Variations in Grading and Instructor Feedback

Anna Marie Bergman
Fort Lewis College

Andrew Kercher
Simon Fraser University

Although developing proof proficiency is an integral part of undergraduate mathematics, little is known about instructor grading practices and feedback conventions in undergraduate courses in which proof is taught. In recent years, small-scale studies have begun to shed light on how individual instructors at single institutions assign points and provide feedback on proof-related activities. Our study extends this work by describing the results of a large-scale online survey of mathematics educators (n = 86) from multiple institutions. Participants were asked to score and give feedback to hypothetical student submissions of a proof-adjacent assignment; we investigate participants' responses and provide a taxonomy of their feedback.

Keywords: Feedback, Grading, Proof-related activities, Teaching Practices of Mathematicians

Proof is widely agreed to be central to the activity of mathematicians, and therefore argumentation and proof-related activities are an integral part of undergraduate mathematics education. Developing proof proficiency "is often the primary goal of advanced mathematics courses and typically the only means of assessing students' performance," (Weber, 2001, p. 101). In light of this, mathematics educators spend significant time evaluating and responding to students' mathematical arguments—and yet, little is known about instructor grading practices and conventions for feedback (Moore, 2016). In fact, until recently, little has been known about the teaching practices of undergraduate mathematics educators in general (Speer, Smith, & Horvath, 2010). Small scale exploratory investigations have begun to shed light on the criteria professors use to grade proofs (Moore, 2016; $n = 4$ $n = 4$), assign points (Miller et al., 2018; $n = 8$ $n = 8$), and provide feedback (Kontorovich, 2021; $n = 1$ $n = 1$). Findings indicate that proof grading is a complex teaching practice requiring difficult judgements that lead to substantial variation in points assigned and a need for further research.

Our study aims to add to the growing body of literature on undergraduate mathematics teaching practices by investigating the scores and feedback provided to a proof-adjacent activity on a large scale. We conducted an online survey of mathematics educators in which we provided 5 different student submissions analyzing a single mathematical statement. Survey participants were asked to score each submission and provide feedback to the student. Responses to the survey were used to investigate the following research question: *What kinds of feedback accompany different numerical scores on an undergraduate, proof-adjacent mathematics task?*

Background

In general, "feedback is conceptualized as information provided by an agent (e.g., teacher) regarding aspects of one's performance or understanding" (Hattie & Timperley, 2007, p.81). There is an extensive body of research on feedback, including reviews of the literature and meta-studies investigating the effectiveness feedback (Black & William, 1998; Shute 2008; Hattie & Timperley, 2007). Many of these studies span disciplines and provide frameworks for describing particular features and functions of feedback. For example, Shute (2008) compiles several frameworks in her review differentiating *directive* feedback which tells the student what to fix; from, *facilitative* feedback which provides comments or suggestions without explicit direction (Black & Wiliam, 1998, as cited in Shute, 2008). Feedback can also contain elements of both *verification* and *elaboration*. Verification acknowledges whether or not the work is correct, whereas elaboration is a description of the work designed to lead the student to a better answer (Kulhavy & Stock, 1989, as cited in Shute, 2008). While many of these frameworks offer general ideas about feedback, they are often drawn from studies based in secondary or primary school settings and feature participants who are engaged in well-defined tasks or well-structured problems. Literature on feedback in undergraduate settings that features ill-structured or open-ended tasks with multiple solution strategies is needed.

Methods

Data Collection

We conducted an online survey of mathematics educators in which we provided 5 different student submissions to the assignment pictured in Figure 1.

For the following statement, indicate whether it is mathematically true or false. If you claim that the statement is false, refute it with a counterexample. Explain the relevance of your example to the statement and how it disproves the statement. If you claim that the statement is true, prove it.

There exist four different prime numbers the sum of which is prime.

Figure 1. The assignment that engendered the student submissions featured in the online survey.

Survey participants were prompted to "Imagine that you were teaching a math class, and you received the following five homework submissions as an argument or justification for whether the statement ['There exist four different prime numbers the sum of which is prime'] is true or false." For each submission, participants were asked to score it out of 10 possible points and to also provide written feedback for the student. No scoring rubric was provided. The student submissions in question are pictured in Table 1.

Table 1. The five student submissions assessed by participants in the study.

Participant	Submission
Alpha	True. In order to prove this exists we only have to show one valid example of the statement. *a, b, c, d, e* in "primes" $a + b + c + d = e$ Example: 2 + 5 + 7 + 17 = 31 2, 5, 7, 17, and 31 are all primes. Therefore, there does exist a case where four distinct/ different primes sum to equal a prime number.
Beta	True. For example, consider the sum of prime numbers 2 + 3 + 7 + 11 = 23 which proves that there exists four different prime numbers the sum of which is prime. But there also exists four different prime numbers the sum of which is not prime. For example, 2 + 5 + 7 + 11 = 25.
Gamma	True. 2, 3, 5, and 7 are all different positive primes. Their sum, 2 + 3 + 5 + 7 = 17, is also a prime number. This example proves the initial statement that there exists a (at least one) solution. Note that 2 must be one of the primes in the sum: 2 is the only even prime number, so if none of the four primes is 2, then we have odd + odd + odd + odd = (2a + 1) + (2b +1) + (2c + 1) + (2d + 1) = 2(a + b + c + d + 2) = 2k

	where k is an integer, hence the result is even and cannot be prime (unless k = 2, except all odd primes are greater than 2, so the sum will be greater than 2 making this not possible).
Delta	True, 2 + 3 + 5 + 7 = 17, all primes.
Epsilon	This statement is true. Proving by contradiction, let us assume that this statement is false, then it becomes, "There don't exist four different prime (positive) numbers the sum of which is prime." We can disprove this with a single example, such as 2, 5, 7, and 17, which sum to 31, a prime number. As the inverse statement is false, we can see that the original statement, that there exists at least one group of prime numbers that sum to a prime, is true.

The student submissions in Table 1 were inspired by real student work and only slightly re-worded for clarity. These particular submissions were chosen because they all correctly identified the statement as true and provided the necessary example. However, they also represented a variety of different, sometimes non-standard approaches to mathematical proving.

In total, 86 participants provided at least a numerical score to each of the 5 student submissions. Of the 430 individual numerical scores, 69 had no accompanying written feedback. Therefore, in total, we analyzed 361 individual *complete responses* in the sense that they contained both a numerical score and accompanying written feedback. Delta received the fewest complete responses (n = 69), whereas Gamma received the most (n = 74). The other student submissions each received 72 complete responses.

Participant Demographics

Two populations of mathematics educators were invited to participate in the online survey through email invitations distributed via listserv. The first population was comprised of mathematics educators and mathematics education graduate students at an R1 institution in Canada; the second population was an international group of professionals who were subscribed to a listserv dedicated to promoting research in undergraduate mathematics education. Participants were only asked demographic questions related to their education and teaching experience. More specifically, participants described the highest levels at which they had taken courses in mathematics and mathematics education. They also provided the highest levels at which they had taught courses in mathematics and mathematics education (see Table 2).

Table 2. Participants' experience as students and instructors of mathematics and mathematics education.

	Graduate	Undergrad.	Secondary	Primary	None of these
Experience as a student:					
Mathematics	71	15	-	-	-
Math Education	62	6	-	-	18
Experience as an instructor:					
Mathematics	25	46	15	-	-
Math Education	43	14	-	1	30

Data Analysis

To analyze the content of the complete responses, they were subjected to two rounds of coding. In the first round, each researcher used open-coding techniques to apply many descriptive codes (Saldaña, 2016) to the data; these codes sought to comprehensively capture the variety and nuance within the data from multiple perspectives and along several axes. The research team then came together to discuss the variety found in the content. Together, they organized their independently generated codes into a unified set of codes that captured the types of feedback given to the student characters. In a second round of coding, each researcher independently recoded the entirety of the dataset, this time using the newly created codebook. Discrepancies between these codes were discussed on a response-by-response basis until consensus was obtained and complete intercoder reliability was achieved.

In this report, we present a subset of feedbacks that the research team recognized as *actionable feedbacks*. We interpreted feedback as actionable if the student to whom it was directed could interpret that feedback as suggesting a course of action that would bring the submission closer to an implied mathematical standard for proof-writing.

Findings

We identified 62% of written feedback as containing an actionable element. At the broadest level, we categorized actionable feedback as either *prompting an addition, prompting a reduction*, or *prompting a change*. Note that these three types of feedback (and each of their respective subcategories, described in the following sections) are not treated by the research team as mutually exclusive categories. In this study, however, overlap was observed to be minimal.

Feedback Prompting an Addition

Feedback prompting an addition recommends, implicitly or explicitly, that something be added to the student's submission to improve its overall quality. Within this type of feedback, four prominent subcategories arose. First, and most broadly, some feedback that prompted an addition simply asked the student to add more without any specific direction about what type of addition would be appropriate; this was a general prompt.

When the feedback prompting an addition made a specific recommendation to the student, it fell into one of the three remaining subcategories. The addition could be an *epistemological justification*; this was a common suggestion when the participant did not feel that the student submission adequately demonstrated why the proposed statement was in fact true. The addition might instead be a *structural justification*, which (in the eyes of the participant providing the feedback) would lead to a submission that had clearer internal logical structure and purpose. Finally, some participants directed the students to provide *additional insight*—that is, the feedback suggested that the submission would be improved if the student added details of their problem solving, examined related corollaries, or demonstrated a deeper understanding of the mathematics than might be required to establish the truth of the statement. Examples of each type of feedback prompting an addition are provided in Table 2.

Table 2. The types of feedback prompting an addition.

Feedback Type	n	Example
General	6	"You should explain a bit more."
Epistemological Justification	24	"Why is it sufficient for you to identify just a single example to establish the statement as true?"
Structural Justification	33	"The argumentation is correct but you should express it through a discourse, and make logical relationships clearer."
Additional Insight	69	"You have answered the question correctly. Can you expand by giving another three examples? What do all of your examples have in common?"

Feedback Prompting a Reduction

Feedback prompting a reduction pinpointed a particular area in which the student's submission could be cut down in some way, rather than expanded. This type of feedback manifested in two ways. First, the reduction might target *excess information*; this type of feedback can be interpreted as the antithesis of feedback prompting the addition of additional insight. That is, whereas some participants posed questions in their feedback that solicited mathematical details that did not directly contribute to a proof of the statement at hand, other participants specifically requested that students remove such details. Second, the reduction might require removing *excess machinery*. In this type of feedback, the respondent typically felt that the information presented in the student submission was relevant but that the proof structure or level of mathematical formalism used to convey that information was unnecessary and could be reduced. Examples of each type of feedback prompting a reduction are provided in Table 3.

Table 3. The types of feedback prompting a reduction.

Feedback Type	n	Example
Excess Information	44	"It suffices to show that one case exists. It's not necessary to show that there are sums of primes for which the property does not hold."
Excess Machinery	51	"Your proof is correct, but it's unnecessarily cumbersome. The original statement is a claim about a single case existing. The most direct proof then just gives an example. The extra layer of disproving the negation is a bit confusing for a reader."

Feedback Prompting a Change

Finally, feedback prompting a change suggested that some already existing aspect of the work was relevant, and thus should not be removed, but instead should in some way be modified. Feedback prompting a change typically targets one of two types of perceived shortcomings. On one hand, the reviewer may have felt that the mathematical content of the student's submission needed to be changed. This type of feedback typically attended to a student's misuse of mathematical vocabulary or symbols; often, it pointed out what the respondent perceived to be a "typo" on the part of the student. On the other hand, the reviewer may have called attention to a way in which the organizational or grammatical formatting of the submission was inadequate. Examples of each type of feedback prompting a change are provided in Table 4.

Table 4. The types of feedback prompting a change.

Feedback Type	n	Example
Mathematical Content	34	"Terse and correct except for the error in mathematical grammar. The equation is a true assertion about primes, but an equation is not a prime."
Formatting	9	"Nice addition. In the future, please separate the proof from additional insight, and mark the proof as such."

Frequency of Actionable Feedback Types

In Table 5, we present the frequency with which of each type of actionable feedback was provided to each student submission. In this table, we also coordinate the type of actionable feedback with whether or not the participant awarded the student submission in question with a full or partial score (FS and PS, respectively). For example, of the 19 feedbacks that prompted Alpha to include additional insight, 9 of these still awarded his submission a full score; another 10 awarded him less than a full score.

Table 5. The frequency of each type of actionable feedback given to each student submission.

Feedback Type	Student									
	Alpha		Beta		Gamma		Delta		Epsilon	
	FS	PS	FS	PS	FS	PS	FS	PS	FS	PS
Prompting an Addition										
General	-	-	-	-	-	-	-	6	-	-
Epist. Justification	1	1	1	3	-	-	2	15	-	1
Str. Justification	2	1	4	3	1	1	1	12	-	8
Additional Insight	9	10	6	9	7	3	6	9	4	6
Prompting a Reduction										
Excess Information	-	-	16	16	8	4	-	-	-	-
Excess Machinery	3	1	-	-	-	-	-	-	25	22
Prompting a Change										
Math Content	3	5	-	1	7	3	-	3	-	12
Formatting	1	1	2	-	2	-	-	3	-	-

Discussion

Miller (2018) observed that, when instructors allotted scores to student-submitted proofs, "the points that they assigned was dependent upon their perceptions of participants' understanding of the proof that they submitted" (p. 30). As seen in Table 5, certain types of actionable feedback are more likely to be attached to partial numerical scores; prompts for additional epistemological justification, for additional structural information, and for a change in mathematical content. These types of actionable feedback appeared 83%, 76%, and 71% of the time (respectively) with a partial numerical score. In light of Miller's finding, the association of these types of feedback with partial scores could indicate which types of errors instructors tend to recognize as indicative of a fundamental misunderstandings on the part of the student author.

On the other hand, actionable feedback prompting a reduction of excess information or machinery did not clearly correlate to either a full or partial score. Respondents in our study did not clearly agree on whether or not "too much" of something demonstrated that the student fundamentally misunderstood the underlying mathematical logic. Beta and Epsilon received prompts for a reduction in information and machinery (respectively) more than any other student submission received any other type of feedback. However, despite how frequently respondents provided these types of feedback, they did not clearly correspond to a partial score.

We note that prompts for additional insight were also more common on partially scored submissions, but only slightly (at 54%)—and this trend was reversed in the case of Gamma's feedback. In fact, because of Gamma's overall high scores, this was not the only trend they reversed. Gamma's submission was the only student's work to receive mostly full scores even in the presence of a prompt for a change in mathematical content which, as discussed in the previous paragraph, was often attached to partial scores. As described in Moore (2016), some instructors, when grading proof-like student submissions, do not deduct points for each error; instead, they attempt to assign a score "by judging the overall quality of the proof" (p. 263). Thus, while Gamma's mathematical typo generated a number of feedbacks prompting a change in mathematical content, they still often received full credit because respondents perceived that Gamma understood the underlying mathematics of the proof.

Prompts for additional insight comprised a large and especially nuanced portion of feedback. This type of actionable feedback occurred more often than any other, and in 30 of its 69 instantiations, it was the only type of actionable feedback offered to a student who received a partial score. Of these 30 feedbacks, 6 are very open-ended prompts (e.g., "Can you dig deeper?"). The rest are guiding questions that ask the hypothetical student to consider or provide material that is beyond the scope of the original prompt (e.g., "What are the reasons of the choice of your primes?"). Occasionally, the respondent instead presents their prompt for additional feedback in the form of an observation; for example, one feedback explains that "this [Delta's submission] is correct but it offers very little in the way of showing where your numbers come from or telling me how you understand this must be true."

Miller (2018) found that, by some instructors' standards, "correct proofs would not necessarily receive full credit" (p. 29). Our study corroborates this finding. Student submissions were chosen for this study on the basis that they each correctly attended to the task at hand. As Miller observed, however, "correctness is not the only criterion that is used by mathematicians for grading proofs" (ibid, p. 30); the student submissions in this study were also often scored based on how well respondents felt they demonstrated the student authors' understanding of why their work was sufficient and correct. Sometimes, this is clear in the accompanying written feedback—see the example of feedback requesting additional structural justification in Table 2.

Conclusion and Limitations

Our main contribution is a taxonomy of the feedback mathematics educators provided to students in proof-adjacent activities. Beyond a more detailed description of the feedback presented above, the presentation will also explore a type of feedback adjacent to the actionable feedback described in this report: that is, when no actionable written feedback is provided and yet the student submission is only awarded a partial score. We wonder, when instructors provide such feedback, what socio-mathematical norms (if any) are they attempting to reinforce? A future study could also explore what actions students take in response to actionable feedback.

We would like to recognize two limitations to our study. First, respondents were not allowed to physically grade the student submissions—that is, they weren't provided with submissions that they could "mark up" with non-textual feedback. Spiro et al. (2019) found that annotations such as checkmarks or arrows account for 50% or more of the recorded feedback when professors grade assignments. Our survey required participants to type feedback in the form of sentences, which might be different from their usual method of providing feedback to their students. On the other hand, our study could contribute to a better understanding of how instructors' feedback practices change in an online environment.

Second, we agree with the sentiment expressed by Sommerhoff & Ufer, who argue that "the generalizability of results in the context of proofs are always questionable, as they rely on local socio-mathematical norms" (2019, p. 729). From that perspective, our study was limited by the decontextualization of the student work from a greater classroom culture. In fact, some respondents indicated that their treatment of student submissions would depend on the socio-mathematical norms that had been established in the classroom. The feedback given by these participants may not be representative of their actual feedback practices.

References

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: principles, policy & practice*, *5*(1), 7-74.

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

Kontorovich, I. (2021). Minding mathematicians' discourses in investigations of their feedback on students' proofs: a case study. *Educational Studies in Mathematics*, *107*(2), 213-234.

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.

Saldaña, J. (2021). The coding manual for qualitative researchers. *The coding manual for qualitative researchers*, 1-440.

Shute, V. J. (2008). Focus on formative feedback. *Review of educational research*, *78*(1), 153-189.

Speer, N. M., Smith III, J. P., & Horvath, A. (2010). Collegiate mathematics teaching: An unexamined practice. *The Journal of Mathematical Behavior*, *29*(2), 99-114.

Spiro, B., Hanusch, S., Miller, D., Moore, R. C., & Fukawa-Connelly, T. (2019). Categorizing Professors' Feedback on Student Proofs in Abstract Algebra and Real Analysis. *Proceedings of the 22nd Meeting of the MAA Special Interest Group on Research in Undergraduate Mathematics Education*, Oklahoma City, OK, United States.

Sommerhoff, D., & Ufer, S. (2019). Acceptance criteria for validating mathematical proofs used by school students, university students, and mathematicians in the context of teaching. *ZDM*, *51*(5), 717-730.

Strauss, A., & Corbin, J. (1994). Grounded theory methodology: An overview. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (pp. 273 – 285). Thousand Oaks: Sage Publications.

Weber, K. (2001). Student difficulty in constructing proofs: The need for strategic knowledge. *Educational studies in mathematics*, *48*(1), 101-119.