Assessment of Student Understanding: Exploring Faculty Optimization Problem Grading

David Miller
West Virginia University

Jesse Cook
West Virginia University

Kaitlyn Roxby
West Virginia University

William Robertson
West Virginia University

Edwin (Ted) Townsend
West Virginia University

In this study we examine how faculty members grade optimization problems from calculus. Eight faculty members were interviewed and asked to grade three student solutions for three different optimization problems. In addition, faculty members also discussed their grading philosophies and practices for both optimization problems and in general. We used open thematic analysis to analyze all faculty interviews and two major themes of reflection of understanding and key steps in optimization solutions emerged from the data. These themes that was supported by faculty were further analyzed for the interplay between them. In addition, we delve deeper into reflection of understanding using Skemp's instrumental and relational understanding to establish when faculty assessed good student understanding of optimization problems. Our results provide insight into how mathematics faculty grade student solutions for optimization problems, how they emphasize student understanding, and offer possible insights into the variations in scores we observed in this study.

Keywords: Assessment, Calculus, Optimization, Understanding

Faculty in many mathematics departments are needed to teach calculus courses so they can cover all mathematics courses they offer each semester. As part of their teaching, faculty normally grade homework, quizzes, and exams. Computerized homework that is graded automatically is commonly used, but faculty is still responsible for grading paper and pencil assignments such as quizzes and exams. Grading student solutions is important and crucial as grading communicates instructor's anticipated standards to students (Stiggins & Chappuis, 2005; Yerushalmi et al., 2015), as well as guiding students on how to enhance their work to improve their understanding of the subject material. Black and Wiliam (1998) noted that "feedback on tests, seatwork, and homework should give each pupil guidance on how to improve" (p. 144). Further, effective grading practices can provide feedback that can foster learning, communicate to students what practices and skills help them: learn a discipline, develop problem-solving skills, prepare for formative and summative assessments, and communicate to students about what to focus on to better prepare for future activities (Black and Wiliam, 2009; Marshman et al., 2018; Hattie & Timperley, 2007; Mulliner & Tucker, 2015; Wiliam, 2006). Having consistent grading is also important for ensuring fairness among students. Thus, the feedback and grading provided to students by instructors on their solutions plays an essential role in shaping their learning experiences.

Despite the importance of grading practices of instructors, there have been few studies on the topic. Several prior studies have focused on the evaluations of students' proof in theoretical mathematics classes by professors (e.g., Moore, 2016; Miller et al., 2018) as well as the grading practices of physics professors on students' problem solutions in calculus-based introduction to physics classes (e.g., Docktor et al., 2016; Henderson et al., 2004). None of these studies investigated the grading practices of mathematics faculty in calculus. Thus, our study will specifically investigate the grading practices of mathematics faculty, and in particular, how they grade calculus students' solutions to optimization problems. The following questions in the context of optimization problems in calculus guide our research: 1) How do mathematics faculty assign points to student solutions?, 2) What features or details do mathematics faculty value when grading?, and 3) How do mathematics faculty assess student understanding when grading?

Literature Review and Theoretical Framework

While studies regarding the grading of student solutions have been on the rise, most prior research lies in the realm of physics (Docktor et al., 2016; Henderson et al., 2004; Marshman et al., 2020). One of the earliest such studies by Henderson et al. (2004) tasked 30 collegiate physics instructors to evaluate two student solutions resulting in the same (correct) answer. This study indicated that instructors were apt to deduct points for flawed intermediate logic with minimal supporting work, while graders may be reluctant to deduct points from a student solution that might be correct. Further, it was noted graders tend to project correct thought processes, indicating that the graders were placing the 'burden of proof' on themselves rather than the student. Of the six graders interviewed in depth in the Henderson et al. (2004) study, only one grader placed the 'burden of proof' on the student and awarded points to a student solution when the student showed explicit evidence of understanding. The other graders placed the 'burden of proof' on themselves and only deducted points when there was clear evidence the student did not understand. Thus, it was observed that clear evidence of understanding was not always needed to award points.

To obtain more consistency in grading solutions, Docktor et al. (2016) developed a problem-solving rubric encompassing five aspects of a desired student solution including descriptions of the problem, mathematical procedures, and appropriate logical progression. Such a rubric provides a solid foundation for other practitioners to implement either holistically or analytically.

In a similar vein, Marshman et al. (2020) implemented a professional development seminar for GTAs enrolled in physics doctoral programs. During their first semester, new GTAs met with each other and the authors of the study to discuss grading strategies and philosophies. Interestingly, the GTAs involved in this seminar believed grading was a formative assessment designed to enhance student learning with feedback. As the seminar progressed, however, it became evident that the GTAs focused on the correctness of a solution rather than treating the assignment as a learning experience. This result corroborates the study by Henderson et al. (2004) who found that graders can experience internal conflict when assigning grades.

Another study conducted by Marshman et al. (2017) implemented a similar professional development seminar as Marshman et al. (2020). The Marshman et al. (2017) study focused on how GTAs developed teaching examples, and if the assistants valued the students implementing the same problem-solving strategies found within the teaching examples when grading student solutions. In addition, they investigated if the GTAs' grading strategies differed between introductory physics and quantum mechanics. They found that the GTAs valued correct answers over explicit problem-solving strategies when grading introductory physics, but the reverse was true for quantum mechanics. Moreover, the GTAs graded the quantum mechanics solutions harsher than the introductory students (Marshman et al., 2017). The GTAs stated that they graded this way because the students in quantum mechanics were more experienced and should know the basic physics concepts. Echoing the findings previously discussed, the GTAs viewed grading as a summative assessment and not a formative teaching tool (Marshman et al., 2017). Therefore, it appears important to investigate if mathematics instructors grade calculus students for correct answers or provide formative feedback.

Other studies researched how professional mathematicians grade students' proof-writing skills. Moore (2016) interviewed four professors and noted remarkable variation in scores contingent upon the perceived student's intent within the proof. The four professors discussed the importance of logical correctness, clarity, and fluency in good proof writing and differed in their evaluations due to seriousness of errors, the students' demonstration of understanding of the proof, the validity of the proof and the associated mathematical concepts. Similarly, Miller et al. (2018) presented nine mathematicians with sample proofs from a transition-to-proof course. This study indicated that professors offer 'the benefit of the doubt' to student proofs that exhibit logical gaps. This 'benefit of the doubt' extended to instances where generic proofs would be accepted as full proofs, and proofs deemed "correct" received less than full credit. Such results indicate inconsistency in how professional mathematicians grade in proof level courses, so it would be prudent to investigate how instructors grade problems in calculus.

We use a theoretical framework of understanding by Skemp (1976), where student understanding is broken down into two categories, 'instrumental' and 'relational'. Instrumental understanding refers to students having certain rules or facts in mind, such as specific formulae or procedures, without having the underlying knowledge or understanding of why these rules work. Conversely, relational understanding refers to a deeper type of understanding, where the underlying connections and rationale behind knowledge is known, allowing students to be able to more easily and appropriately apply their knowledge to new situations. Skemp maintains that students gaining relational understanding is preferred over students only having an instrumental understanding. However, many students only acquire instrumental understanding as many classroom environments do not promote relational understanding. We will use this framework to analyze the participant's grading of the optimization problems and their interpretations of student understanding of these solutions.

Methodology

Participants

For this study, eight mathematics faculty members from a large state university in the United States consented to participate in this study during the Spring of 2024. The participants all had experience teaching calculus for the university that included grading exams and other assignments. The participants were representative of all the levels of tenure-track faculty and the assistant and associate levels for teaching professors. Also, one participant was a postdoctoral researcher with three years' experience teaching calculus, along with additional experience as a graduate student at a large state university. Participants were from a variety of fields in mathematics such as Algebra, Graph Theory, Discrete Mathematics, Partial Differential Equations, and Undergraduate Mathematics Education.

Procedure

The eight participants were interviewed in the spring of 2024. They are listed as F1 to F8 in this study and referred as a neutral gender to mask their identity. The interviews were conducted via Zoom. Each participant was presented with three optimization problems to grade, with three solutions (labeled S1, S2, and S3) per problem. The interviews were conducted in two sessions, each taking approximately an hour, depending on the length it took for the participants to grade the papers. A draft of the interview transcripts was automatically generated via software using the recorded video and then polished up by the researchers after reviewing each recorded interview. The solutions for the problems were generated by the authors and the prior studies in physics, and the participants were asked to grade them as if they were grading a calculus assignment in the online grading system called Crowdmark. While they were grading the questions, the participants were encouraged to provide justifications on what they were doing, how the grade was determined, and their final score. Additionally, before the participants graded the student solutions to the three optimization problems, they were asked semi-structured open-ended general interview questions about their grading philosophies and practices, as well as specific approaches when grading optimization problems.

Materials

The nine solutions we devised for the interviews consisted of three problems with three solutions S1, S2, and S3 (see Figure 1 for an example solution of the rectangle problem). One of these problems, the Rectangle Problem, is a standard problem asking how to minimize the cost of a fence enclosing a rectangular, partitioned area. The other two optimization problems were the Barrel Problem and the Pipeline Problem, from a previous Calculus exam conducted at the university and are common types of problems in Calculus textbooks. The Barrel problem dealt with optimizing the volume of a cylinder with specific surface area and the pipeline was about building the minimum cost pipeline to a refinery over land and across a river. While the fundamental structure of all three problems is similar, the difficulty in understanding the scenarios and the algebra involved varies between the problems, with the Rectangle Problem being the easiest, followed by the Barrel and Pipeline problems. The problems and their solutions can be shared upon request.

The solutions were designed to highlight the issue of how the faculty members would assign partial credit and motivated by the Physics Education research of how GTAs and faculty grade calculus-based physics problems (Henderson et al., 2004). For the Rectangle and the Pipeline problems, all the final solutions are accurate, however the amount of work shown varies between the solutions, with many of the solutions having what the authors deemed as an insufficient amount of work. The solutions for the Barrel problem on the other hand has a variety of errors and different answers due to those errors. Note that all solutions have approaches which are generally correct. They each find a formula for the quantity to be optimized as a function, compute the derivative, and then determine where the derivative is equal to zero to find the critical points. However, further steps such as verification of the solution and showing appropriate algebra or calculations to justify the work are sometimes brief or omitted. The variety in the amount of work shown and correctness was done to mimic real student work on optimization problems in order to provide the graders with authentic, realistic grading scenarios they are likely to encounter with optimization problems.

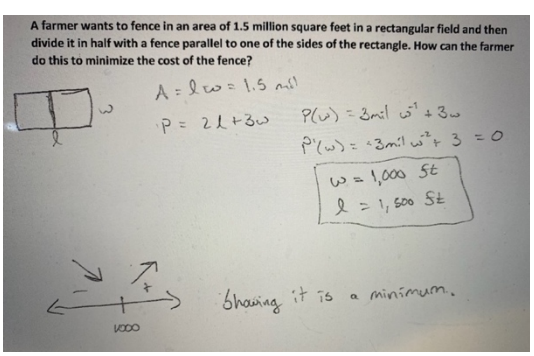

Figure 1. Optimization Solution S1

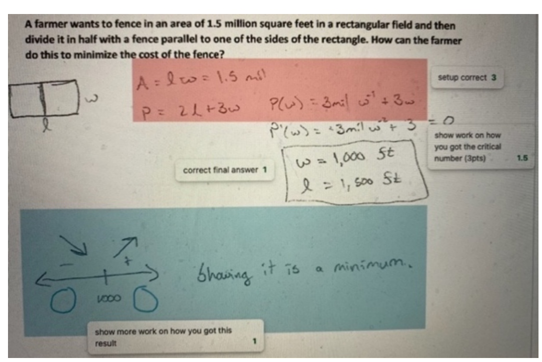

Figure 2. Solution S2 on the rectangle problem

Analysis

The authors engaged in open thematic analysis (Corbin & Strauss, 2014) in the following way. First, all authors individually read over all the transcribed data while identifying and highlighting passages that might be of theoretical interest concerning grading. All authors met to discuss and compare their findings and developed a list of codes for further study. Next, each author reread the transcripts again, searching for each topic of interest; we then put these excerpts into a file of comments related to that code. Through this process, we developed criteria and descriptions of these themes. Finally, the authors individually went back through the data one last time to identify passages that would satisfy the criteria for each of these themes.

Results

Quantitative Data

Table 1 below shows the scores that the eight-mathematics faculty assigned for the three different optimization problems solutions, S1, S2, and S3, on the three optimization problems. In addition, the average, median, and variances for each solution are shown at the bottom of the table. We can see that the largest variances occurred on the rectangle optimization problem with student solutions 1 and 3 and the smallest variances occurred on the barrel and pipeline problems with student solution 3. The rest of the variances were somewhere in between. Furthermore, four of nine solutions typically scored high with an average of over 8, two were more problematic solutions with an average score below 5.5, and the remaining three of nine solutions were somewhere between.

Table 1. Faculty Member Scores for the 3 Optimization Problems and 3 Solutions S1, S2, and S3.

	Question 1 (Rectangle)			Question 2 (Barrel)			Question 3 (Pipeline)		
Grader	S1	S2	S3	S1	S2	S3	S1	S2	S3
F1	4.5	7	6.5	8	9	9.5	4	10	7.5
F2	5	6	8	7	10	8	5	10	8
F3	4	8	7.5	9	9	9	4	10	7
F4	6	9	9	9.25	9.5	9.5	6	10	7.5
F5	4.5	8	9.5	9	10	9	7	8	8
F6	7	7.5	7	6.5	8	8	4	8.5	7
F7	4	7	5	7	8	8	4	7	6
F8	9	10	9	9	10	9	6	10	8
Average	5.5	7.813	7.688	8.094	9.188	8.75	5	9.188	7.375
Median	4.75	7.75	7.75	8.5	9.25	9	4.5	10	7.5
Variance	3.071	1.567	2.281	1.249	0.710	0.429	1.423	1.424	0.482

Qualitative Data

We report on two major themes that emerged from the data: 1) Key steps in Optimization Solutions and 2) Reflection of Understanding. These themes were supported by all participants, and we will show support for the themes with excerpts from the faculty interviews.

Key Steps in an Optimization Solution. All participants talked about key step(s) while grading the optimization problems. While grading the optimization problems, one of the first things that most participants mentioned were checking the key steps of the solution that consisted of constructing the optimizing function, finding the derivative, solving for the critical points, and verifying which critical point solves the problem for the maximum or minimum. For example, F9 states while they are grading, "there are kind of key steps that I look for along the way" and

> "they're setting up the problem which typically has an optimizing function and a constraint. I don't like the word constraint, but because really it helps us to put the optimizing function into one variable. And then I want to see them take the derivative of the optimizing function, find the critical points, and then I want to see them determine whether or not those critical points really are the things we're looking for."

Therefore, checking key steps served as signposts to quickly assess if the solution is correct.

Some participants went further by combining key steps into chunks. They would look at the chunks as important pieces and would determine the number of points they designated to each chunk depending on their views on the importance of the chunk. For example, F10 comments that *"I solve the problem, show all the steps, and then I just go through and block off each little piece of what they do, and decide… How much does knowing this part account for? How much does knowing this part account for? So, I kind of block it off into pieces*" and F7 also states *"in an optimization problem they set up the function, right? That's one batch of stuff. They take the derivative and find the critical numbers. That's one batch of stuff. They test the critical numbers in the first or second derivative test. That's a batch of stuff, and then they answer the problem. That's a batch of stuff."* This ensures the appropriate allocation of points for important work in the student solutions and assists them when grading. However, not all participants grouped key steps in chunks and just developed their rubrics around the key steps. This is illustrated by F8 with *"[I] break my solution into steps depending on how important it is, and I usually have in mind that this step they should get 2 points. And this step has these points."*

At that point and to varying degrees, the majority would assess the details of the solution to further assess the student solutions and determine how many points they earn on key steps as well as overall. It is during this stage that participants expected the solutions to provide more detail to justify the work (see the next section of reflection of understanding).

Therefore, key steps and key chunks played an important role for participants when building their grading rubrics and while they were grading student solutions for the optimization problems. This ensured each participant graded consistently and fairly on each of the three solutions for the three optimization problems. Although this result was in the context of optimization problems, participants discussed this is how they grade on most topics.

Reflection of Understanding. We now use Skemp's (1976) relational and instrumental understanding to further explore how participants assess whether students understand the mathematics for the optimization solutions. We first discuss participants assessment of instrumental understanding and then move to the assessment of relational understanding.

Reflection of Instrumental Understanding. When analyzing the participant's optimization grading, we found instances where participants assessed the solution for instrumental understanding by not expecting any justification in the solution. They were assessing the solution based on formulas, rules, or facts, without displaying any further underlying understanding.

For example, all participants did not need to see justification for the calculation of the derivative for at least the optimizing function on these three solutions. For example, when F9 is grading the rectangle optimization problem, she states as she reads the solution*, "And then we're taking the derivative, which I think is good." For another example,* F10 states *"I think it can be common for students to just write something down because they know it's supposed to be a minimum, and so they just draw the picture like they know it's supposed to be without doing any actual thinking at all."*

Reflection of Relational Understanding. When analyzing the participant's optimization problem grading, we found many instances where participants assessed specific steps of the solution for relational understanding by expecting justification in that step. That is, they were looking for a reflection of relational understanding in a step that showed a deeper understanding of the rationale and underlying reasons for the mathematics. One prime example of the reflection of relational understanding occurred on the verification key step for the maximum or the minimum. For example, on S2 for the rectangle problem (see Figure 2), F7 states

> "… let's call this the next batch of stuff [boxes verification work in blue] which I think is the most important batch. I still would not call that enough work. If you want it to be a minimum, you know you want it to go down and then up. So, you're not showing me that you actually checked that. You're just showing me that you know what it's supposed to be so, I would imagine that there would need to be like, in my opinion, what I would call enough work, would be put a number over here [draws circles on both sides of sign chart].Put a number over here, you know, like 10 or one, or whatever, and then like 1,500, or whatever over here, and show me that you plugged it back into the first derivative and got this decreasing increasing. So, I would say, show more work on how you got this result."

For another example, F12 wants the hypothetical student to explain why there is a maximum or minimum at the critical point instead of just saying the $V''(r) < 0$, "My concern when there's just so little is that they're just regurgitating a fact. Like, 'Oh, it's less than 0, so that means it's a maximum.' Like, well, do you actually understand why it's a maximum." Therefore, F7, as other participants, expected the solution to display justification for key steps in the solution to show a reflection of relational understanding.

Discussion

In this paper we investigated how faculty assess student understanding on three optimization problems and found three main results. The first result was that for the most part, professors assigned points consistently, except on the rectangle problem. The second result was that professors wanted to see key steps in the solution. When a key step or steps were missing from a solution or there was insufficient work, professors would assess a lack of understanding and deduct points. This parallels the results of Raman (2003) for mathematical proofs where she found that mathematicians valued key ideas in a proof. The key idea is the big picture of why the proof works and can be used to construct the final written proof. Likewise, key steps give an outline or big picture of what needs to be produced to solve an optimization problem.

The third result was that professors wanted to see a reflection of understanding in the student solutions. They interpreted that students did not have a complete understanding when the solution did not include key step(s) or there were errors in between key steps. Prior research in proofs by Miller (2018) and Moore (2016) found a similar result when professors graded mathematical proofs. In contrast, Henderson et al. (2004) found that this was not the case with physics graduate students. This reflection of understanding further evolved into the final result where professors were interested in seeing that students were fluent with the mathematical rules (instrumental understanding) but also wanted to see students providing justification to show they understood the mathematical concepts involved in their solutions (relational understanding).

References

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. Assessment in Education: principles, policy & practice, 5(1), 7-74. https://doi.org/10.1080/0969595980050102

Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. Educational Assessment, Evaluation and Accountability (formerly: Journal of personnel evaluation in education), 21, 5-31. https://doi.org/10.1007/s11092-008-9068-5

Corbin, J., & Strauss, A. (2014). Basics of qualitative research: Techniques and procedures for developing grounded theory. Sage publications. https://doi.org/10.4135/9781452230153

Docktor, J., Dornfeld, J., Frodermann, E., Heller, K., Hsu, L., Jackson, K. A., Mason, A., Ryan, Q. X., & Yang, J. (2016). Assessing student written problem solutions: A problem-solving rubric with application to introductory physics. Physical Review Physics Education Research, 12, 1-18. https://doi.org/10.1103/PhysRevPhysEducRes.12.010130

Hattie, J., & Timperley, H. (2007). The power of feedback. Review of educational research, 77(1), 81-112. https://doi.org/10.3102/003465430298487

Henderson, C., Yerushalmi, E., Kuo, V., Heller, P., & Heller, K. (2004). Grading student problem solutions: The challenge of sending a consistent message. American Journal of Physics, 72, 164-169. https://doi.org/10.1119/1.1634963

Marshman, E., Sayer, R., Henderson, C., & Singh, C. (2017). Contrasting grading approaches in introductory physics and quantum mechanics: The case of graduate teaching assistants. Physical Review Physics Education Research, 13(1), 010120. https://doi.org/10.1103/PhysRevPhysEducRes.13.010120

Marshman, E., Sayer, R., Henderson, C., Yerushalmi, E., & Singh, C. (2018). The challenges of changing teaching assistants' grading practices: Requiring students to show evidence of understanding. Canadian Journal of Physics, 96(4), 420-437. https://doi.org/10.1139/cjp-2017-0030

Marshman, E., Maries, A., Sayer, R., Henderson, C., Yerushalmi, E., & Chandralekha, S. (2020). Physics postgraduate teaching assistants' grading approaches: conflicting goals and practices. European Journal of Physics, 41, 1-27. https://doi.org/https://doi.org/10.1088/1361-6404/ab9890

Miller, D., Engelke Infante, N., & Weber, K. (20187). How mathematicians assign points to student proofs. Journal of Mathematical Behavior. https://doi.org/http://dx.doi.org/10.1016/j.jmathb.2017.03.002

Moore, R. (2016). Mathematics Professors' Evaluation of Students' Proofs: A Complex Teaching Practice. International Journal of Research in Undergraduate Mathematics Education, 2, 246-278. https://doi.org/10.1007/s40753-016-0029-y

Mulliner, E., & Tucker, M. (2015). Feedback on feedback practice: Perceptions of students and academics. Assessment & Evaluation in Higher Education, 1-23. https://doi:10.1080/02602938.2015.1103365

Raman, M. (2003). Key ideas: What are they and how can they help us understand how people view proof?. Educational studies in mathematics, 52(3), 319-325. https://doi.org/10.1023/A:1024360204239

Skemp, R. R. (1976). Relational understanding and instrumental understanding. Mathematics teaching, 77(1), 20-26. https://doi.org/10.1080/00255807600100070

Stiggins, R., & Chappuis, J. (2005). Using student-involved classroom assessment to close achievement gaps. Theory Into Practice, 44(1), 11–18. https://dx.doi.org/10.1207/s15430421tip4401_3,

Wiliam, D. (2006). Formative assessment: Getting the focus right. Educational Assessment, 11(3/4), 283–289. https://doi.org/10.1207/s15326977ea1103&4_7
Yerushalmi, E., Marshman, E., Maries, A., Henderson, C. R., & Singh, C. (2015). Grading practices and considerations of graduate students at the beginning of their teaching assignment. 2014 Physics Education Research Conference Proceedings, 287–290. https://doi.org/10.1119/perc.2014.pr.068