

# "Could You Explain My Grade?" The Pedagogical and Administrative Virtues of Grading Sheets

Sally Roever, University of California, Berkeley  
Paul Manna, College of William and Mary

The scenario is as unpleasant for the teacher as it is for the student. After spending countless hours trudging through a stack of mind-numbingly repetitive bluebooks, the teacher has finally returned graded midterm exams to her class. With grading safely in the past, she gleefully returns to other teaching, publishing, and personal responsibilities. At least she tries to—until a concerned student appears at the office door asking her to explain the points she assigned to a given answer. She flips to the relevant page of the bluebook and discovers no comments in the margin. The teacher is now on the spot to rack her brain for the reasoning that justified the score she assigned several days or even weeks before and explain to the student clearly and accurately why he did not receive full credit. The student is in a likewise awkward position, wondering how much thought went into the grade the first time around.

Few conscientious teachers would insist that numerical grades on paper assignments or in-class exams are infallible measurements of writing quality. Nonetheless, awkward confrontations between graders and students can be minimized with the preparation of a useful classroom tool: a grading sheet. Grading sheets are one- or two-page rubrics that list a carefully developed set of criteria for answers, match numerical scores

---

*Sally Roever is a Ph.D. candidate in political science at the University of California, Berkeley and former visiting assistant professor of government at the College of William and Mary. Her research and teaching interests focus on comparative politics, research methods, and the politics of the informal sector in Latin America.*

*Paul Manna is assistant professor of government at the College of William and Mary where he also teaches in the Thomas Jefferson Program in Public Policy. His research and teaching interests center on American politics and public policy, in particular the politics of federalism and K-12 education in the United States.*

with qualitative descriptions of answer quality across the range of potential scores, and provide space for brief comments linked to those descriptions. Instructors can use grading sheets to evaluate in-class exams, take-home assignments, and longer term papers. Though disputes over scores are probably inevitable, we believe that a thoughtfully conceived, carefully used grading sheet can help avoid such uncomfortable situations. In our own teaching we have found that grading sheets improve grading accuracy and consistency, provide helpful feedback to students who genuinely want to improve their work, and reduce the amount of time needed to evaluate a set of exams or papers.

## The Challenges of High-Volume Grading

Teachers with experience in large classrooms with several teaching assistants or smaller courses where the professor does all the grading widely acknowledge the challenges inherent in evaluating assignments that entail a substantial amount of student writing.<sup>1</sup> First, assigning accurate scores is not always as straightforward as we would hope. Grades are attempts to measure an often subjective concept called a "good answer." The quality of a paper or exam answer can vary widely on a number of different dimensions, such as accuracy, specificity, structure, use of course themes, reference to readings, and clarity of prose, to name just a few. Arriving at "the right" score on an essay even after carefully weighing these various dimensions can be difficult, and the pressure to return assignments in a timely fashion makes accurate grading all the more challenging.

Second, grading consistently across students, across answers, and across teaching assistants in large classes requires heroic effort and is probably rarely achieved. The higher the volume of essays to grade, the more onerous the task, since grading large numbers of

exams or papers can be both boring and exhausting. Greater fatigue makes it more difficult to differentiate between answers, but spreading the grading out over several days or weeks to avoid weariness might cause the teacher's standards for good answers to shift over time (Chisholm 1990, 603).<sup>2</sup> Teaching assistants, who may enjoy working with students, are nonetheless under substantial pressure to keep their own research on track and thus face paltry incentives for providing accurate and complete feedback on student work. Professors fielding complaints from students about their TAs' grading, essentially a problem of inter-coder reliability, face unpalatable courses of action: some faculty members support their teaching assistants no matter what to maintain their TAs' credibility with students, while others easily cave to student complaints and bump the grade up a few points to usher the students out of their office more quickly.

Finally, the incentive structures in large universities rarely prioritize good teaching over good research (Chisholm 1990), and the reward for accurate grading is often nothing more than a lack of complaints from students. Yet, in all but a small number of colleges, grading is an inevitable component of the educational process in the United States. For the most part, students take grades very seriously; thus, as graders, we should treat students fairly and not simply as unfortunate victims of our own time constraints (Glenn 1998).

## A Potential Solution to the Grading Malaise: Grading Sheets

Grading sheets help both teachers and students because they identify clear criteria that will be used to evaluate student writing. We structure our grading sheets around two types of criteria. *General criteria* identify the key variables we will be looking for in the students' work, such as use of evidence. *Explicit criteria* describe qualitatively different answer possibilities under each

variable; for example, a student's use of evidence might range from "exceptional," which would earn 10 points, to "nonexistent," which would earn zero points. When handing out a writing assignment or announcing an exam we find it helpful to present students with the general criteria so they have a basic understanding of our expectations. Professors may, and often do, hold tacit ideas about what they expect their students to show on an exam or in a paper. (We often hope that students will make certain connections or seize on specific concepts from our course readings, for example). We believe it is unfair to hold students accountable for such unspoken criteria. Doing so encourages students to try to guess what is in the teacher's head so they can conform to some unspoken ideal. We find that students can develop more persuasive and well-informed pieces of writing, and improve their writing skills overall, when they work from an explicit framework of expectations.

Grading sheets serve several functions: they help teachers arrive at fair, accurate scores, they give students valuable feedback on their work, and, because of the presence of pre-typed explicit criteria, they allow instructors to avoid writing the same comments repeatedly. Teachers can use grading sheets for evaluating identification, short-answer, and essay questions on bluebook exams, and with longer term papers or essays.

Even though the use of grading sheets may not be common, many academic researchers employ a similar process when they systematically examine a text for content. In analyzing presidential rhetoric, for example, scholars often begin with general concepts in mind, such as the presence of domestic policy issues in major speeches. That general concept may include more explicit categories, such as education, welfare, health care, and Social Security. Good coding sheets (or computer programs that guide machine coding systems) break down the content in this way and then contain explicit decision rules for how to code particular nuggets of speech.<sup>3</sup> (Does a job training program count as a reference to education, or is education exclusively for K-12 schooling and college? Should Medicare fall under Social Security or health care?) The analogy between devising sets of criteria to code a presidential speech and a student essay is not perfect; however, both of these activities flow from common principles about how to read and systematically assess for content a large number of similar

texts (e.g., speeches, newspaper articles, or student exam answers).

There are many ways to devise a grading sheet. We have found the following approach quite useful for exams and paper assignments in our intro- and mid-level courses in comparative and American politics.<sup>4</sup> First, we begin with the general criteria we expect in the students' work. A recent essay assignment contained four such criteria: (1) clarity of the overall argument and responsiveness to the essay question; (2) use of class readings as evidence to develop the argument; (3) clarity of prose; and (4) grammar and spelling.

The second step is to assign weights to each of these general criteria. In the current example, the weighting followed the pattern described in Table 1. Of course, instructors who choose to emphasize the mechanics of writing could weight categories 3 and 4 more heavily, an approach that might be more useful in a freshman writing seminar, for example.

The third step is to elaborate each general criterion with a set of explicit criteria and corresponding point values. This step is essential for papers that assign points to compute a grade. In other words, if students are to understand why they received 26 rather than 30 points on their use of class readings, there should be some standard that distinguishes between these two scores. Table 2 outlines one possible approach. For the general criterion "use of readings as evidence," papers that scored a 30 made exceptional use of the course readings; those that scored a 26 were headed in the right direction, but suffered from one of several possible problems, including discussion that revealed a conceptual error (e.g., the paper made mistakes in discussing "divided government") or a lack of explanation to demonstrate understanding of key concepts (e.g., the discussion of divided government was not necessarily incorrect, but was too thin to show thorough understanding of the term).

Unless the instructor has used the grading sheet previously, refined it

through the years, and circulated it to students with the assignment, it is valuable to take a fourth step: testing a first cut of these explicit criteria on a few randomly selected student papers or exams. Consider this step akin to field testing a public opinion survey. In surveys, questions that appear clear to researchers may actually confuse respondents. Likewise, a set of explicit criteria for assigning points to an essay may seem clear in theory, but inadequate in practice. During this field test, it is important, however, not to delete or significantly alter the general criteria or their weights. Doing that would be unfair to students who used those criteria in crafting their essays (Damron 2003).

A field test also helps teachers gauge the clarity of their paper assignments or exam essay questions. If students' responses take several different tacks, while the teacher expected a set of papers or bluebooks that would address a more narrow range of topics, then there was likely a breakdown in communication. Rather than penalizing students for a gap in the professor's upfront explanation about the assignment or exam, the professor can alter the explicit criteria used to assign points.

A field test is also invaluable for training teaching assistants how to apply the criteria consistently. As in content coding for research, we concur with Chisholm (1990, 603) that grading systems, especially when several assistants are involved, should be as reliable as possible. In cases where several TAs are grading an assignment, we suggest field testing the criteria by having all assistants read the same set of three responses and use the draft grading sheet to assign points. Professors and TAs can then meet as a group to discuss changes to the grading sheet or to clarify the distinctions between major and minor errors. Spending this extra time up front may seem taxing, but it actually tends to save time in the end, both during the actual grading process and when students seek clarification of their scores.

After completing the field test on a

**Table 1**  
**Example of Weighted General Criteria for Evaluating Student Essays**

- |  |
|--|
| 1. Overall argument and responsiveness to the assignment (40 points) |
| 2. Use of readings as evidence to develop the argument (30 points)   |
| 3. Clarity of the prose (20 points)                                  |
| 4. Grammar and spelling (10 points)                                  |

Note: Readers may access the complete grading sheet, accompanying assignment, and other examples at Paul Manna's web site.

**Table 2**  
**Example of Explicit Criteria for Scoring Students' Use of Readings as Evidence**

Use of readings as evidence to develop the argument (30 points)

The paper . . .

- 30—clearly draws on specific concepts from readings to develop a persuasive argument, and demonstrates excellent conceptual understanding.
- 28—clearly draws on specific concepts from readings to develop a persuasive argument, but in discussing the readings, reveals minor misunderstanding.
- 26—draws on specific concepts from readings, but evidence of 1 significant conceptual error or gap in explanation is present, or sometimes uses concepts without clarifying their substantive meaning, or sometimes offers unnecessary summaries of readings that tend to crowd out the paper author's voice.
- 24—draws on specific concepts from readings, but at least 2 of the following are present: evidence of 1 significant conceptual error or gap in explanation; sometimes uses concepts without clarifying their substantive meaning; sometimes offers excessive summaries of readings or direct quotes that crowd out the paper author's voice.
- 22—draws on specific concepts from readings, but at least 2 significant conceptual errors or gaps in explanation are present, or frequently uses concepts from the readings without clarifying their substantive meaning, or frequently offers excessive summaries of readings or direct quotes that tend to crowd out the paper author's voice.
- 20—draws on specific concepts from readings, but at least 2 of the following are present: 2 significant conceptual errors or gaps in explanation are present; frequently uses concepts from the readings without clarifying their substantive meaning; frequently offers excessive summaries of readings or direct quotes that tend to crowd out the paper author's voice.
- 18—even though readings may be mentioned or cited, the paper makes only tangential or no use of specific concepts from the readings in a substantively significant way.

Note: Readers may access the complete grading sheet, accompanying assignment, and other examples at Paul Manna's web site.

evaluation, "some professors forget to stress simple guidelines for their assignments," and handouts that outlined general criteria made this student "feel prepared for any assignment." Moreover, students with parents who teach, or older siblings whose work can serve as a reference point, may enjoy a certain advantage in understanding what goes into A-level work. Circulating clear standards in advance and sticking to them during the grading process provides students equal access to their instructor's expectations.

Second, returning grading sheets to students along with their assignments gives them specific and concrete feedback. Grading sheets provide cues that indicate areas for improvement, while costing the teacher very little time. For students unable or unwilling to approach their instructor to discuss grades and ways to improve their work, grading sheets provide useful information that can help them progress on their own from one assignment to the next. Also, with the criteria listed explicitly on the grading sheet, the instructor can denote areas of strength or weakness simply by adding checkmarks or circles, rather than writing out the same general comment on several papers. In settings where professors know students well (at small colleges, for instance) not having to write generic explanatory comments, such as "inaccurate use of concept," affords professors more time to tailor specific substantive feedback to each individual student. Finally, written comments can be disproportionately negative, as when teachers use them to justify lower grades (Chisholm 1990, 603). A checkmark over one criterion with the annotation "super" and a circle around another with the word "inadequate" may make it faster and easier for teachers to balance the positive with the negative.

Third, we have found that using grading sheets improves scoring consistency across students. Whether an instructor spaces grading out over several days or attempts to plow through the pile in one fell swoop, having a set of explicit and detailed criteria to refer to helps ameliorate the effects of fatigue and memory loss about what constitutes a good or bad answer. In addition, the criteria promote grading consistency across teaching assistants by developing common decision rules. Inevitably, some TAs earn reputations as either "easy" or "hard" graders, and undoubtedly there will always be some inconsistencies across graders in high-volume settings. Establishing a target average score on an exam or writing assignment is one

small number of papers and integrating needed adjustments, it is time to begin grading. In setting up the actual grading sheet, it is important to leave space on the sheet for anecdotal comments. White spaces enable instructors to provide additional feedback to students, or to direct them to marginal comments on their papers or bluebooks. For example, using the criteria in Table 2, students found it most helpful not only when they could learn that they had made a conceptual error but also when they received some signal about the specific nature of the mistake. Circling a 26 and checking the "conceptual error" phrase from the criteria in Table 2 could be accompanied by this quick sentence: "Divided government is about partisan divisions across institutions, not simply differences in opinion between elected officials; see Mayhew reading."

During the grading itself, we find that a fifth and final step, which involves keeping an anecdotal set of comments about how the explicit criteria worked, can pay important future dividends.

Hopefully, the field testing worked out all of the major bugs. However, like any measurement tool, no grading sheet is ever perfect. Thus, it makes sense to monitor gaps and problems so one can integrate refinements into the grading sheet for the next course writing assignment or for use in a subsequent course. Although initially developing grading sheets is time consuming, in the long run it pays off.

### The Virtues of Grading Sheets

In our experience, using grading sheets has several advantages over alternative grading methods. First, and perhaps most importantly, grading sheets set clear standards for students (Glenn 1998; Zeiser 1999). Though years of experience give teachers a clear idea of what constitutes good and bad writing, students lack the benefit of that experience because in practice they rarely read one another's work. As one student recently put it in a course

common way to limit those inconsistencies, but a common grading sheet could limit them even further.

Fourth, grading sheets can help reduce student complaints. In one semester in which we both used grading sheets, across all graded assignments students complained less than 1% of the time about how we scored their work. Those who did had specific questions, and the grading sheets helped focus complaints. In general, the exchanges we had were reasonable discussions of legitimate concerns.

## Not a Panacea, But Still a Huge Help

Overall, we have found grading sheets quite useful. Not only do they make the grading process more consistent and fair, an important goal by itself (Glenn 1998), they also help students more carefully assess their own strengths and weaknesses as writers and analysts of politics. Like any potential remedy, though, grading sheets are not a panacea for all of the challenges associated with grading student exams or papers.

Some professors may consider grading sheets that use explicit criteria to assign points (as described in Table 2) inappropriate in courses where students are writing longer research papers. Likewise, professors may find the potentially formulaic and impersonal quality of grading sheets overly restrictive in small classroom settings. In these situations, using general criteria with weights is

still useful because they allow instructors to clarify their expectations. Omitting the explicit criteria, however, would leave room for rewarding creativity, originality, and extra effort, which are all important traits that may be more difficult to calibrate on an explicit point-based scale.

Grading sheets also do not eliminate the need to make judgment calls about student work. It is impossible to define every possible contingency when crafting a grading sheet that assigns points. The examples in Table 1 and Table 2, for instance, were crafted for an assignment in which students could choose from up to eight different essay topics. Rather than developing eight different grading sheets, the one here was written in general terms so that it could apply in all cases. The result was that during the grading process the criteria needed to be considered with many different possible answers in mind.

A final caution arises: Is the instructor working backward from a grade that he or she has in mind or working forward to a grade based on the numbers that the explicit criteria generate? In other words, as professors we often have an intuitive sense about a piece of student writing. Comments such as "This one feels like a B and that one isn't quite an A" are not uncommon when colleagues discuss their students' work. Even though the strictest use of a grading sheet would always have instructors working from criteria to a final grade, there may be some value in

working backward. In those situations, the grading sheet can still help professors communicate their reactions to students.

The danger in working backward, of course, is when instructors allow impressions of students or their past work to color how they grade the present assignment. We all have our favorite students and those we find less than inspiring. Believing that an "A" or "C" student *must* have produced work similar to that produced in the past can lead professors to use grading criteria in a biased way to justify allocating those scores. But that approach would violate one of the key premises of using a grading sheet in the first place: to help professors grade all students consistently and fairly regardless of their prior reputation or performance in class. Thus, like any grading tool, grading sheets do not eliminate biased grading unless instructors make a conscious effort to grade fairly.

Despite these caveats, we have found grading sheets to be valuable tools, both pedagogically and administratively. They have allowed us to communicate substantial amounts of information to students about the quality of their writing and about their knowledge of key political science concepts. We agree with Zeiser (1999) that instructors who assign papers and writing on exams should be mindful of assessing final products, but should also gear instruction toward helping students become better writers. Grading sheets make both goals achievable.

## Notes

1. Instructors sometimes sidestep these challenges by assigning no writing at all, an approach that has problems of its own.

2. Shifting standards may be less of a problem if one grades an exam section-by-section (e.g., grading all identification questions in one

sitting). However, grading longer exam essay answers or papers typically requires several days of work.

3. Survey researchers go through a similar process when they develop codes for open-ended questions.

4. At the College of William and Mary, these courses contain approximately 40 students per section. A professor has a total of roughly 80–100 students per semester and does all of his or her own grading.

## References

Chisholm, Donald. 1990. "Between Leibniz and Voltaire: Exams and Grading in a Less than Perfect World." *PS: Political Science and Politics* 23 (December): 600–604.

Damron, Danny. 2003. "Facilitating Critical Feedback on Capstone Papers through

Student Poster Sessions." *PS: Political Science and Politics* 36 (October): 777–780.

Glenn, Brian J. 1998. "The Golden Rule of Grading: Being Fair." *PS: Political Science and Politics* 31 (December): 787–788.

Zeiser, Pamela A. 1999. "Teaching Process and Product: Crafting and Responding to Student Writing Assignments." *PS: Political Science and Politics* 32 (September): 593–595.