
Research Paper Evaluation Forms: A Better Mousetrap?

L. Michael Couvillion

The chore of evaluating research (or term) papers and providing meaningful feedback to student authors while minimizing faculty burnout is particularly difficult with large class sizes. Over the years, I have typically resorted to handling this problem by simply writing a letter grade, or its numerical equivalent, together with a one-sentence "executive summary" of my thoughts, on the cover sheet. I have identified specific problems in the text with a liberal use of red ink and question marks.

Problems

Many problems have been associated with this approach, however. Some of the more serious ones include:

1. Assuring consistency in marking papers among different students in the same class, as well as the same students in different years. It becomes difficult to explain why the paper of a student who received a "B+" is inferior to that of a roommate who received an "A."

student must choose a paper topic which I then approve.

It quickly became apparent that students were not only curious about this new grading strategy, but also perplexed. While their initial reaction was not unmitigated joy, most seemed intrigued by the novelty of this grading scheme and willing to give it a chance. To address their concerns, and to turn anxiety into a positive rather than negative motivator, I devoted two hours of class time to planning the paper.

The first hour was spent going over the form and answering questions. As a stimulus to class discussion, I learned that just distributing the form elicited more questions and comments than I have ever received. Some sample questions were:

- What's the difference between footnotes and bibliographies?
- Are we penalized if the spelling checker doesn't catch spelling mistakes?
- Do we have to have graphs and charts?
- What do I do if the paper is too short?
- What do I do if the paper is too long?
- Why do we have to use headings?
- I don't understand how to outline a paper: Can you give us some good examples?
- Is the *Wall Street Journal* a journal for grading purposes?
- If we are supposed to use scholarly journal articles, how do we find them, given the available library resources?

The second hour was devoted to specific research procedures that represented problem areas for student authors. I worked with the class to completely outline a hypothetical research paper on Social Security and wrote the headings on the board so students could see the logical flow of ideas. Of course, I subsequently received six papers on the subject, all amazingly alike in organization! I explained the use of the New Palgrave Dictionary of Economics and copied a sample subject heading on Human

Capital so that they could see how valuable such a resource could be. But the heart of this discussion centered on the proper way to most efficiently use the many indexing tools available in the Journal of Economic Literature. Most students had no idea that such a resource was even available in the library.

Results

This form was given to the students again in November, during the student evaluation period, for their input and to remind them of the writing task ahead of them. In December, I actually used this form to grade all papers. The results are given in Table 1 below:

Table 1	
Term Paper Evaluation Scores	
Raw Scores	Statistics
97	High Score
94	
93	Class Size = 25
90	
89-4	Class Average = 80.11
87	
85	Standard Deviation = 9.87
A 84	
80-2	Coefficient of Variation = 12%
77-3	
76-3	Curve Added = +7 points
75	
B 74	
72	
71	
C 65	
61	
D 60	Low Score

As you can see, I converted the numerical scores into their quality point equivalents or letter grades by adding seven points to each student's raw score. Why seven points? The mean score on the summary item (Overall

Impression) was 86.8 while the class average was 80.11. I assume that my overall subjective impression should set the curve for the class as a whole. In other words, the average paper grade should have been a "B+" and turned out to be a "B-."

Outcomes

At the end of this course, I gave the students the form once again (after returning their papers to them) and asked for comment. Table 2 presents representative student observations. Some of the feedback was negative, since I asked for it and since frustration levels were peaking as each student saw the consequences of his or her actions. For example, most suggested reassigning weights away from areas where they lost points to areas where they achieved perfection.

Table 2
Student/Author Observations

- Don't change anything.
- Topic should be approved first; therefore, it is a moot point to grade it as an afterthought.
- I think overall impression should be the major basis on grading.
I feel some topics don't have much graphics.
- I like the grading system.
- I feel that since this is an economic course not English 101 you should grade the paper on topical area, overall impression and items within that realm. Having such a large breakdown is overwhelming for the student.
- Can the whole thing—JUST GRADE IT!
- I think it's a fair distribution.
- I think the overall impression is what counts. The others should be considered in the overall impression, but the overall impression is what counts.

A second outcome was my discovery of a better way to curve scores. In the past, I used some arbitrary method (such as adding the difference between the class average and 80, assuming that a "B-" was average for a junior/senior level class) which produced somewhat arbitrary results. Now, I use the last item on the evaluation form to gain an "overall impression" of the students' written work. The average of that line becomes my standard for setting the grading curve for the class. I am convinced

this “global” assessment is a more appropriate gauge of the worth of the stack of papers on my desk than any artificial standard would be.

Perhaps the most significant outcome of all, however, was the satisfaction of the students with their grade. Only one student out of 25 challenged his grade. I added 5 points to his score when he pointed out to me that his first paragraph was meant to be an abstract. I pointed out to him that without proper headings, it is difficult for the reader to find items of interest. For perhaps the first time in his college career, this student learned the value of outlining and topical subdivisions.

Advantages

The beauty of this strategy for evaluating research papers is its flexibility: each instructor can (and, I feel strongly, *should*) modify this format to meet his/her evaluative needs. Categories can be combined, items which are irrelevant for the course can be omitted, and pet peeves can be addressed before it is too late.

One additional, and unexpected, advantage of this exercise was the unique opportunity to set priorities and clarify expectations. Is organization really more important than appendices? Do titles matter, and if so how much? How does writing style affect the entire paper’s evaluation? Can a hopelessly flawed paper be detected by this system, and how? Is top-quality work sufficiently valued?

My end product was a set of papers which were returned to the student authors with more feedback of value to them. They now at least understand why their work was evaluated as it was, and they are capable of independently conducting an academic autopsy to find out what went wrong and, equally importantly, what went right. Much to my surprise, I found that I was actually able to plow through the 400-600 pages of text, right before final exams, spending (investing?) *less* time in grading than with my old system.

Conclusions

Table 3 contains the descriptive statistics for each individual grading item. Note the difficulty index variation. My students had the most trouble handling appendices (59) and the abstract (65) last semester. Not surprisingly, these two items had the highest coefficient of variation. Why? I was effective in communicating their importance to some, but by no means all, of the authors. I am encouraged by other paper dimensions, though, since 94% of the class footnoted perfectly and 87% provided impressive coverage of the topic.

Item	Weight	Mean	Standard Deviation	Coefficient of Variation	Difficulty* Index
Title	5	4.24	0.70	16%	85
Abstract	5	3.24	1.96	61	65
Headings	5	3.84	1.45	38	77
Organization	10	8.60	1.15	13	86
Exhibits	10	7.68	1.68	22	77
Spelling	5	3.80	1.22	32	76
Grammar	5	3.56	1.36	38	71
Topic	10	8.64	1.47	17	86
Coverage	10	8.67	1.53	18	87
Footnotes	5	4.72	0.66	14	94
Bibliography	5	4.12	0.82	20	82
Appendices	5	2.96	1.88	64	59
Writing Style	10	8.28	1.23	15	83
Overall Impression	10	8.68	0.90	10	87

*Difficulty Index = Mean/Weight

I am using this form, with minor editing, this semester in my Development of Economic Thought course. This time, I will be able to complete a cross-sectional multiple regression analysis of the form's data with the overall score as the dependent variable and each item score as independent variables. Based on this analysis of which factors really determine the final score, I will be in a position to make a wholesale revision and condensation of the form for future course grading.

This tentative evidence from actual classroom use of the Research Paper Evaluation Form means to me that the strategy is a successful one. I encourage other faculty to modify the form to best meet the needs of each different course.

L. Michael Couvillion, an associate professor of Economics, has taught a variety of upper-level elective courses in the Business Department for ten years. Recently he has worked as a consultant to Houghton-Mifflin, writing examination questions for economics textbooks.