's proof comprehension is beyond the scope of this paper, we note that these additional data provide valuable insights into Ben's comprehension that were not salient in the 1st submission, which suggests that multiple rounds of feedback and revision could support lecturers' diagnosis of their students' proof comprehension. In addition, we found discrepancies between the facets of proof comprehension that Mike and Ben addressed in their respective feedback and revision. In particular, we found that in some cases Ben responded to holistic-oriented feedback with local-oriented revisions, and responded holistically only after Mike reiterated his comment. Thus, our analysis indicates that multiple rounds of feedback and revision could support lecturers in providing feedback that students could leverage to produce a proof that the lecturers would find acceptable.

This case study also illustrates that it may be necessary to support students in harnessing lecturer feedback to develop their holistic proof comprehension. Research indicates that it is difficult for students to develop holistic comprehension of proofs even when instructors' explicitly highlight holistic aspects of proofs in their lectures (Lew, Fukawa-Connelly, Mejía-Ramos, & Weber, 2016). Cycles of feedback and revision may offer affordances for holistic comprehension. Mike's first feedback on Ben's alternate proof scheme was not prescriptive—he did not tell Ben how to correct his proof. Providing open questions as feedback is a risky move in a traditional single-round assessment, as students may fail to answer these questions and revise the proof, as Ben's second submission illustrates. However, because Mike was able to monitor Ben's revisions and provide further feedback, he was able to provide Ben the opportunity not only to reflect on the inadequacies of his own submissions, but also on the ways in which they are addressed in the lecture-proof that he eventually reproduced (e.g., the role of the definition of real numbers in the proof). The discrepancies between Mike's holistic feedback and Ben's local revisions, along with Mike's comments on these discrepancies, highlighted gaps in instructor-student communication that were likely to remain hidden in a traditional single-round assessment of Ben's first submission.

We now turn to discuss a potential affordance of multiple cycles of feedback and revision from the student perspective. In his first submission, Ben attempted a proof scheme different from the one presented in class. In his re-submissions, Ben postponed attending to some feedback, splitting his responses to Mike's 4th feedback over his 5th and 6th submissions. While we can only speculate on Ben's considerations, we recognize in his readiness to submit an original proof and to prioritize his revisions an indication of agency that in our experience is not common in undergraduate mathematics courses, particularly in relation to proof writing. A consequence of instructors' tendency to grade proofs by reducing points for deficiencies rather than assigning points for merit is that students might avoid taking risks and constructing original arguments that expose their thinking and (mis)comprehensions. In contrast, if students know they will have the opportunity to resubmit their work, and that their work will be graded on the quality of their final submission, then they have little to lose and something to gain from constructing an original argument. Thus, students can leverage multiple cycles of feedback and revisions as an invitation to explore and take risks in their proving.

Concluding thoughts

We have discussed various potential affordances of an assessment scheme that include multiple rounds of feedback/revision for formative assessment of students' comprehension of mathematical proof. We note an advantage of such a scheme in relation to the scheme proposed by Mejia-Ramos et al. (2012); whereas Mejia-Ramos et al. propose designing sets of questions to probe various facets of student comprehension, multiple rounds of feedback and revision seem more compatible with traditional practices of assessment grounded in proof validation, and thus may be more accessible and appealing for some instructors. We acknowledge that multiple rounds of feedback and revision are demanding for both instructors and students, because each new homework assignment entails reviewing/revising any number of prior assignments. Furthermore, it is not clear when the process should end. Ben's 6th and final submission included several substantial deficiencies. Did Mike deem it "pedagogically acceptable"? Did he decide that in spite of its flaws, his time and effort would be better spent commenting on Ben's submissions of more recent assignments? Or did he simply tire of the process? Cutting the feedback process short could have serious consequences, because refraining from comment on a student's proof could be misconstrued as implicit endorsement of its correctness. It is also important to recognize that although multiple iterations of feedback and revision may provide invaluable opportunities to engage students with the various facets of proof comprehension, lecturers and students will not necessarily capitalize on these opportunities. Mike noted, in retrospect, that in some cases students ended up copying proofs from the textbook. In the final interview, he mentioned several examples where students' revisions of their proofs, in light of his comments, revealed that "they did not understand what they were talking about."

Thus, there is still much we need to learn about enacting assessment schemes based on multiple rounds of feedback and revision effectively. We call for further research that will develop and validate various formative assessment schemes, towards promoting student proof comprehension.

References

Alcock, L., & Weber, K. (2005). Proof validation in real analysis: Inferring and evaluating warrants. *Journal of Mathematical Behavior, 24*(2), 125–134.

Cuoco, A., Goldenberg, P. E., & Mark, J. (1996). Habits of mind: An organizing principle for mathematics curriculum. The *Journal of Mathematical Behavior, 15*(4), 375–402.

Knapp, J. (2005). Learning to prove in order to prove to learn. https://mathpost.asu.edu/~sjgm/issues/2005_spring/SJGM_knapp.pdf.

Lew, K., Fukawa-Connelly, T. P., Mejía-Ramos, J. P., & Weber, K. (2016). Lectures in advanced mathematics: Why students might not understand what the mathematics professor is trying to convey. *Journal for Research in Mathematics Education*, *47*(2), 162–98.

Mejia-Ramos, J. P., Fuller, E., Weber, K., Rhoads, K., & Samkoff, A. (2012). An assessment model for proof comprehension in undergraduate mathematics. *Educational Studies in Mathematics, 79*(1), 3–18.

Moore, R. C. (2016). Mathematics professors' evaluations of students' proofs: A complex teaching process. *International Journal of Research in Undergraduate Mathematics Education, 2*(2), 246–278.

Moore, R. C., Byrne, M., Hanusch, S., & Fukawa-Connelly, T. (2016). When we grade students' proofs, do they understand our feedback? *https://digitalcommons.andrews.edu/pubs/422*.

Pinto, A. & Karsenty, R. (2018). From course design to presentations of proofs: How mathematics professors attend to student independent proof reading. *Journal of Mathematics Behavior, 49*, 129–144.

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.