2016

When we Grade Students' Proofs, Do They Understand our Feedback?

Robert C. Moore
Andrews University, moorer@andrews.edu

Martha Byrne
Earlham College, byrnema@earlham.edu

Sarah Hanusch
Texas State University - San Marcos, sh1609@txstate.edu

Timothy Fukawa-Connelly
Temple University, timatunh@gmail.com

When we grade students' proofs, do they understand our feedback?

Robert C. Moore
Andrews University

Martha Byrne
Earlham College

Sarah Hanusch
Texas State University

Tim Fukawa-Connelly
Temple University

Instructors often write feedback on students' proofs even if there is no expectation for the students to revise and resubmit the work. It is not known, however, what students do with that feedback or if they understand the professor's intentions. To this end, we asked eight advanced mathematics undergraduates to respond to professor comments on four written proofs by interpreting and implementing the comments. We analyzed the student's responses using the categories of corrective feedback for language acquisition, viewing the language of mathematical proof as a register of academic English.

Keywords: Proof Writing, Proof Grading, Proof Instruction, Proof Revision, Student Thinking

Introduction and Research Questions

Rav (1999) claimed that proofs "are the heart of mathematics" and they play an "intricate role […] in generating mathematical knowledge and understanding" (p. 6), and multiple authors have claimed that proof is how the discipline grows (cf., Lakatos, 1976; de Villiers, 2003). Consequently, proof is perhaps the dominant feature of the advanced undergraduate mathematics curriculum. For example, Mills (2011) estimated that approximately half of all class time consists of instructors presenting proofs at the board while the presentation of definitions, examples, algorithms and theorems (as well as explanations of the same) together constitute less than half of class time. Similarly, student homework and exams in advanced mathematics classes generally include a number of prompts asking them to produce written proofs of given claims (cf., Fukawa-Connelly, 2015). That is, as Weber stated, developing proof proficiency "is often the primary goal of advanced mathematics courses and typically the only means of assessing students' performance" (2001, p. 101).

While proof proficiency may be the primary goal of advanced undergraduate mathematics courses, a significant body of research has demonstrated that students have great difficulty writing proofs, and a body of writing about mathematicians' personal teaching experiences supports this claim as well. For example, Epp (2003) claimed:

> Often their efforts consisted of little more than a few disconnected calculations and imprecisely or incorrectly used words and phrases that did not even advance the substance of their cases. My students seemed to live in a different logical and linguistic world from the one I inhabited, a world that made it very difficult for them to engage in the kind of abstract mathematical thinking I was trying to help them learn. (886)

Due to space constraints, we do not report all of this research here, rather we note that there are multiple handbook chapters (e.g., Harel and Sowder, 2007) describing the literature, and a comprehensive summary of undergraduate mathematics majors' difficulties with proof is given in Selden and Selden (2008). This difficulty is further illustrated by the emphasis on reforming proof instruction as evidenced by the large number of "how-to" textbooks on proofs (e.g., Chartrand, Polimeni, & Zhang, 2012; Franklin & Daoud, 2011; Smith, Eggen & St. Andre, 2014), and practitioner articles about improving student proof writing (e.g., Strickland & Rand, in press; Zerr & Zerr, 2011). Of particular importance to the present study is Moore's (1994) claim that students are unfamiliar with the language of mathematical proof. Yet, the language of proof writing is largely unstudied, and thus, there is little information about how mathematicians understand this language; instead, we rely on individual reflections and theoretical analyses.

We argue that the language of proof is a particular academic register of English. A register is a variety of language used for a particular purpose or within a particular social context. Specifically, it "is the set of meanings, the configuration of semantic patterns, that are typically drawn upon under the specific conditions, along with the words and structures that are used in the realization of these meanings" (Halliday, 1978, p. 23).To support our claim that the language of mathematical proof constitutes a register, we note that Scarcella (2003) claimed that academic English was a particular register "used in professional books and characterized by the specific linguistic features associated with academic disciplines" (p. 9) and that each academic discipline had its own particular subregister. Halliday (1978), Pimm (1987) and Moschkovich (1999) have similarly argued that the language of mathematics constitutes a particular register. Thus, "we can refer to a 'mathematics register,' in the sense of the meanings that belong to the language of mathematics […], and that a language must express if it is being used for mathematical purpose" (Halliday, 1978, p.195).

When considering how students might learn and use this register, exposure and practice is certainly important (Pinker, 2009). Yet, without feedback on this practice, students are unlikely to improve their proof writing. Moreover, mathematicians act on the belief that giving students feedback is critical to their learning by making marks and notes on student proof productions (e.g., Hattie & Timperley, 2007; Moore, 2016). Yet, this feedback improves student learning only if students read, make sense of, and incorporate it into their future work. Few, if any, studies have explored students' understanding of this process of giving feedback as a means to improve their proficiency. Thus, in this study we investigate the following questions:

1. What do students claim to do with the professor's feedback on their proofs?
2. How do students interpret and explain the rationale for the professor's marks and comments on student-written proofs?
3. How do students' responses to the professor's comments align with the way that is normative in the discipline, as described by mathematically enculturated individuals?

Literature and Theory

Theoretical orientation

The theoretical orientation for this study is acquisition of language through which we consider the register of mathematical proof. In the case of proof in an advanced undergraduate mathematics classroom, there is an interesting duality: there are the proofs that the professors present and those that the students produce. Professors report that they often present proofs with small gaps in them; giving multiple reasons for doing so, such as not allowing technical details to obscure the big ideas, conserving limited class time, and providing the gaps as learning opportunities for students to help them better understand the content and support their proof writing (Lai & Weber, 2014; Lai, Weber, & Mejia-Ramos, 2012).

Language acquisition

While there is little research on how students learn formal mathematical language, there is a significant body of research on language acquisition (both first and second language acquisition) which has relevance to the acquisition of a register. Pinker (2009) described four criteria for a language, or in this case, a register, to be learnable:

1. From among the class of all possible languages, the target needs to be identified.
2. Learners need an environment in which to learn it.
3. Learners need a learning strategy, meaning a learner-created algorithm that uses information in the environment to create "hypotheses" about the target language, and then to determine whether they are consistent with the input information from the environment.
4. Learners determine a success criterion, meaning the hypotheses are related in some systematic way to the target language. A learner may waver among a set of hypotheses, one of which is correct. They may arrive at a hypothesis identical to the target language, or they may arrive at an approximation to it.

All of these are fulfilled by the advanced mathematics classroom: there is a particular register to be acquired, an environment in which to learn it, means for trying out usages of the language (proof writing), and a success criterion (sufficient fluency to earn passing marks).

Pinker continued by noting that language learning was a case of induction—developing uncertain generalizations from the observed instances of the language used in the environment. In the case of the formal mathematical register, students often get specific instruction describing the use of some phrases as well as the syntax, and they observe their professors using the register, although mixed with colloquial mathematics (Lew, et a., 2016). Moreover, professors frequently write only the algebraic steps of a proof on the board while saying aloud many of the connecting phrases and logical underpinnings (Fukawa-Connelly, 2012). In contrast, when reading proofs the professors focus on the phrases and logical descriptions in the written language, acting as if the algebraic manipulations are of secondary importance (Hodds, Alcock, & Inglis, 2014). Thus, the sample of the register on which the students are inducting is problematic and suggests an over-importance of the symbolic argument, thus increasing the number of generalizations likely to be present in the superset of the language that the students might develop. It is from these uses that students must develop hypotheses about the symbols, phrases and syntax, and yet as noted above, unlike in colloquial language, formal mathematical language must be unambiguous and correct, thus increasing the challenge for students.

Most important, language acquisition research suggests vital roles for learners' language production, including that learners "try out new language forms and structures as they stretch their interlanguage to meet communicative needs; they may use output to see what works and what does not" (Swain, 1998, p. 68). That is, students learning the mathematical register will write proofs in which they will commonly use correct grammar and syntax but also incorrect grammar and syntax, and they do this, partially, in order to try to learn the language and solicit feedback. This testing of language is not typically a conscious goal, rather we appear to be biologically disposed to do so. While to a mathematical expert, novice proof productions are filled with errors of logic, grammar, syntax and more, under the lens of language learning we instead view them as attempts to communicate in the style of the community where the rules are, at best, partially mastered and often tacit. On the basis of this research, we assert that professors' feedback on student proof productions has perhaps an unparalleled role in students' learning to produce proofs.

The role of feedback in language acquisition

We adopt Leeman's (2007) definition of feedback, meaning a mechanism that provides a learner with information regarding the success or failure of a given production, such as a written proof. Much research has focused on negative feedback, which is feedback that in some way indicates an error or issue with the production (Gass, 2008; Herschensohn & Young-Scholten, 2013). In particular, Lyster and Ranta (1997) described six types of corrective feedback, each of which can also be seen as indicating issues with the learner's production. Their types of feedback included:

1. Explicit correction: the teacher indicates that the utterance is incorrect and provides the correct form.
2. Recast: the teacher reformulates the student's production and remediates the error or provides the correction, such as writing out the correct sentence, without specifically indicating the error in the student's work.
3. Clarification request: the teacher asks the student to reformulate their production without specifically indicating what the error is.
4. Metalinguistic clues: the teacher poses questions, typically with a yes/no answer, or gives comments related to the student's production but does not provide the correct form or specifically indicate the error (e.g., Do we use this symbol here?").
5. Elicitation: the teacher directly elicits the correct form from the student by asking questions or starting the phrase. These questions typically expect more than a yes/no answer, such as, "Where do we use this hypothesis?"
6. Repetition: the teacher repeats the student production but highlights the error in the delivery (e.g., highlighting).

Lyster and Ranta found that recasts and explicit corrections did not result in subsequent improvement in student productions and hypothesized that it was because students are then limited to repeating the correct form the teacher provided (Tedick & de Gortari, 1998). Gass (2008) argued that explicit negative feedback can only teach about surface-level phenomena and cannot teach abstraction. Consequently, students may use the correct grammar and syntax in the future but without developing reasons to support their choices. The other five types of responses do not provide learners with the correct form and require that the learners provide it. Lyster and Ranta claim that when learners must restate, they are having to actively engage with the feedback, which is a critical feature in learning from feedback.

Student and mathematician misunderstandings

While we argue that professor comments on student proof productions have a significant role in students' learning to produce proofs, we also have reason to believe that students are likely to misinterpret them. Research on student understanding of lectures suggests that students sometimes develop significantly different understandings of the presented material and meaning for professor actions than the professor intends (cf. Lew, et al., 2016; Weinberg, Weisner, & Fukawa-Connelly, 2014). The work of Ko and Knuth (2013) and Selden and Selden (2003) has shown that students often fixate on the form, such as the presence of mathematical symbols, rather than the content of the proof, when reading mathematical arguments.

We use this prior research on misunderstandings and "misses of understanding" to form hypotheses about how students are likely to interpret a professor's comments on proof productions. (To learn about the kinds of comments professors leave on proofs, see Moore, 2016). In particular, we hypothesize that they are likely to:

- not apprehend some feedback,
- develop only a surface-level understanding of some feedback, and
- interpret feedback in ways that differ from what mathematical experts would do.

Moreover, we argue that the latter two of these hypotheses are supported by the language acquisition literature described above.

Methods

Participant selection

The participants were 8 students, 4 men and 4 women, with advanced undergraduate standing at two institutions, four from each institution. Each participant had taken at least two proof-based undergraduate mathematics classes, including a transition-to-proof course. We purposely selected participants who had experience with writing proofs and receiving feedback from their professors so as to give the best possible chances for their success in understanding the proofs and interpreting the professor's comments in this study.

Data collection

We engaged each participant in a 90-minute task-based interview where the primary task was to describe and interpret a professor's comments about proofs. The interviews were audio-recorded and pencast with Livescribe pens. Each interview began with basic demographic questions and reflective questions about the participant's typical use of their professors' feedback, as follows:

1. What is your major(s) and which college-level math courses have you taken? Which math courses have emphasized proof writing?
2. When a professor returned homework papers, how often were you asked to revise and resubmit your work?
3. When you were not asked to revise and resubmit, what did you do with the feedback your professors gave you? Can you provide a specific example?
4. If you were asked to revise and resubmit, what did you do with the feedback?
5. Do you think it is important to read the comments on your proofs? Why or why not?

Subsequently, the main part of the interview consisted of a sequence of three or four proofs, as time allowed. We ensured that each proof had a mix of professor marks and comments related to notation and presentation, and logical issues. The proofs and professor comments were taken from a previous research project exploring four mathematics professors' proof grading practices (Moore, 2016). From among the various marks and comments that the professors in Moore's study wrote on the proofs, we selected the ones that appeared on the proofs in the present study. An example proof, Proof A, with comments, is shown in Figure 1. (Note that the participants did not see the numbers beside the comments; we inserted them later for our convenience in referring to the comments).

First, we handed a proof to the participant, told her it had been written by a student, and asked her to read and understand it as best she could. Next, we presented a marked proof to the participant with a professor's feedback written in red ink. To determine whether the participant's interpretation of the mark or comment matched our own, we asked the participant to explain why the professor had written each individual mark or comment and what changes she thought the professor wanted. Finally, we asked the participant to rewrite the proof in order to allow us to further explore her interpretation of the comments and see how she implemented the professor's recommendations.

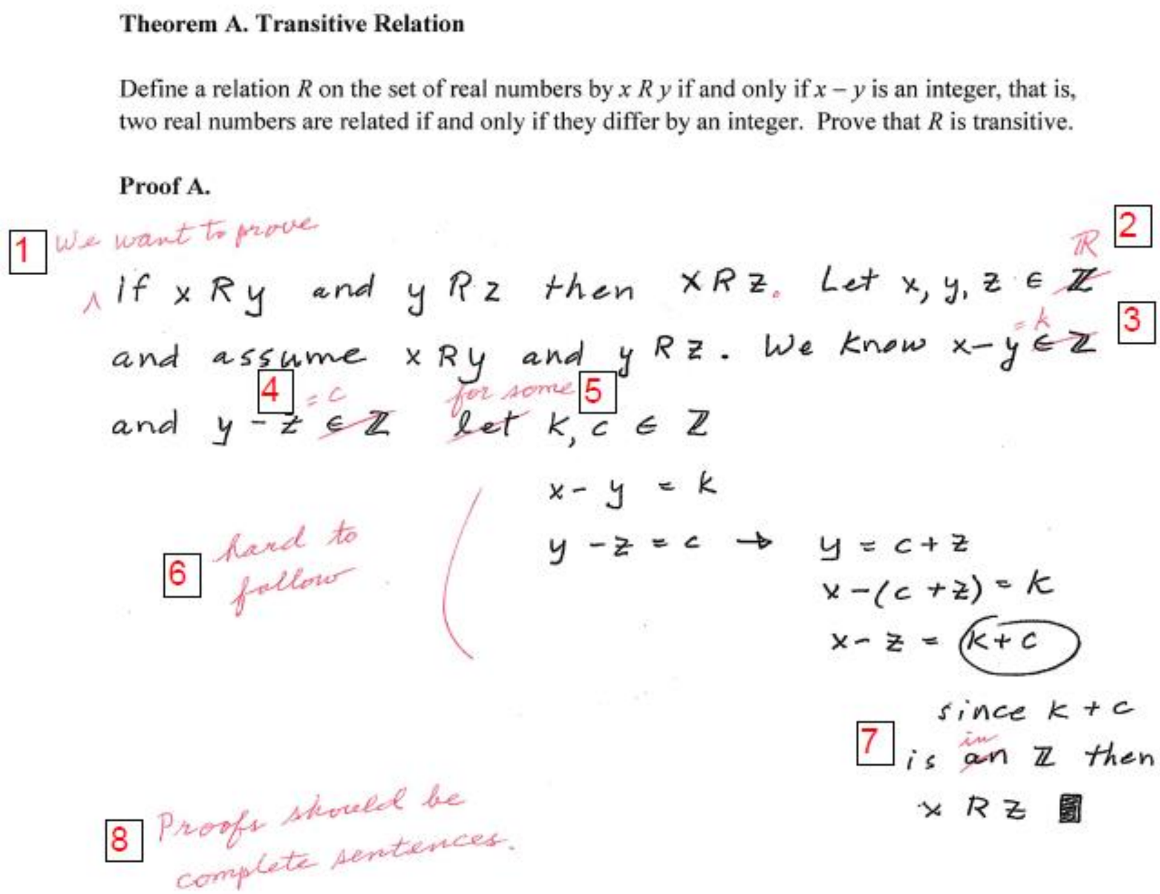

Figure 1. Proof A with the professor's comments.

Data analysis

For each comment in each proof, each of the researchers wrote a description of what change he or she believed the professor wanted and a rationale for the change. Based on these individual notes, we created a consensus description of what each mark and comment was asking the participant to change and the reason for the change. We also classified each comment using Lyster and Ranta's (1997) types of corrective feedback. Thus, each comment was classified along three dimensions: what change was requested, the reason for the change, and the type of corrective feedback.

We transcribed each interview and then chunked it at a number of levels. We parsed the demographic and reflective questions in one piece and the participant's discussion of each proof in additional pieces. We partitioned the discussion of each proof by identifying the participant's initial reading of the proof as a whole, and then her conversations about the individual comments. In cases where participants discussed multiple comments in the same utterance, we looked across interviews, and when it was common, we treated the comments as a single unit to parse all interviews similarly. We made a final block of the talk-aloud proof-writing process, for which we chunked the participant's utterances around the comments and linked those to what she wrote when she revised the proof.

To code the participant's utterances about each comment we first wrote a brief holistic summary, and then we developed a more detailed coding sheet that recorded:

- what the participant identified as the part of the proof the comment addressed,
- what the participant's response suggested should change in the proof,
- any reason the participant gave to explain the intended change and the underlying logic,
- a summary of what the participant changed in her revised proof,
- a comparison of each of the above points to our consensus expert interpretation, and
- an explanation of how an unanticipated change exhibited during the proof writing could be understood as a logical interpretation of the professor's comments.

We then created summaries, first by looking at individual proof comments across participants, and then by looking across the different categorizations of comments. For example, in the case of explicit corrections we described how the participants interpreted the requested changes, the kinds of reasons the participants provided for them, and the changes the participants included in their revised proofs. Similarly, we aggregated all of the changes that, according to our expert consensus, were recommended by the professor for a particular reason, such as "cultural convention," described the kinds of reasons the participants provided for them, and noted the changes the participants included in their revised proofs.

Results

Overall, the participants were very successful at interpreting what a professor wanted them to do in response to the comments. For example, for each of the eight comments on Proof A, 100% of participants correctly identified an acceptable part of the proof to be changed, and they all executed a change in a manner logically consistent with our understanding of the comment. However, their explanations of the rationale for the comments were not always consistent with our expert understanding. In the sections that follow, we explore the participants' work and thinking about the professor's comments on Proofs A and B.

Explicit corrections and recasts: feedback that specified the change

When the professor's comment specified a change to make to the proof, such as in an explicit correction or recast, the participants were consistently able to identify and state the change the professor wanted, and in the case of explicit corrections, what the professor considered incorrect. In their revised proofs, all of the participants consistently adopted the professor's suggested changes. For example, six of the professor's eight comments on Proof A (see Figure 1) were explicit corrections or recasts. Five of them indicated that something in the proof should be crossed out and replaced. In these instances, the participants always identified what they believed the professor wanted them to revise and implemented the revisions in a way that conformed with expert understanding.

Comment 1, the only recast comment on Proof A, was also specific but suggested the addition of new text, namely, the phrase "We want to prove," rather than the replacement of existing text. Seven participants added the recommended phrase, and one participant, Adam, showed some individuality by assuming that *xRy* and *yRz* and then writing, "We will prove that *xRz*," which we judged to be consistent with our understanding of the professor's comment.

In summary, when the professor's comment was an explicit change of or an addition to the proof-text, the participants' identification and implementations of the recommended changes were largely consistent with the expert consensus, or made a related change that addressed the intent of the comment. Thus, it appears that the participants demonstrated that they could appropriately identify and use explicit corrections and recasts on proofs.

Elicitations and clarification requests: feedback that did not specify the change

The professor's comments on the four proofs included eleven elicitations and clarification requests. None of these comments gave specific directions about how to revise the proof. For example, comment 6 on Proof A, which says "hard to follow," was a clarification request that did not indicate how to revise the proof. The participants generally interpreted this comment to mean that the algebraic steps lacked readability. Adam said about this comment:

> When I was reading the original proof, I guess it took me a little bit of time to follow how they laid it out, and maybe the layout could have been a little bit easier to follow. So maybe we should change some of this so it's a little bit easier to get from certain points to certain points.

Our interpretation of Adam's statement is that the comment asked for a change in layout in order to help a reader understand how the proof flows from point to point.

Bella and Don interpreted comment 6 differently and focused on the idea that the professor was asking for more detail. Bella focused on adding algebraic steps, whereas Don suggested clarifying and justifying the steps: "I would take that to mean use more detail, I suppose, describing what I'm actually doing at each step, even it were just to label, you know, 1, 2, 3, and then preferably say why 1, why 2, why 3."

Two participants, Ruby and Nancy, noted specific changes that they would make to the proof. Ruby suggested that, "If there were sentences there, it wouldn't be as hard to follow." Nancy explained how she would add transition phrases and justifications, when she said:

> Normally you say more instead of just writing a bunch of random equations. ... You could say like "using substitution, we know this part" ... instead of just writing down random arrows and saying "Ha, ha! We got it!" … You can figure out what it's trying to say, it's not hard to follow in that sense, but it's hard to follow in the sense of formal proof writing. It doesn't have the normal, yeah, words, English.

Nancy has claimed that text is a "normal" part of proofs and a "bunch of random equations" is not appropriate, and her interpretation of the comment appears to be influenced by this belief.

On Proof B, shown in Figure 2, we classified comments 2 and 3 as clarification requests. For comment 2, the professor circled a part of the student's work and wrote, "right idea, bad notation." For comment 3, the professor crossed out two parts of the student's proof and wrote, "Bad syntax. Sets can't imply sets." Note that in comment 2, the professor did not specify the error in what was written, and for neither of these two comments did the professor indicate how to revise the proof, which is consistent with the definition of a clarification request.

When the participants read comment 2, "right idea, bad notation," all but one were able to identify that the professor was indicating that the student's arrow notation was incorrect and needed to be changed, but most of them had difficulty articulating exactly what was incorrect and how to correct it. For example, Charles said, “I suppose this tells me they did it wrong. I would have to look up to see what the proper notation was, if I didn’t know what it was off hand.”

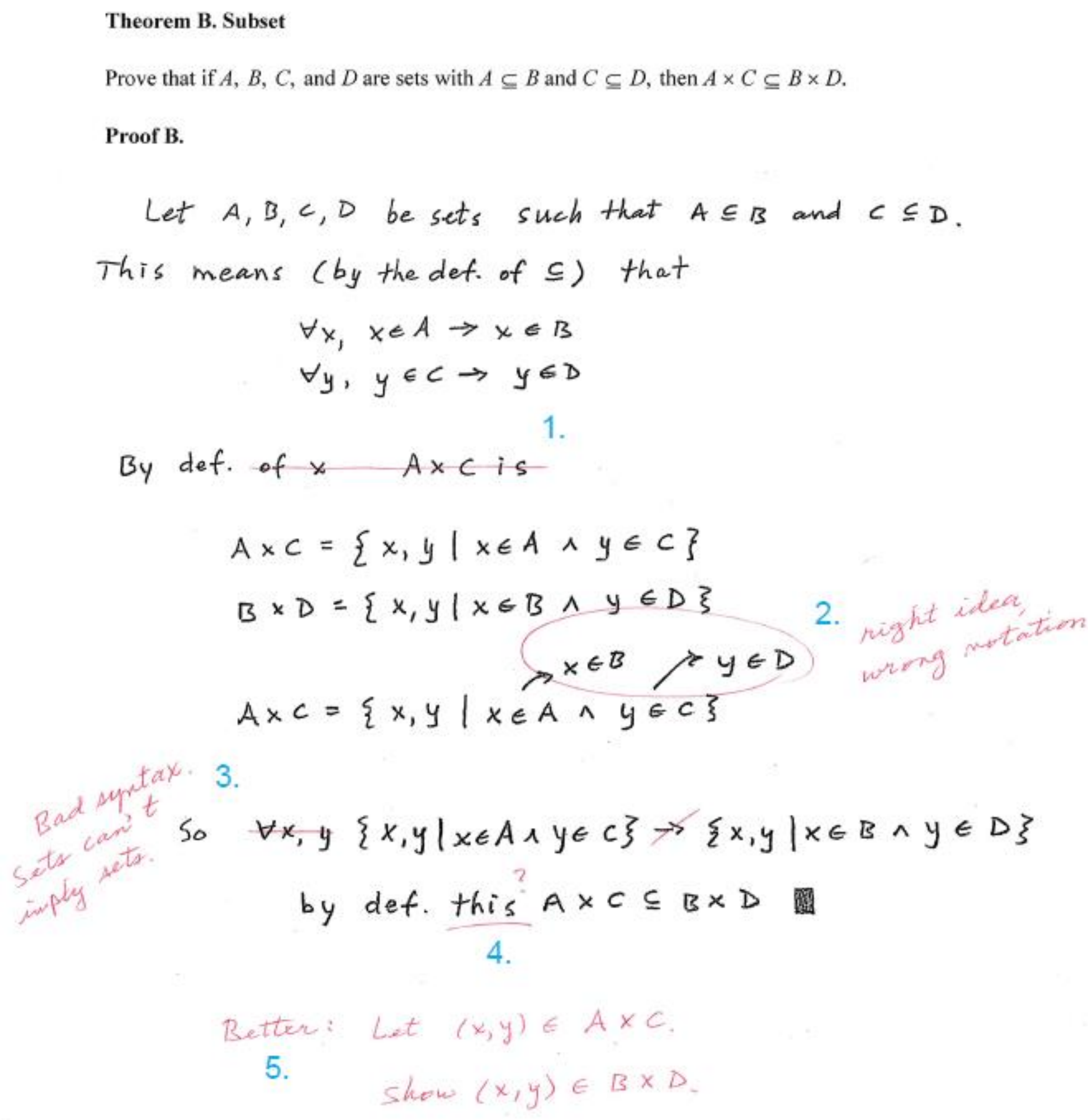

Figure 2. Proof B with the professor’s comments.

Two participants suggested a means to deal with the comment by changing everything into written language without using symbols. Genevieve claimed, “that should be sentences instead of arrows.” Nancy noted that the text is “not written in a mathematically formalized way. It’s more written in a shorthand of reminding yourself” and suggested “actually writing it out in a way that says ... it in a mathematical proof kind of way.” Nancy’s response provided insufficient evidence to determine what she thought would make the argument more formal, but when asked, suggested writing out the inclusions in words: “Here I might write, ‘by definition of subset,’ and then write, ‘x is in B and y is in D,’ or something like that.”

The participants’ responses to comment 3 were similar to those for comment 2. They recognized the arrow between the two sets as problematic but did not give much more information. Bella stated, “It’s not something you can do, it’s not like, it’s just maybe a rule.” Multiple participants commented that the arrow indicated implication, such as Adam, “Yeah, the arrow means implies. Syntax. … you have to make sure you’re using the right ones to mean what you want it to mean. And implies doesn’t work for sets. You just don’t do that.” Genevieve said, “It’s not grammatical when you substitute implies in place of arrow.” Thus, although the participants generally recognized the problematic usage of the arrow, their verbal explanations of their interpretations of this comment offered little insight into what was wrong about the arrow and what they would change. Moreover, all of the participants avoided the problematic arrow usage in their revised proofs, typically by writing the inclusion in sentence form, often significantly restructuring the last few lines of text.

Participants’ descriptions of the logic of specified changes

The participants were also asked to describe why they thought the professor had specified each of the changes to the proof text, and their answers revealed a variety of ways of thinking about the language of formal mathematics. For most of the comments, some participants were able to express the logic supporting the comment and some were not, and this was consistent across comment types.

With regard to the “We want to prove” recast comment at the beginning of Proof A, the expert consensus was that a proof should not begin with the conclusion that is to be proved, and the correction indicates that the first sentence expresses the goal of the proof. Two participants clearly articulated the logical issues with first line of the proof. Here is Adam’s response:

> I think what the professor meant by this is you want to make a statement saying this is what we are going to prove or trying to prove.... They didn’t say we’re trying to prove this or this is the conclusion we’re going to come to. And you want to make that clear.

In contrast, Bella explained that she understood the reason for the comment as “that’s just one of the proper ways to start a proof, that from what I’ve learned, yeah, it’s just the way to start a proof.” That is, her thinking appeared to focus on the form of proofs, rather than on the logical function of the statement. The other five participants described the added phrase as clarifying the presentation, but they did not clearly articulate the logical issue. For example, Don said, “This is just to me good syntax. It’s a way of setting it up to be understand better and to be read more easily.” Thus, we suggest that the logic the professor intended to motivate by this comment was not successfully communicated to six of the participants.

The participants also initially showed mixed understanding of the reasons for the professor’s explicit corrections on Proof A (Figure 1). For comment 2, which specified changing $\boldsymbol{Z}$ to $\boldsymbol{R}$, the experts agreed that the change was logically necessary because the relation R is defined on the set of real number. Six of the participants gave an explanation that approached that of the experts, including Genevieve who said, “there is no reason to believe that x, y, z are in the integers. The theorem never states that they are in the integers. [The theorem states] on the set of real numbers.” Don gave a mixed explanation, initially saying “they [$\boldsymbol{Z}$ and $\boldsymbol{R}$] are both correct but the real numbers are more applicable, in most cases,” but later in the interview noted that the theorem specifies that *x*, *y*, *z* are real numbers. Don's explanation for the requested change did not reject the original statement as inappropriate for the proof.

In reference to comment 5, only two participants, Don and Nancy, articulated the distinction between *let* and *for some,* whereas the other participants said little about this comment or gave explanations that did not fully align with the expert consensus. For instance, Charles said: "... when you make the mark *for some k, c,* that's traditional writing. But it also helps clarify how *k* and *c* are related to the two previous corrections. Um, whereas the student seems to be setting it up as its own separate idea, it's important to clarify that *k* and *c* are related to *x, y*, and *z*." Charles's comment about "traditional writing" suggests that he viewed the explicit correction as stylistic, i.e., to bring the writing into alignment with writing norms, although he also noted the logical issue.

The participants' interpretations of the clarification requests were similarly met with mixed success. Consider comment 2 on Proof B (Figure 2). The participants generally did not know the correct notation, and as a result, gave only an explanation about the importance of correct notation. For example, Adam said, "Wrong notation. Let me think here. (pause) I don't necessarily know if I know the right notation to put this in now (laughs), but I knew that wasn't the right notation." Genevieve, on the other hand, was able to articulate the notational issues with this part of the proof by saying "The arrows are sort of sometimes ambiguous with people because it generally means implies, but sometimes people use arrows in different ways, and so I think to actually put it in a sentence and make it clearer would be a lot better."

Similarly, for comment 3 on Proof B the participants agreed that "sets can't imply sets" but struggled to articulate the reason. Nancy and Ruby attempted to give a rationale in terms of *if-then* statements but had difficulty expressing themselves clearly and succinctly.

The participants' revisions of the proofs

When the professor gave an explicit correction or recast comment, the participants always incorporated it, or revised the proof in such a way as to avoid the problem. When the professor used an elicitation or a clarification request, the participants had to develop their own change. One such comment was "proofs should be complete sentences," which is somewhat directive in that it suggested a course of action. Four of the participants rewrote the entire proof in paragraph form, eschewing the string of equations, while the other four displayed the string of equations and wrote the concluding part of the proof in complete sentences. Both are reasonable interpretations of the professor's note, and either is stylistically acceptable.

In sum, the participants 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.

What the participants claimed to do with professor feedback

To begin the interview, we asked the participants to describe what they typically did with the professors' comments on their graded proofs. Generally, they claimed to make relatively little use of them. Bella noted that she did not consistently read the comments, noting that "sometimes it would just go on the pile of homework, to be honest." Five more claimed to consistently read them. Nancy claimed to do more than simply read the comments:

> I would read through it, um, and I would try to kind of make mental notes. I would often not try to go back and rewrite it, but if I then did another proof that was similar, I would ... try to

> make a mental note of it and use that. I did try to use it in other proofs. I wouldn't always redo those, but I definitely would use that information the next time that I wrote a proof. And I would actually get it out and look at it the next time I was writing a proof and try to see what did I do, what could I do differently to make this better.

Adam, too, talked about making additional use of the comments, and in his case it was in the context of his abstract algebra course. During this course, the teacher would sometimes return proof papers in class and go over them, and in that case Adam would write notes on his proofs for later reference. Adam claimed that he would use the written and oral comments to help him prepare to rewrite the same proofs during exams. But it seemed that without that motivation, he would normally do no more than read the comments.

Six of the participants claimed to make use of their professors' comments when they were asked to revise and resubmit proofs, but they all noted that they were seldom asked for revisions. This evidence suggests that students generally do not meaningfully engage with the comments on their graded proofs unless required to do so. The language acquisition literature (e.g., Lyster & Ranta, 1997) and composition research (e.g., Bean, 2011) argue that such engagement is critical for learning from feedback.

Conclusion

This study contributes to undergraduate mathematics education in that it is the first study to describe and analyze how students interpret and respond to their graded proof papers. Given that mathematicians consider proof grading to be an important means of teaching students to write proofs (Moore, 2016), the study opens a new line of research on the teaching and learning of mathematical proof at the undergraduate level.

We report three principal findings in response to the research questions. The first important finding is that when the professor wrote an explicit correction or a recast comment on a proof, the participants correctly identified the changes recommended and generally, but not always, could provide some rationale for the changes, including what was incorrect about the original proof. We suggest that the participants were able to invent reasons by comparing the original text and the corrected text; that is, they were able to develop a hypothesis about usage via induction (Swain, 1998). In contrast, when the professor's comment was a clarification request or an elicitation and did not provide new text, the participants struggled to provide a rationale for the change and fell back to more general claims that did not explicitly identify what was incorrect about the original proof. We note that the literature on second language acquisition suggests that cognitive engagement with the feedback is important for the students to incorporate the new information, and that clarification requests and elicitations promote cognitive engagement better than explicit corrections and recasts (Lyster & Ranta, 1997). Yet in this study, when the professor wrote clarification requests and elicitations, our participants were not reliably successful in responding to the comments because they did not appear to fully understand the comment and the mathematical language to correct the problem. Instead, they relied on general claims that are insufficient to support future proof-writing attempts. On the other hand, in the context of an actual course, if the professor requires students to revise their work, clarification requests and elicitations may promote cognitive engagement by directing students to use available resources to learn how to address the issues raised by the comments on their proofs.

The second important finding is that, regardless of the type of feedback the professor offered, the participants could not reliably describe normatively correct logic for the changes and why the original proof production was incorrect or problematic. While explicit corrections and recasts gave the participants a means to develop hypotheses by contrasting the two proofs, their inferences were sometimes incorrect or incomplete. It appeared that, generally, the participants could only explain the logic for requested changes when they already could identify the issue prior to reading the professor's comments. Moreover, it appears that they would often fall back to claiming that a requested change was "cultural" or "how you do it in formal mathematics." This observation suggests that the way professors currently write comments is not an effective way to communicate the reasons underlying the comments, which they often claim is the most important thing they are attempting to convey in their instruction (Lew, et al., 2016). Our data suggests that professors should write more about the logic that they are attempting to convey to the students, as well as distinguish between logical errors and readability concerns. What form such feedback might take to be most effective and efficient, whether direct statements, clarification requests, or elicitations, is a productive direction for future research.

A third finding is that students' written proofs are insufficient to distinguish between those who have some level of conceptual understanding and those who work only procedurally. When they revised the proofs, the participants successfully implemented the suggested changes in nearly all instances, regardless of the type of comment, even when they did not fully understand the rationale for the changes. This ability to write a correct proof without an understanding of the underlying logic is problematic for the teacher who may conclude that students understand more than they actually do. This potential mismatch suggests that perhaps professors have good reason to focus on teaching students the important mathematical logic and let the language, symbols, and grammar sort themselves out over time.

We recognize that this is a single, exploratory study with a small number of elementary proofs, a small number of participants, and only analytical generalizations. Moreover, we note three significant limitations of this study that suggest the need for further work. First, the participants were reading and writing proofs on mathematical topics that most of them had not worked with in some time, possibly since their introduction to proof class. Second, we asked the participants to interpret comments on proofs they had not written, thus imposing a need to make sense of another student's proof attempt prior to interpreting the comments. More research is needed to explore students' ability to interpret feedback in the context of their own proof writing. This first exploratory study provides a body of empirical evidence for future directions and more theoretical work. We note a third limitation: initially the four experts did not always agree on the reasons for the changes. While we could come to a consensus interpretation, there were significant differences in our initial interpretations, which means that different researchers, or a different mix of researchers, might have arrived at a different consensus interpretation of the professor's comments. This limitation of the study suggests an avenue for future research, motivated by Weber's (2014) argument that proof is a cluster concept. We hypothesize that while professors might share instructional goals about proof and use similar notes and language to communicate with students, in reality they may be attempting to convey very different content via the same notes, which has significant implications for students.

References

Bean, J. C. (2011). *Engaging ideas: The professor's guide to integrating writing, critical thinking, and active learning in the classroom* (2nd ed.). Hoboken, NJ: John Wiley & Sons.

Chartrand, G., Polimeni, A. D., & Zhang, P. (2012). *Mathematical proofs: A transition to advanced mathematics* (3rd Ed.). Boston, MA: Pearson.

De Villiers, M. D. (2003). *Rethinking proof with the Geometer's Sketchpad*. Oakland, CA: Key Curriculum Press.

Epp, S. S. (2003). The role of logic in teaching proof. *American Mathematical Monthly*, *110,* 886-899.

Franklin, J., & Daoud, A. (2011). *Proof in mathematics: An introduction.* Sydney, Australia: Kew Books.

Fukawa-Connelly, T. P. (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.

Fukawa-Connelly, T. (2015). Responsibility for proving and defining in abstract algebra class. *International Journal of Mathematical Education in Science and Technology*, 1-17. 10.1080/0020739X.2015.1114159

Gass, S. (2008). Input and interaction. In Doughty, C. J., & Long, M. H. (Eds.). (2008). *The handbook of second language acquisition* (Vol. 27). Malden, MA: John Wiley & Sons.

Halliday, M. K. (1978). *Language as a social semiotic: The social interpretation of language and meaning*. Baltimore: University Park Press.

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* (p. 805-842). Charlotte, NC: Information Age.

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of educational research*, *77*(1), 81-112.

Herschensohn, J., and Young-Scholten, M. (eds.) (2013). *The Cambridge handbook for second language acquisition*. Cambridge: Cambridge University Press.

Hodds, M., Alcock, L., & Inglis, M. (2014). Self-explanation training improves proof comprehension. *Journal for Research in Mathematics Education*, *45*(1), 62-101.

Ko, Y., & Knuth, E. J. (2013). Validating proofs and counterexamples across content domains: Practices of importance for mathematics majors. *Journal of Mathematical Behavior,* 32, 20-35.

Lai, Y., & Weber, K. (2014). Factors mathematicians profess to consider when presenting pedagogical proofs. *Educational Studies in Mathematics*,*85*(1), 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.

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

Leeman, J. (2007). Feedback in L2 learning: Responding to errors during practice. *Practice in a second language: Perspectives from applied linguistics and cognitive psychology*, 111-137.

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), 162-198.

Lyster, R., & Ranta, L. (1997). Corrective feedback and learner uptake. *Studies in second language acquisition*, *19*(01), 37-66.

Mills, M. (2011) Mathematicians' pedagogical thoughts and practices in proof presentation. In S. Brown, S. Larsen, K. Marrongelle, & M. Oehrtman (Eds.), *Proceedings of the 14th Annual*

Conference on Research in Undergraduate Mathematics Education (pp. 283-297), Portland, Oregon. Retrieved from http://sigmaa.maa.org/rume/Site/ Proceedings.html

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

Moore, R. C. (2016). Mathematics professors' evaluation of students' proofs: A complex teaching practice. *International Journal of Research in Undergraduate Mathematics Education, 2*(2), 246-278. doi:10.1007/s40753-016-0029-y

Moschkovich, J. (1999) Supporting the participation of English language learners in mathematical discussions. *For the Learning of Mathematics*, 19(1), 11–19.

Pimm, D. (1987) *Speaking mathematically: Communication in mathematics classrooms*. London: Routledge.

Pinker, S. (2009). *Language learnability and language development, with new commentary by the author* (Vol. 7). Harvard University Press.

Rav, Y. (1999). Why do we prove theorems? *Philosophia mathematica-Series 3*, *7*(1), 5-41.

Scarcella, R. (2003). *Academic English: A conceptual framework*. UC Berkeley: University of California Linguistic Minority Research Institute.

Selden, A., & Selden, J. (2003). Validations of proofs considered as texts: Can undergraduates tell whether an argument proves a theorem? *Journal for Research in Mathematics Education,* 34, 4-36. Washington, DC: Mathematical Association of America.

Selden. A. & Selden, J. (2008) Overcoming students' difficulties in learning to understand and construct proofs. In M. P. Carlson & C. Rasmussen (Eds.) *Making the connection: Research and teaching in undergraduate mathematics education* (p. 95-110).

Smith, D., Eggen, M., & St. Andre, R. (2014). *A transition to advanced mathematics* (8th ed.) Boston, MA: Cengage.

Strickland, S., Rand, B. (in press). Learning proofs via composition instruction techniques. In *Beyond Lecture: Techniques to Improve Student Proof-Writing Across the Curriculum.* Washington, DC: Mathematical Association of America.

Swain, M. (1998). Focus on form through conscious reflection. *Focus on form in classroom second language acquisition*, 64-81.

Tedick, D. J., & de Gortari, B. (1998). Research on error correction and implications for classroom teaching. *ACIE Newsletter*, *1*(3), 1-6.

Weinberg, A., Wiesner, E., & Fukawa-Connelly, T. (2014). Students' sense-making frames in mathematics lectures. *The Journal of Mathematical Behavior*, *33*, 168-179.

Weber, K. (2001). Student difficulty in constructing proofs: The need for strategic knowledge. *Educational Studies in Mathematics, 48,* 101-119.

Weber, K. (2014). What is proof? A linguistic answer to an educational question. In T. Fukawa-Connelly, G. Karakok, K. Keene, & M. Zandieh (Eds.), *Proceedings of the 17th Annual Conference on Research in Undergraduate Mathematics Education* (pp. 1-13), Denver, Colorado.

Zerr, J. M., & Zerr, R. J. (2011). Learning from their mistakes: Using students' incorrect proofs as a pedagogical tool. *PRIMUS, 21*, 530-544.