Categorizing Professors' Feedback on Student Proofs in Abstract Algebra and Real Analysis

Ben Spiro
Temple University

Sarah Hanusch
SUNY - Oswego

David Miller
West Virginia University

Robert C. Moore
Andrews University

Tim Fukawa-Connelly
Temple University

Abstract: Mathematics faculty spend considerable time scoring and providing feedback on student-generated proofs. Yet there is very little research on the feedback professors provide on proofs during the grading process. In this paper, we discuss a coding scheme developed for categorizing the feedback professors write on student-generated proofs in abstract algebra and real analysis. We then explore the types of annotations that professors make on student proof attempts. The results show that professors generously use annotations (like checkmarks) as informal grading tools or to signify things they have read when grading, most feedback focuses on a particular part of the proof that is no more than a few lines, and the majority of feedback does not convey why the feedback was given.

Keywords: Proof Evaluation, Feedback On Proofs, Proof Instruction, Abstract Algebra, Real Analysis.

Introduction and Literature Review

Proof writing is difficult for undergraduate students (cf., Stylianides, Stylianides, & Weber, 2018) and one of the primary proficiencies mathematicians want students to learn in advanced mathematics classes (cf., Weber, 2004, Lew, et al, 2018). Mathematics faculty report spending considerable time marking and commenting on student proof-attempts (Moore, 2016), believing that the significant amounts of time spent augmenting grades with detailed comments and corrections will improve student proof-writing. Although prior studies have focused on how professors grade their proofs (cf., Moore, 2016; Miller, et al, 2018), the content and efficacy of written feedback remains largely unexplored. This study explores the types of annotations professors make on student proof attempts, and, offers implications for student learning.

Three studies have explored how undergraduate mathematics faculty evaluate student proof-productions. Moore (2016) identified features that professors value in well-written proofs, including logical correctness, clarity, fluency, and demonstration of understanding. Two studies have found that there are disagreements among mathematics professors about how to assign points to incorrect proofs (Moore, 2016; Miller, Infante & Weber, 2018). For example, Moore (2016) found that disagreements about how severely individual errors should be penalized led to wide score variations among graders who identified the same errors. Miller, Infante & Weber (2018) found that faculty might agree that a vital portion of the proof was missing, but showed little consistency regarding how many points should be deducted. Both studies suggested that the scoring inconsistency – among graders who essentially agree on which portions of proof are in error – can be connected back to a professor's perception of student cognition. Moore (2016) explained that professors "sometimes differ in their evaluation of a student's proof because they differ in their perceptions of what the student was thinking, and consequently they arrive at different judgments on the seriousness of errors" (p. 269). Miller, et al. claimed that "a sizeable minority of the participants would give the benefit of the doubt to a student who they perceived to be strong while the majority of participants would be suspicious of a high quality proof written by a student of low perceived ability" (p. 8). These findings suggest that proof scoring relies on both a professor's perceptions about what constitutes a correct proof, as well as assumptions about the abilities of the student.

There is virtually no research on proof grading as an instructional practice. However, in moving from an analysis of scoring to an analysis of the annotations that professors make on proofs there are a number of concerns that arise. For example, some annotations might be "ticks" to indicate that a particular section has been read (e.g., mainly for the professor's organization) while others might be meant to communicate ideas to the student. We might explore the form and content of these annotations to better understand how and why professors annotate proofs, what they intend to convey to students, and, as a means to investigate what students might learn by reading the annotations professors leave on their proofs. The goal of this study is relatively modest, namely, to explore the form and content of the annotations professors make on student proof attempts. We then draw some inferences about what students might learn from the feedback, but these inferences are meant to be hypotheses that form the basis for further study.

Framing--The Nature of Feedback

Evans (2013) proposed a constructivist model in which an exchange between a professor and a student exists on a Feedback Landscape. When the professor comments on a proof, these comments are created in a buffer zone of mediating social and cognitive factors. The professor's perception of the student's understanding is one such factor. The student receiving the feedback parses the information through a similar buffer zone. It is within these buffer zones that the meaning and utility of feedback can be misconstrued or lost entirely. Glover and Brown (2006) argued the students often cannot derive actable meaning from feedback. In proof-based mathematics, Byrne, Hanusch, Moore, and Fukawa-Connelly (2018) found that students reading professor comments on proofs in a transitions-to-proof course could not reliably describe normatively correct logic for the professor's requested changes, suggesting that even when they could make the changes, that they would not derive transferable learning.

The present study developed and used a coding scheme for professors' annotations as a means to make claims, independent of professor intent or student interpretation, about the form and content of these annotations. We used the coding scheme to address the following questions:

- What types of annotations do professors commonly write on student work?
- What meaning might be conveyed to students by common annotations?
- How do the annotation patterns change over a semester-long course?

Participants & Coding Methods

The study was conducted at a medium-sized, rural, 4-year, public university in the Northeast. The participants consisted of four professors who were identified as teaching a single semester, proof-intensive abstract algebra (n=2) or real analysis (n=2) course. Both courses have an introduction to proof course as a prerequisite, and different semesters of the course should cover the same material.

Each professor volunteered to participate in the collection of students' homework, quiz, and test papers throughout the fall semester of 2017 and the spring semester of 2018. The number of homework assignments collected for fall algebra (n=4) and fall analysis (n=11) was smaller than the number of homework assignments collected for spring algebra (n=24) and spring analysis (n=26). This discrepancy has implications for using raw counts to draw inferences. To protect the students' identities, each student was assigned a numerical identifier that was kept consistent across a semester. The scores of each homework, quiz, and test item, as well as the overall scores on the students' papers, were redacted prior to coding. Each piece of written feedback was numbered so that the coders would be able to uniformly identify what constituted a separate piece of feedback. For each student paper, all the professor's annotations were analyzed by two coders working independently. The separate codes would then undergo a tie-breaking process wherein a third coder would reconcile any discrepancies.

We began the creation of our coding system using Vardi's (2009) coding system that analyzes three aspects of each item of instructor feedback, namely, characteristic, manner and scope. The characteristic category is our evaluation of the content of the professor's annotation; for example, it might be about proof structure, mathematical notation, or validity. While our focus is on the annotation, we also review the student's work in the evaluation. Vardi's characteristic codes proved inadequate for proof writing, so we used a thematic analysis through several iterations to develop new codes. In our system, each characteristic code is hierarchical, with a general group code and a detail subcode. We present a summary table of our characteristic codes in Table 1, and Table 2 shows an excerpt of our coding manual for three detail subcodes.

Table 1. General Characteristic Codes with Definition and Detail Characteristic List

Characteristic - General	Characteristic – Detail
General Academic Feedback (GAF) - Feedback that is non-specific to either proof or subject.	Mechanics (Mech)
	Completion (Comp) Sources (So)
	Fundamental Math Skills (FMS)
General Proof Feedback (GPF) - Feedback relating to the clarity and logical construction of proofs	Mathematical Language & Notation (MLN)
	Proof Framework (PF)
	Proof Presentation (PP)
	Referencing (Ref) Validity (V)
Content Specific Feedback (CSF) - Feedback on the nature and usage of content specific to the class being analyzed.	Subject Matter Notation (SMN)
	Def/Thm Content Statement (CS)
	Def/Thm Choice (DTC)
	Def/Thm Operationalization (Op)
Other Feedback (OF) - Feedback which fits none of the categories above.	Omission (Om)
	Overall (All)
	Other/Unclear

Table 2. Excerpt of coding manual

General Characteristic	Detail Characteristic	Detail Definition
General Academic Feedback (GAF)	Fundamental Math Skills (FMS)	Feedback which addresses skills and symbols from prerequisite, non-proof, mathematics courses, including algebraic manipulation and trigonometric facts.
General Proof Feedback (GPF)	Mathematical Language Notation (MLN)	Feedback on math language and notation, including idioms and phrases, set theory notation, functions and notation of non-specific symbolic logic.
Content Specific Feedback (CSF)	Subject Matter Notation (SMN)	Feedback on notation specific to the subject or course, or that repurposes previous notation in a subject specific way.

The manner codes describe the methods by which the professor expresses feedback, and the scope codes describe the breadth of application of an item of feedback within the proof. Many of Vardi's manner and scope codes were transferrable to the context of proof writing despite originally being designed for traditional academic writing. We show our manner codes in Table 3 below.

Table 3. Manner Characteristic Codes with Definition

Manner Codes	**Description**
Direct Edit (DE)	Feedback where the instructor directly edits the student's work. Something must be crossed out or inserted.
Explanation (Exp)	Feedback that explains why a change is required, that explains a mathematical concept, or explains the marker's reasoning or thinking
Prescription (Pre)	Feedback that prescribes a change to be made by the student. The change needs to be described, but not done for the student.
Question (Q)	Feedback in the form of a question.
Question Mark (?)	A question mark without other text, possibly accompanied by an underline, circle or other indication.
Comment (Com)	Feedback that makes an observation about the proof production, but does not indicate a specific correction.
Indication (Ind)	Feedback that indicates an aspect of the proof, but provides little other information, such as underlining or circling.
Evaluation (Eval)	Feedback that provides an evaluation of the student's work, such as "good" or "weak." X's over student work will fall into this category.
Personal comment (PC)	Feedback that addresses issues outside of the work, e.g., "I hope you're feeling better."
Checkmarks (Chk)	Feedback in the form of indicative marks. Often used for scoring purposes.
Other/Unclear	Any feedback that does not fit into the above categories

For this report, the scope codes we report on are *local* and *global*. *Global* feedback was directed to the proof as a whole, whereas *local* feedback appeared to be directed at a piece of text that was part of a sentence, an entire sentence, or a few lines. The critical distinguishing feature of *local* feedback was that the text could not be considered a proof of any proposition on its own (e.g., could not show that an operation was closed on a particular set), even if that proposition might be a subproof in context of the proof task. We note that our coding system includes an intermediate, *regional*, code, which we do not analyze here.

We made the decision to allow multiple codes within the characteristic and manner dimensions for a particular piece of feedback. We found this critical because in the context of abstract algebra or real analysis, there are both content-specific and “generic” proof proficiencies required to produce a proof.

At the end of the coding process, we analyzed the collected codes in several different ways. Initially, we computed the frequencies and relative frequencies of the four main code types (general characteristic, detail characteristic, manner, and scope) on each assignment (homework, quiz, test). We used these frequencies for a longitudinal analysis. We also explored coding patterns on different types of assignments such as homework, quizzes, and tests.

Data and Results

Examples of Coding

We first illustrate our coding scheme with 2 annotations that the professor made, then explain some patterns in coding we observed. For example, consider the professor’s note shown in Figure 1 which makes an observation about the student’s proof.

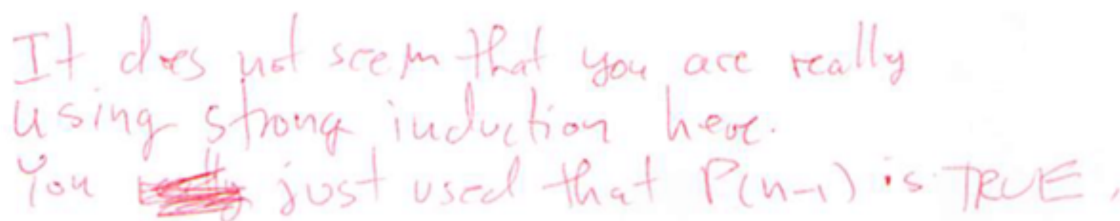

Figure 1. Professor’s note about induction

We code this as *general proof feedback* (GPF) with a detail code of *proof framework* (PF) because the focus of the professor’s comment is on the difference between strong and weak induction. We assigned the manner code of *explanation* because it cites a specific fact that distinguishes strong from weak induction, “you’ve just used that P(n-1) is TRUE” to justify the need for a change. Finally, we label it *global*, because the comment applies to the entire proof.

As a second example, consider the annotation shown in Figure 2, which edited the student’s proof by inserting a symbol for union between *S* and *T*.

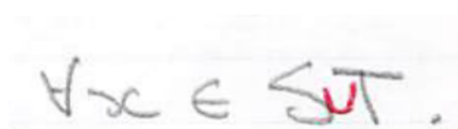

Figure 2. The professor inserted ‘∪’ between *S* and *T*

The characteristic is *general proof feedback* (GPF) with a detail code of *mathematical language and notation* (MLN). We coded this as GPF because set theory is part of an introduction-to-proof course, and MLN because it focused on the symbolic language of mathematics. While we acknowledge that one could read this as a logical issue in that without the annotation, the sentence is not grammatical and therefore could not be interpreted, we argue that the professor appears to treat it as a "typo" where the student's meaning was clear, but missing a "word." We coded this as a *direct edit* (DE) because the professor edited the student's work by inserting the needed symbol, and as *local* (L) because it addressed a piece of content within a mathematical sentence.

Since the focus of the coding system is instructor feedback–as opposed to student errors– the coders attempted to divorce their choices of codes from the content of student proofs whenever possible. However, student content was considered in cases where the meaning of the feedback would be altered by the context of the proof. For instance, when an instructor added a symbol, such as in the above example, it would be impossible to identify the content without making an interpretation of the student's work. As a final example, we note that there are a number of annotations that professors commonly made that we did not feel we could assign content meanings to, such as a single question mark, a checkmark, or even the question "What?" and, as a result, we would code such annotations as *other* or *unclear*.

Checkmarks

One pattern shared by all professors across this study was the usage of annotations such as checkmarks which tended to make up 50% or more of the recorded comments for each coded assignment. These annotations are often informal grading tools which instructors use to tabulate scores rather than attempts at purposeful feedback to students, or they might simply be indications by the professor that she has read a particular exercise. For these reasons, this form of acknowledgement was specifically rejected as meaningful feedback for the purposes of this coding system.

Manner and Scope Types Make it Difficult for Annotations to Convey Information

The vast majority of all feedback made by the instructors from this study was *local.* No professor gave less than 78% *local* feedback, while *global* comments ranged from 3% to 17%. This may be indicative of the fact that most students had correct "big picture" ideas and structures, which might be because most items were on homework, allowing students to spend significant time and even ask for help from the professor and classmates. We further note that most local feedback did not convey information about why the professor made the annotation. The procedural was often emphasized over the conceptual, meaning that the content of annotations was, for example, focused on correct use of notation or the presentation of the proof, while explanations for why were generally absent. Even written comments focused on individual steps being taken or errors being made rather than broad feedback about how concepts were being understood. For each instructor *direct edit* was the most popular choice of manner, as exemplified in Figure 2. When *direct edits, prescriptions, and evaluations* were combined, they ranged from 43% to 71% of feedback manners, none of which explained why a change was needed. Additionally, nearly all feedback was made about incorrect work to convey a needed change. The most common positive responses were nonspecific comments such as "OK" or "Good," again without explanation of what caused the professor to evaluate it that way.

The most common characteristics were *proof presentation, validity,* and *operationalization*. We interpret the prevalence of these characteristics to be connected to the local, task-oriented nature of most of the coded comments. It is possible to make global, conceptual comments about the structure of how a proof should be written, or about the types of logical arguments that are valid. However, most often these characteristics were used for line-to-line error correction or identification.

Variations Over Time

In analyzing the occurrence of general characteristics over the course of a semester, we identified two patterns which we interpreted to indicate substantive differences between how feedback is given in algebra and analysis. The relation between *content specific feedback* and *general proof feedback* seemed to follow two subject specific patterns. All four courses began with a relatively small percentage of *content specific feedback* (less than 25%). We interpret this to be connected to early lessons designed to reintroduce students to format, structure, and logic of proofs rather than specific subject-based notation, theorems, and definitions. In both algebra classes the rate of content specific feedback rose to about 50% with a corresponding decrease in *general proof feedback*. In contrast, both analysis courses developed "bubbles" of content specific feedback which began in small percentages early in the semester, peaked around the middle of the semester, and shrank toward the end. Without making any specific causal links, these patterns could be interpreted as showing that future assertions about the quality of proof feedback for a given subject may not be immediately transferable to another proof-related class with different feedback needs. Further research is needed to better understand these patterns and whether faculty are purposeful about them. We note, for example, that the professor, Dr. T., in Weber's (2004) study wanted students to quasi-mechanically write proofs in the beginning of the semester in real analysis and only later develop the ability to write proofs with meaning. A professor like this might, at the beginning of a semester, purposefully make "behavioral" comments on proofs, editing or prescribing changes without explanation and as the semester progresses focus more on content and use explanatory concepts.

Discussion

The primary work here, of developing and implementing a coding scheme for professor proof annotations, has allowed us to note trends in annotating among four professors who teach abstract algebra and real analysis. In particular, we noted high rates of annotations that cannot meaningfully convey worthwhile information to students (e.g, checkmarks, question marks, ...). There were also a large percentage of annotations that indicated an error, and sometimes a correction (e.g, a *direct edit*), but without an explanation either of what the mistake was or why the correction was needed. Prior research by Byrne et al. (2018) suggests that students can correctly use these types of annotations to revise the given proof, but cannot explain why the corrections are needed. As a result, it seems unlikely that the annotations will lead to students changing their practice on future proofs, although this certainly warrants further study. Moreover, the trends identified were among faculty at a single university, so we should further explore how common they are. Similarly, we call for significant research exploring why faculty annotation students' papers, what they intend to convey to students, and what, if anything, they hope students will do in response.

References

Byrne, M., Hanusch, S., Moore, R. C., & Fukawa-Connelly, T. (2018). Student Interpretations of Written Comments on Graded Proofs. *International Journal of Research in Undergraduate Mathematics Education*, *4*(2), 228-253.

Evans, C. (2013). Making sense of assessment feedback in higher education. *Review of educational research*, *83*(1), 70-120.

Glover, C., & Brown, E. (2006). Written feedback for students: too much, too detailed or too incomprehensible to be effective?. *Bioscience Education*, *7*(1), 1-16.

Lew, K., Fukawa-Connelly, T. P., Mejia-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.

Miller, D., Infante, N., & Weber, K. (2018). How mathematicians assign points to student proofs. *The Journal of Mathematical Behavior*, *49*, 24-34.

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.

Stylianides, A., Stylianides, G., & Weber, K. (2017). Research on the Teaching and Learning of Proof: Taking Stock and Moving Forward, In Cai, J. (ed). *Compendium for Research in Mathematics Education* (pp. 237-266). Reston, VA: National Council of Teachers of Mathematics.

Vardi, I. (2009). The Relationship between Feedback and Change in Tertiary Student Writing in the Disciplines. *International Journal of Teaching and Learning in Higher Education*, *20*(3), 350-361.

Weber, K. (2004). Traditional instruction in advanced mathematics courses: A case study of one professor's lectures and proofs in an introductory real analysis course. *The Journal of Mathematical Behavior*, *23*(2), 115-133.