Pass-Fail Grading of Technical Writing in a Material Science Laboratory

Prof. John R. Rogers, Benedictine College

Dr. John Rogers is a member of the Benedictine College School of Engineering faculty where he teaches courses in mechanical engineering including materials engineering laboratory, mechatronics, vibrations, and design.

He earned a doctorate in mechanical engineering from Rensselaer Polytechnic Institute in 2003. He has a range of engineering and teaching experience. As an ocean engineer, he developed towed systems for underwater listening, and structures that reduce ship noise. As a structural engineer, he verified the strength of foundations for shipboard equipment. As a design engineer he developed fluid power systems and designed industrial controls for aircraft and automobile plants. As a college instructor, he taught a variety of courses in design, mechanics, and mechatronics, and he advised numerous student projects and independent studies.

Dr. Rogers' PhD research work at Rensselaer was the optimization of stepper controls to reduce unwanted dynamics in machinery. He was awarded a patent for this work. His scholarly interests include microcontroller-based system design, biomechanics, instrumentation, and teaching and learning. He holds a Professional Engineer license.

Taylor Goring
Joel Michael Iwanski

Pass-Fail Grading of Technical Writing in a Material Science Laboratory

Abstract

Grading requires considerable effort, and often is slow due to the time it takes to judge the quality of student work. The question studied was: does low-resolution grading reduce the time it takes to grade written work while maintaining fairness? A new grading scheme was developed to address these challenges and it was implemented in an undergraduate laboratory course in material science. The goal was to reduce the time spent grading while maintaining consistent grading. The course is taught to juniors in a mechanical engineering program; enrollment is 30 students. One of the course objectives is to help students become proficient in technical writing, and so most graded events are technical memos and reports. Student graders were employed to grade student submissions using the new scheme. Eight aspects of technical writing were emphasized in the course: grammatically correct, proper format, clear, precise word choice, concise, technically sound, thorough, and insightful. Each of these aspects was graded on a pass-fail basis in the memos and reports that the students submitted. The use of pass-fail grading was introduced to help make it easier and quicker to judge the quality of student work. Students received feedback in the form of pass or fail in each aspect with written explanation from the grader. To be consistent in assigning grades, student graders and the instructor developed a rubric, agreed on definitions in each aspect, and established an understanding of what constitutes passing, and what constitutes failing. The student graders and the instructor calibrated passing thresholds by iteratively grading, discussing, and regrading a subset of student submissions until grades were consistent. It was found that, once the decision criteria were commonly understood, grading was faster because it is easier to consider one aspect at a time, judge the work, and make a binary decision: pass or fail. The turnaround time was noted for six written assignments in two successive semester offerings, one with the traditional grading, and one with the new scheme. The result is faster turnaround with the new pass-fail scheme. Additionally, the instructor and student graders subjectively note improvements in writing quality over the course of the semester.

Introduction

Grading is an essential duty for college instructors because it is an important means of providing feedback to students about their learning. Feedback from the instructor (or the grader) to the student often comprises a grade and written comments or corrections. Ideally, students will use the feedback to learn, and to improve future efforts.

Grading can be a challenge for instructors in cases where there are large enrollments, or the work

to be graded is complex or lengthy. In these cases it can take considerable time for the grader to process the work and provide feedback. Lengthy feedback times are suboptimal from a learning perspective since the student may miss opportunities to learn from the feedback. Faster feedback results in better learning because the feedback has better connection to the work when the memory of the work is fresh.

One way to reduce grading time is to employ low-resolution grading, that is, grading methods that use low numbers of possible grade levels. Grading on a scale of 100% without fractional percentage points has 100 levels. Grading on an A-B-C-D-F scale without pluses and minuses has five levels. Miguel and Larson[1] recommend using the lowest number of grading levels that allows an accurate assessment of student learning, and they state that grading is generally faster with fewer levels. Hjelmstad and Baisley[2] completely restructured mechanics courses, implementing mastery-based assessment. An important feature of their work was the use low resolution assessment for work that is graded by graduate teaching assistants; this makes it easy for the teaching assistants to make rapidly judge the work and provide timely feedback. They note good consistency among multiple graders.

This study explores the question of whether low-resolution grading reduces the time it takes to grade written work while maintaining fairness. It was conducted at the engineering program of a small Midwestern liberal arts college. The study concerned the method of grading in a materials laboratory course. The course contributes to the college's written communication goal, and so most of the assessments are written works: technical reports and technical memoranda. The course is typically taken in the junior year of the mechanical engineering degree. Lectures on technical writing are included to reinforce written communication skill. Students in the course will have had a physics course that requires written lab reports, and an English composition course.

Methods

Turnaround time was used in this study as a direct measure of how rapidly feedback was provided to students. For this study, turnaround time was defined as the average number of days between an assignment due date and the date the grade was entered in the learning management system. Turnaround time was recorded for six assignments: two technical reports and four technical memoranda. Data was collected over two semesters, one semester with traditional grading, and one with the experimental method.

A key part of the method was the creation of a clear, easy-to-use rubric. A rubric typically has a number of elements to evaluate, and each element will have a number of discrete levels. Determining the number of rubric elements and the number of levels in each element is a balance between thorough assessment and quick assessment. More elements allow the grader to more thoroughly consider different aspects of the work, but it takes longer. The same can be said of the number of levels in each element. Grading generally takes less time when there are fewer elements to assess, but this may not provide sufficient feedback.

An eight-element rubric was developed to balance thoroughness and speed. The original rubric was developed by the instructor, and then revised by the student graders and the instructor as a group. Each rubric element has two levels: pass or fail. Examples of passing and failing in each

rubric element were provided to the students. The elements of the rubric correspond to eight aspects of technical writing.

The eight elements of the rubric are:

- grammatically correct
- proper format
- clear
- precise word choice
- concise
- technically sound
- thorough
- insightful

These aspects are often cited in the literature and in technical writing books,[3][4][5][6][7][8] and they were each selected based on their importance to the class learning objectives. The rubric was provided to the students with commentary and examples to communicate what constitutes passing and what constitutes failing in each aspect of technical writing. On every student submission, the graders evaluated the eight rubric elements on a pass-fail basis.

The graders were undergraduate students who had previously taken the class. They were selected due to their writing proficiency during their time in the class. The graders and the instructor spent time on calibration to develop grader-to-grader consistency. Nine out of 36 submissions for the first assignment were randomly selected for inclusion in a calibration set. All the graders and the instructor independently assigned tentative grades using the rubric for all the submissions in the calibration set. All discrepancies in tentative grades were discussed as a group. This cycle was repeated on the same calibration set until grades were consistent. The cycle of grading and discussing discrepancies was repeated three times before inter-grader reliability was judged to be adequate. Grader-to-grader consistency is high when the rubric elements are well defined. After the first iteration only, the team revised the rubric and added examples for clarification.

The low resolution pass-fail structure was used (rather than structures with more levels) to test whether it could improve grading turnaround time.

The eight elements of the rubric are described below.

Grammatically Correct

The grammar criterion evaluated the word and sentence structures within the paper. Misplaced commas, improper capitalization, improper subject-verb agreement, inappropriate tense and inappropriate use of punctuation constituted minor errors in this category. Major errors included consistently inappropriate verb tenses, run-on sentences, and sentence fragments. One major error resulted in a fail for the paper, while numerous minor errors or a wide variety of minor errors gave a fail. An example of a failing grade in the grammar rubric due to a sentence fragment can be seen below.

“It would also be helpful to take more replicates at each level, especially at the larger angles. Where the data stops following the trend of decreasing max strain.”

Proper Format

The format criterion covered the visual appearance of elements in the paper. Minor formatting issues consisted of inconsistent font, inappropriate white space or margins, incorrectly formatted equations, missing page number, and small mistakes in table and figure presentation. Major errors included missing section headers, figure captions, or table titles, and poor resolution figures.

Clear

The clarity criterion determined how clear, organized, and easy-to-read the paper was. Many of the minor errors were determined by the graders’ judgment when it was difficult to follow the author’s train of thought. Major errors included content that was placed in the wrong section and sentences or phrases that do not follow what preceded them. A failing example in clarity is provided below. In the excerpt, the clause “which can either increase. . . ” does not follow from the preceding clause. A passing example would specify that the carbon fiber strength depends on the orientation of the weaves relative to tension forces present.

“Since carbon fiber is a weaved cloth, there are multiple orientations that the strands can lay in a piece which can either increase or decrease the parts strength.”

Precise Word Choice

The precise word choice criterion evaluated a consistent use of words which the reader understands to mean precisely what the author intends. It also involved the use of appropriate engineering and technical terms, especially those which related to the experiment. Failing grades were given when there was more than one instance of ambiguous wording or the use of the wrong word. The example below contains a minor error of precise word choice due to the use of the word “deviance” instead of the proper term “deviation.” Any subsequent errors would result in a failed grade.

“The smaller the deviance in degree from the tensile forces the higher the stress forces were resisted until a brittle fracture.”

Concise

The concision criterion determined whether the author used a reasonable minimum of words to communicate clearly. In this case, the distinctions between minor and major errors were made by the graders’ judgments based on the gravity and frequency of the mistakes. Errors against concision included unnecessarily wordy writing, inclusion of irrelevant content, the inclusion of content that the reader is expected to know, and the repeated use of multi-word phrases that could be expressed in fewer words. The example below displays a minor error, since the reader could assume that data was transferred to the computer in a proper manner. Subsequent offenses would lead to a failed grade.

“A USB flash drive saves the exported Excel file of the data from the machine.”

Technically Sound

The technically sound criterion analyzed whether the author had a sufficiently deep technical understanding of the topic. A paper that was technically sound would have no technical or engineering errors. All claims and discussion would be supported by engineering fundamentals and evidence. Notably, a false conclusion that follows logically from the results and discussion of the paper was not a cause for failure. A minor error against technically sound included inappropriate numbers of digits reported and a mismatch between citations and reference numbers. Major infractions included significant engineering errors, plot axes not matching the data, missing equations, missing analyses, or unjustified claims of success.

Thorough

The thorough criterion evaluated whether the paper addressed all of the requested material in experiment handouts. It also covered whether all of the necessary sections of the paper were present, and whether the conclusion logically followed the content of the paper. Failing thresholds included missing the sections outlined in the memo and report specifications, missing references, missing figures, missing assignment deliverables, and insufficient conclusions. For example, the excerpt below displays a sample conclusion in a memo. Although the information provided in the quotation is useful, it contains no mention of the results, discussion, or purpose of the experiment. Therefore, the conclusion is insufficient and the paper will fail the thorough category.

"In order to get more consistent results, the jaw spikes should be sharpened or replaced. Multiple groups tightened the jaws as much as they were physically capable and yet the test pieces were slowly coming out of the jaws. Preventing the test piece from slipping will reduce the displacement distance recorded."

Insightful

The insightful criterion evaluated the author's original thoughts or conclusion based on the evidence. An insightful paper could provide an explanation of anomalies, evaluate nuances within the results, or discuss observations beyond the basic treatment of the subject devised in the experiment handout. Failing grades were given to papers with a marginally adequate treatment of the subject.

Rubric Weighting and Clarifications

A spreadsheet was used for grading, see Figure 1. The spreadsheet is copied for each submission that is graded. "P" or "F" (for *pass* or *fail*) is entered on each row. The spreadsheet uses the weights to calculate the grade. Graders comments can optionally be entered directly on each row of in the spreadsheet. This spreadsheet was sent along with marked up work to the student for feedback.

In the scoring of reports, the *technically sound* and *thorough* categories were each weighted as 20% of the grade, while the other categories were weighted 10%. *Technically sound* was weighted higher due to the importance for the students to conduct their experiments safely and correctly. Thorough was weighted higher due to the importance in technical writing of having a whole, well-rounded paper which addresses all of the main concerns of a paper, such as the abstract, background, methods, results, discussion, conclusion, and references. It is important to note that 80% of the grade was based on technical writing; only the *technically sound* category evaluated the engineering competence.

StudentName			AssignmentTitle	
scoreMatrixPassFail.xls	Revised:	16-Dec-22		
	Weight	Pass or Fail	Score	Out of
Grammatically Correct	10%	p	10	10
Proper Format	10%	p	10	10
Clear	10%	p	10	10
Precise Word Choice	10%	f	0	10
Concise	10%	p	10	10
Technically Sound , Accurate	20%	f	0	20
Thorough	20%	p	20	20
Insightful	10%	p	10	10
		Points (rounded)	totalPoints	%
	GRADE	70	100	70%

Figure 1: A scoring spreadsheet was used to calculate a grade based on pass or fail and the weight in each category.

Results

Turnaround time is shown in Table 1. Grading of three out of four memos took marginally longer with the new pass-fail grading method. Turnaround time on one of the memos and both reports were substantially improved. The average turnaround time for memos improved by 3.5 days with the pass-fail grading. The average turnaround time for reports was improved by 20.5 days.

Table 1: Turnaround Time in days; traditional vs. pass-fail grading

Assignment	Traditional	Pass-Fail	Improvement
Memo 1 (Individual)	7	9	-2
Memo 2 (Individual)	15	19	-4
Memo 3 (Group)	26	29	-3
Memo 4 (Group)	37	14	23
Report 1 (Individual)	40	21	19
Report 2 (Group)	37	15	22

Discussion

Creating a Pass-Fail system ultimately led to many benefits for the instructor, graders, and students. In this study, the pass-fail system significantly decreased turnaround time on longer papers like technical reports, and marginally decreased turnaround time on shorter papers like technical memos. With a well-defined rubric, a pass-fail system decreases turnaround time for grade delivery. The system also encourages consistency between all graders and the instructor, so that the instructor does not have to inspect the work of graders as often.

The three memos that took longer to grade with the pass-fail system were the first assignments graded. The longer turnaround is attributed to grading these three memos while the team was

revising the rubric and calibrating for grader-to-grader consistency. It is expected that future semesters will be improved in this regard since the rubric will be ready at the start of the semester.

Creating an efficient pass-fail system can take more time up front with creating a rubric and having a calibration period, but it helps ease workload later in a semester. Furthermore, if it is used for a class that is repeated every year, the same rubric can be used, further decreasing an instructor's prep time.

Having a detailed rubric that included examples for each category greatly improved students' writing and set clear expectations for the assignments. The instructor and graders spent considerable effort on calibration. The first iteration of calibration resulted in significant discrepancies between graders, and it was determined that the first draft of the rubric was inadequate. The rubric was refined for the benefit of students and graders. The revised rubric had more examples of cases that were marginal: just barely passing or just barely failing. The more clear definitions and examples in the revised rubric helped to reduce overlap between categories. Graders were discouraged from failing a paper in multiple elements for single offences. Where there was overlap during grading, graders were encouraged to use their judgement to pick the most appropriate element to apply a penalty.

Compared to grading for other classes, a pass-fail system helps reduce the judgement time required of graders. When a grader is given a specific set of solutions to check work against, the grading is quick. When a grader decides on a score between 1 and 100, having to constantly use their own judgement adds significant time to grading. Using a pass-fail system for grading papers helps reduce judgement time, since the rubric specifies what mistakes fall into which category, and how many mistakes constitutes a pass or fail. A pass-fail system is a good solution for classes that cannot use an exact solutions sheet, but still need quick and equitable grading.

Limitations

Throughout the semester, certain limitations of the pass-fail grading system became apparent. This study was conducted over only two semesters, and therefore the data is limited. Running a study over a longer period would be helpful. The second semester began with a draft rubric but it became clear during calibration that the rubric needed revision. This caused the calibration period to be longer than expected. The revised rubric went out to students in week 4 of the semester, in time for memo 3.

For the rubric to remain clear for students, separate rubrics should have been made between a technical memo and a report. It was difficult to quantify the acceptable number of mistakes when the lengths of different assignments varied so much. Technical memorandums, two to three pages in length, were the first assignments submitted, and grading calibrations were based on these. However, it seemed that a modified standard was needed for the technical reports, which were ten or more pages in length. For instance, a single run-on sentence in an otherwise well-written report seemed more forgivable to the graders than in a shorter, more concise memorandum. Therefore, on categories such as grammar, format, and precise word choice, the graders agreed that reports would not receive failing marks without two or three of the errors, unless they were egregious mistakes.

Additionally, the nature of a pass-fail rubric resulted in some graders feeling that certain students received higher grades than they deserved, while others received lower grades than they deserved. This was because a paper might not be good overall, but if they only had one mistake in each defined category, the student probably would not fail that section. Other students had overall very good content, but their grammar mistakes resulted in low scores. One could argue, however, that this is the exact purpose of a rubric - to ensure that a student will not be failed overall for many mistakes in a single category.

A final limitation is that the instructor and graders felt that grading was fair across the whole class, but this was not measured.

Areas for Improvement

For continued use of the pass-fail grading system, there are a few areas with room for improvement. Creating separate rubrics to differentiate technical memos from full-length reports would help clarify what constitutes a pass or fail in each category. Since the issue of "in-between" grades was most prominent with longer technical reports, some of the rubric categories could be broken down further. Experimentation would have to be done to see if creating more rubric categories increases grading time by a considerable amount.

The students were not surveyed about their perceptions of the system, and none of them mentioned it in course end evaluations. It would be interesting to find out how students feel about this method of grading: whether they felt it was fair, and whether they felt it helped them improve their technical writing.

Another possible improvement to the pass-fail system could be in the rubric itself. It could be helpful to have a clearer definition of how many offenses constitute a pass or fail in each category. This could also be made clearer by including more examples of good and bad writing.

The calibration exercise was useful to ensure that graders and instructor could assign consistent grades. The team simply used their judgement to say when inter-grader consistency was adequate, but it would have been helpful to document the convergence of standard deviation of grades among graders over the multiple iterations.

The focus of this study was the reduction in grading time to provide rapid feedback, but the the graders and the instructor noted that the introduction of the detailed rubric was followed by a noticeable improvement in writing quality. Studying this improvement would be of interest.

References

[1] Eric Larson and Agnieszka Miguel. Efficient and effective grading of student work. In *2006 Annual Conference & Exposition*, number 10.18260/1-2–674, Chicago, Illinois, June 2006. ASEE Conferences. https://peer.asee.org/674.

[2] K. D. Hjelmstad and A. Baisley. A novel approach to mastery-based assessment in sophomore-level mechanics

courses. In *2020 ASEE Virtual Annual Conference Content Access*, number 10.18260/1-2–34028, Virtual On line, June 2020. ASEE Conferences. https://peer.asee.org/34028.

[3] Gold R.M., Reilly A., R. Silberman, and Lehr R. Academic achievement declines under pass-fail grading. *The Journal of Experimental Education*, 39(3):17–21, 1971. doi: 10.1080/00220973.1971.11011260.

[4] Sorcinelli M. D. and Elbow P. (editors). *Grading Student Writing*, pages 127–140. Josey-Bass, Hobojken, NJ, 1997.

[5] L. Finklestein. *Pocket book of technical writing for engineers and scientists*. McGraw-Hill's BEST. McGraw-Hill, 2000. ISBN 0072468491.

[6] D. Beer and D. McMurrey. *A Guide to Writing as an Engineer*. john Wiley and sons, 2009. ISBN 978-0-470-41701-0.

[7] D.F. Beer. *Writing and Speaking in the Technology Professions: A Practical Guide*. Wiley Online Library. Wiley-Interscience, 2003. ISBN 9781119134633.

[8] J.E. Aaron. *The Little, Brown Handbook*. Pearson Longman, 2015.