WPA: Writing Program Administration, Volume 13, Numbers 1-2, Fall/Winter 1989
© Council of Writing Program Administrators

Team-Grading in College Composition

Nedra Grogan and Donald A. Daiker

Nearly every director of a first-year college writing program can anticipate problems related to writing assessment. The sources of these problems may be varied—a department chair or dean concerned that grades in college composition courses deviate unsatisfactorily from the all-department or all-college average; colleagues who believe that graduate assistants are undermining the quality of education by consistently assigning grades that are too high; graduate assistants who are unsure of their grading criteria; and students who are absolutely convinced that their instructor is paid by the amount of red ink she uses and promoted by the number of papers she fails.

Here at Miami University we have tried to address several of these problems by instituting a program in team grading within the first-year composition sequence.* For the past four years now, a group of instructors has met twice each semester in order to evaluate papers written by students in composition sections other than their own. In what follows, we would like to outline a procedure for establishing a program in team grading and to suggest how such a program benefits students and teachers while contributing significantly to learning.

To get started requires a writing program director willing either to become program coordinator or to recruit one. The coordinator's first step is to secure the approval of the department chair and other influential administrators—both because their endorsement encourages others to participate and because they are able to pick up the minimal costs (duplicating, secretarial help, refreshments) that team grading entails. Once the coordinator makes clear that team grading is a powerful yet inexpensive form of faculty development that simultaneously addresses problems of grade inflation and grading consistency, administrative support generally follows. An optional next step is for the coordinator to appoint a faculty member or graduate student as associate coordinator or chief reader.

Recruiting Colleagues

To achieve the sense of community that Edward White reminds us is so vital to a successful essay reading (24), program leaders should recruit only colleagues who want to participate. Forced participation will not work. Indeed, our only unsuccessful experience in team grading occurred

WPA: Writing Program Administration, Volume 13, Numbers 1-2, Fall/Winter 1989
© Council of Writing Program Administrators

when we *required* all first-year graduate assistants to take part in the program. The result was a day of long faces, resentful comments, and low reliability scores. Now we *invite* participation—with a departmental memorandum, sent well before the semester begins to enable instructors to include information about team grading in their course syllabus. The memo gives two dates for team grading and describes the key points of the program:

1. **Instructors may participate in one or both team-grading sessions.** We schedule one session on a Saturday at mid-term and a second session on the Saturday before final exam week. The two sessions are run identically but independently. For instructors using the portfolio approach, mid-term team grading enables their students to get an early evaluative response to their papers. For all instructors, participating in a finals-week session can reduce the drudgery of semester-end grading done alone.

2. **Instructors may require or make optional student participation in team grading.** Some instructors require their students to submit a paper or two for team grading; some even specify which paper is to be submitted. But most of us make student participation completely optional. Each student decides for herself whether to participate and, if so, how many papers to submit.

3. **Instructors may use the results of team grading in whatever way they choose.** Although participating instructors must attend a training session and eventually "agree to common standards for the sake of the test" (White 25), they are free to determine how the results of team grading will affect course grades. For some instructors, particularly those who have not previously evaluated the paper, the team grading score becomes the paper score; for others, the team grading score is averaged with the original score; and for most of us, the team grading score counts only if it is higher than the original score.

4. **Each paper will be read as a response to an open assignment**—unless the assignment is specified (25 words or less) at the top of the first page.

5. **All papers will be graded holistically**—that is, with a "quick, impressionistic qualitative procedure for sorting or ranking samples of writing" (Charney 67)—on a scale of A+ to F following standard practices of the Educational Testing Service. Readers will be trained in scoring procedures on the morning of the grading session following procedures described in Odell and Cooper (36-38), White (163-67), and Myers (42-46). Each paper will be read and scored by two different readers and, in instances of a discrepancy (more than a full grade difference in scores), by a third reader. All scores will be reported to the instructor—and to no one else.

6. **Each instructor may submit up to 60 papers.** Limiting submissions prevents the unfair and awkward situation of one colleague's submitting three times as many papers as another.

Pre-Scoring Procedures

Before papers submitted for team grading can be evaluated, they must be collected, randomized, and coded. Although every program can devise its own procedures, careful arrangements need to be made to ensure anonymity of the writer and the classroom teacher and to make sure each graded paper finds its way back to the right teacher and student. This can be done by assigning instructor codes (a two-digit number) and paper numbers (a six-digit random number either computer-generated or copied from tables found in statistics textbooks). To give each submitted paper two readings, the classroom teachers who will be participating are divided into two roughly equal "teams," with each instructor's papers going over to the opposite team. The two teams' papers must be kept separate so that no instructor reads a paper from her own class.

In addition to paper preparation, other preliminary procedures include selecting the sample papers—the range finders—to be used to train readers (Daiker and Grogan). Two or three days in advance, the most experienced participating instructors submit several unmarked papers that they find especially appropriate as range finders. The chief reader then duplicates these papers and convenes a meeting to select one or two papers for each letter grade.

Each team-grading program must, of course, determine its own scoring criteria and scoring scale. We depart from the standard ETS six-point and nine-point numerical scales in favor of traditional letter grades—A+ through F. While letter grades are a distraction in most holistic evaluation sessions, they are a known reality in college composition courses; thus, we follow the system familiar to teachers and students alike. That way, students need not learn the meaning of a "5," and teachers need not convert numbers into letter grades.

The Team-Grading Session

The reading follows ETS procedures for a holistic evaluation session: the chief reader introduces the scoring scale and descriptive criteria and then leads raters through the individual scoring and group discussion of the anchor papers. Readers practice making quick, impressionistic judgments without making any corrections or revisions in the papers (Cooper 3). Once readers are calibrated, or in agreement, "live" papers are distributed and the actual scoring begins.

WPA: Writing Program Administration, Volume 13, Numbers 1-2, Fall/Winter 1989
© Council of Writing Program Administrators

As in any holistic scoring session, the length of a reading depends on how many papers each participant must evaluate. If participants each bring sixty papers to team grading, a team-grading session that includes a calibration period of sixty to ninety minutes will take approximately eight hours. If colleagues hesitate to invest an entire work day in reading papers, it's useful to remind them of the time they spend—all alone in offices or dens—grading one, two, or three sets of papers. Team grading may take a full day of work on Saturday, but it leaves Saturday night and all day Sunday free for family, fun, or anything other than grading papers.

Advantages of Team Grading

A team-grading program offers advantages for students, instructors, and the writing program. Students are informed about this optional evaluation plan at the beginning of the term (thus the importance of a memo to faculty before they plan their course syllabi) and are introduced to team grading as an incentive for them to rethink and revise papers that have already been evaluated: students are less likely to regard graded papers as "over and done with" if they are offered a chance to rewrite them and given a shot at a higher grade. Finally, a team-grading program in which student participation is optional empowers students to make a series of important decisions. For each piece of writing students produce, they must decide whether or not to submit it for evaluation by instructors other than their own. And for each paper they submit, they must also decide upon the nature and extent of their revision. Team grading, like portfolio grading, places key educational decisions squarely in the hands of student writers.

Many advantages of team grading for students accrue to instructors as well. Most importantly, team grading helps us do a better job of teaching writing. It does so primarily by encouraging our students to take a second and third look at their own writing in preparation for submitting it to a brand new audience of readers. But it also does so by helping us create a more positive classroom atmosphere. Morale is enhanced not simply because our students have been given more choices but also because the structure of team grading places instructors and students on the same side. In traditional grading, of course, it's difficult to avoid an adversarial relationship, the student submitting the paper on one side and the instructor evaluating it on the other. In team grading, by contrast, there is no reason why instructors cannot openly cheer for their students to come away with high grades. On the Monday when we make known the results of team grading, we warmly applaud our students who have improved their scores and commiserate with those who haven't. Whatever the results, we're together on this one.

Team grading is a further benefit to instructors—and students—in resolving disputed grades. Without team grading, a student who honestly questions a grade has limited options. Either she swallows her disbelief and disappointment, or she approaches the instructor with her problem. The latter poses risks, of course, since—even when tactfully presented—it may be taken as a direct challenge to an instructor's authority and competence. In earlier years when we invited students to resubmit disputed papers for regrading (by another instructor or by a group of students from the class), there were few takers—and they were virtually all male. Team grading removes the confrontational element from grade disputes. No longer is the student challenging or even questioning an instructor's taste or judgment; she is simply submitting her work—usually after it has been revised—to other readers. It is the same privilege reserved for us in the profession when our own work is not as kindly received as we had hoped.

At the same time that team grading affords students new opportunities for writing and response, it allows instructors to retain a significant amount of autonomy and control. Since all participating instructors help determine grading criteria through their responses to sample papers, the grading standards applied at any team-grading session will obviously reflect the group's point of view. Once the team-grading session ends, however, individual instructors are free to use the results of team grading in whatever way they choose. That is, each instructor has the choice of using one or both grades that a student receives from team grading. Indeed, one regular Miami participant guarantees an "A" to any paper that receives an "A" from *either* evaluator. The individual instructor also determines the relationship between the grades received during team grading and the paper's original grade: some instructors average the two grades, but most of us count only the higher between 1) the original score and 2) the average of the two team-grading scores. The latter arrangement enables us accurately to describe team grading as a "no-lose" proposition, a riskless opportunity for students to better their semester average and to experiment boldly with new forms of writing. Each instructor gives different color to team grading by determining how its results affect semester grades.

But since team grading is an effective means of controlling grade inflation and grading inconsistency, its benefits extend beyond the classroom to the writing program and the department. Team grading addresses grade inflation in two ways: by establishing high standards of evaluation and by building teacher confidence. Colleagues tell us that they apply higher standards during team grading than when evaluating papers submitted in their own courses: they estimate the difference to be between a third and half a grade. So the course paper that receives a "B+" is likely to fare no better than a "B" or "B-" from team grading. Anonymous grading tends to depress scores slightly. Even more important for

WPA: Writing Program Administration, Volume 13, Numbers 1-2, Fall/Winter 1989
© Council of Writing Program Administrators

reducing grade inflation is that team grading enhances the self-confidence of participating instructors, especially inexperienced ones. Since an obvious source of grade inflation is insecurity—an instructor who questions her own judgment may hesitate to assign low grades for fear that she cannot justify them, an equally obvious remedy is to build confidence by enabling instructors to compare their evaluations with their colleagues'. Such confidence building is a major function of the calibration period that opens each team-grading session.

Calibration is also the primary instrument for achieving consistency in grading. As instructors evaluate and then discuss a series of sample papers, they invariably move closer to consensus in terms of both what scoring criteria to apply and what scores to assign. Whether an instructor's initial scores are too high, too low, or simply inconsistent, the gentle force of peer pressure characteristically nudges her in the direction of her colleagues. It stands to reason that at least some of what is learned during calibration will carry over to other evaluative tasks. Indeed, White is convinced that the holistic essay reading integral to team grading provides "the most effective in-service training for the teaching of writing yet discovered" (166).

Some Limitations of Team Grading

Notwithstanding its contribution to an effective writing program, team grading has several real limitations. Perhaps the two most crucial involve scoring reliability and student perceptions. Problems with reliability arise because, unlike most holistic scoring sessions, raters are not evaluating responses to a single prompt but to several different ones, including free choice. Indeed, within a typical scoring session, raters evaluate papers that differ not only in topic but in kind as well: a story about the championship soccer game may be followed by an explanatory essay on antique furniture or a persuasive piece on the treatment of AIDS victims. Given such variety in student writing, team-grading raters may not be able to achieve the scoring reliabilities possible with a single prompt (Cooper 19). During actual scoring sessions, our discrepancy rate—the percentage of papers with two scores that differ by more than a full letter grade—ranges from ten to twelve percent.

If scoring reliability is made more difficult when students choose what to write—in part because it complicates the selection of sample papers, that fact has not prevented the widespread use of open assignments in large-scale assessment situations. Students are permitted virtual freedom, for example, in the "best writing" component of the NCTE Achievement Awards in Writing competition. They also have wide choice in the portfolio-based evaluation program of the State University of New York, Stony Brook: one portfolio item is "an academic essay of any sort" (Elbow and Belanoff 98). Even the timid Educational Testing Service has suggested the inclusion of "a paper of the student's choice" within a writing portfolio designed for college admission (Camp 97). But while the absence of a prompt may lower scoring reliabilities, it probably enhances test validity: an assignment requiring students to discover what to say seems a more comprehensive and therefore more valid measure of writing competence than one in which the topic is chosen for them. Even if it were not for enhanced validity, the risk of lower reliability scores may be a small price to pay for gains in student motivation and class morale. Our responsibility as evaluators should not diminish our far more important responsibility as teachers.

A second crucial limitation of team grading involves student perceptions. Unless our students understand what team grading can and cannot do, it may confuse and even intimidate them. After all, the image of a room full of composition teachers grading papers does not bring immediate joy to most students' hearts. They may even feel, despite our talk about "real" audiences, that strangers have no business evaluating their writing. Some students are understandably more comfortable when their classroom teacher, who knows how hard they've worked, is the one dispensing grades. And they have some reason to be apprehensive, too, since—with author and assignment bias removed—team graders tend to score papers a notch lower than classroom teachers. If that's not bad enough, their papers are returned with only a grade—no comments, no justification, not even a suggestion for improvement. Without careful explanation, it would be possible for students to perceive team grading as part of the same pedagogical arsenal that includes standardized testing, detention halls, and dunce caps.

But once team grading is made voluntary, all its intimidating power is lost. So long as students are free to participate—if they want to raise their grades, if they want an early response to a piece in their portfolios, or if they simply want to know how others evaluate their work, they are in a better psychological position to understand what team grading does, and what it cannot do. What it does is to provide students a holistic evaluation of their work by two instructors trained in writing assessment. What it cannot do is explain the evaluation. But this is a limitation that students will accept so long as they know that written explanations would triple the time of grading sessions and cause even the most committed instructor to withdraw from the program. Team grading is feasible only because holistic evaluation allows papers to be read and scored in three to four minutes.

Team Grading in the Classroom

A good way to help students decide whether to participate in team grading is by involving them in a practice session. A class period devoted

WPA: Writing Program Administration, Volume 13, Numbers 1-2, Fall/Winter 1989
© Council of Writing Program Administrators

to team grading not only gives students a realistic sense of the program's strengths and limitations but helps them develop their own evaluation criteria—useful both for responding to the writing of their classmates during peer workshops and for assessing and then revising their own written work. A practice team-grading session requires only three components: 1) a very brief overview of the principles of holistic assessment; 2) a description of the scoring scale and scoring criteria; and, most important, 3) a set of sample papers for students to read, score, and discuss. The overview quickly characterizes the essentials of holism: multiple, independent readings; a single overall score from each reader; anonymous scoring; trained readers from similar backgrounds; and a common scoring scale with explicit scoring criteria. Both the scoring scale and scoring criteria should reflect institutional practice: if the college catalog defines "A" as *excellent* and "C" as *satisfactory*, that's a good starting point for a more detailed description of points on the scoring scale.

But the heart of team grading in the classroom is a set of papers for students to score. Select three to five papers that range in quality from excellent to below average. If possible, they should be papers written by students other than your own; if not, keep their source and identity secret. Begin by distributing a single mid-range paper and asking students to evaluate it in terms of the scoring scale and scoring criteria. Students record their score anonymously on a slip of paper, fold it over, and pass it forward. All the scores are then read aloud one by one and recorded on the chalkboard. As anyone who has participated in holistic scoring knows, this is an exciting time: students are eager to discover how closely their own evaluation tallies with those of their classmates. Once the scores are recorded for all to see, discussion begins. Like a chief reader, the instructor may want to begin calibrating the students—moving both high and low raters toward the middle—so as to illustrate what takes place during an actual team-grading session. Following discussion, other student papers are distributed, scored, and discussed—one by one. By the end of the class period, students have a much clearer sense of what happens during a team-grading calibration period and, for that reason, of how papers are subsequently evaluated. They can now better appreciate the strengths and limitations of holistic evaluation, and they can more intelligently decide whether to submit one or more of their own papers for team grading. And if any student wants first-hand experience with holistic evaluation, she can be invited to observe the next team-grading session.

Although students may be the major beneficiaries of team grading in college composition, the writing program itself gains in substantial ways. First, team grading is a direct and effective way of addressing problems of grade inflation and grading inconsistency. Second, team grading helps build department morale by bringing graduate students and faculty members together to perform a common task. Third, team grading contributes significantly to the professional development of faculty members in composition: extended discussions of student papers with our colleagues not only motivate us to reexamine our own values but to ask ourselves the kinds of questions that can lead to improved teaching and, on occasion, to an important research project. Finally, a program in team grading announces to the department and university that the writing program faculty willingly gave significant time and effort in order to fulfill their professional and teaching responsibilities. A team grading program helps create the kind of environment in which writing teachers and the teaching of writing flourish.

Works Cited

Camp, Roberta. "The Writing Folder in Post-Secondary Assessment." *Directions and Misdirections in English Evaluation*. Ed. Peter J.A. Evans. Ottawa: CCTE, 1985. 91-99.

Charney, Davida. "The Validity of Using Holistic Scoring to Evaluate Writing: A Critical Overview." *Research in the Teaching of English* 18 (1984): 65-81.

Cooper, Charles R. "Holistic Evaluation of Writing." *Evaluating Writing: Describing, Measuring, Judging*. Eds. Charles R. Cooper and Lee Odell. Urbana, IL.: NCTE, 1977. 3-31.

Daiker, Donald A., and Nedra Grogan. "The Selection and Use of Sample Papers in Holistic Evaluation." ERIC TM012942. Forthcoming in *Journal of Advanced Composition*.

Elbow, Peter, and Pat Belanoff. "State University of New York, Stony Brook Portfolio-Based Evaluation Program." *New Methods in College Writing Programs*. Eds. Paul Connolly and Teresa Vilardi. New York: MLA, 1986. 95-105.

Myers, Miles. *A Procedure for Writing Assessment and Holistic Scoring*. Urbana, IL.: NCTE, 1980.

Odell, Lee, and Charles Cooper. "Procedures for Evaluating Writing: Assumptions and Needed Research." *College English* 42 (1980): 35-44.

White, Edward M. *Teaching and Assessing Writing*. San Francisco: Jossey-Bass, 1985.

*The writing of this article was supported by the Miami University Committee on Faculty Research.