16 Students Using Evaluation in Their Writing Process

Jacob S. Blumner
Kent State University, Stark Campus

Francis Fritz
University of Nevada, Reno

Jacob S. Blumner teaches at Kent State University, Stark Campus, in Canton, Ohio.

Francis Fritz is a graduate fellow in the composition and rhetoric program at the University of Nevada, Reno and is involved in the Core Writing program's portfolio project and the development of the Student Core Writing handbook. He currently entertains a perverse obsession with grading and is initiating a research project involving peer evaluation.

"Here you are, Mr. Fritz. This one had a really strong focus," Christine told me.

"Great," I answered, handing her another folder. At this point in the class, students were nearly piling up on each other waiting to get the next folders to score. Even though I was busy with the effort of keeping the folders in order, making sure each paper got two different readings, keeping students moving along, I could still overhear bits of the conversations of the scoring partners: "I don't think so, Joe. This piece has a lot more originality than you say it has. Who would have thought to argue to legalize fake IDs?"

"Good point," Joe responded, "but are originality and craziness the same thing?"

I heard another pair speak: "But there's only one spelling error in the whole paper. I'd still give it a five."

"Yeah, but the paper is only two pages long. Are you sure it's worth a five?"

Two students suddenly shout simultaneously, "I'd give coherence a four!" They both laughed out loud.

Paul, stepping up to me to pick up another folder, sighs, "This stinks! It's really hard."

I smiled, happy to share this understanding with him, and handed him another folder. He looked at me again and sighed as before. On my left, two students were waving their hands and talking quickly.

"Some of these sentences I can barely understand. I thinks it's a three."

"A three? The introduction is pretty good. That first sentence about potholes is hilarious. It's at least a four."

"But what about...."

"Mr. Fritz, can you check this one? We can't seem to agree."

I shake my head and smile. "Sorry. Today it's your job. Just keep talking about it. I'm sure you can work it out." More students sidled up and handed me folders containing the papers they'd scored. I quickly exchanged them for one that required another reading.

Finally, Aaron stepped up, a student who had often shown his strengths as both a writer and a reader. He looked at me earnestly.

"Mr. Fritz, this is tough. Why do I have to do it, instead of you?"

We considered this same question as we prepared to teach our classes, because the comments we made on student papers, the recommendations for improvements in drafts, the explanation of why students received the grade we gave them seemed to us to be possibly work done in vain. Did students actually learn the kinds of writing strategies that would make them more effective writers, or did they merely learn how to give to us, the instructors, another version of the right answer? Did our comments actually invite them to contemplate the intricate weighing of rhetorical and cognitive possibilities, or rather, did they merely add to or subtract from their papers what they interpreted to be words and sentences we prescribed? Were our grading practices sufficient to instill in our students a sense of ownership in their writing, or did the opposite occur—ownership was neutralized by the very grading that we had been using for so long?

We began to think about what use our grading practices were compared with our efforts to create in our classes a greater awareness of writing possibilities, a stronger sense of ownership of their writing practice, a greater understanding of their role in a community of writers, and, most important, a deeper investment in what our students chose to strive for in their efforts to write. In an attempt to address these problems, we drank coffee and discussed how we might resolve these issues. We began by sharing our frustrations, our successes, and our ideal classrooms. While we debated our teaching practices, we discovered we were constantly returning to three major questions: (1) How can students become more integrally involved in the evaluation process? (2) How do we increase student dialogue and community? (3) How do we increase student ownership of writing?

One problem we see with giving grades is that grading can become the driving or sole purpose for instruction. In fact, evaluation

can direct pedagogical practice in potentially negative ways (e.g., instructors teaching to a test). We decided to try to turn this idea on its head. What if we could find a way to make evaluation and grading serve our goals for instruction and process?

The purpose of our composition classes is to assist students in becoming more successful writers, not merely to generate grades. Therefore, we thought, why not involve students in the assessment and evaluation process? Already we have our students in their peer-response groups read and respond to their fellow writers' drafts. Why not let students—as a community of readers as well as writers—grade each other's writing efforts? Wouldn't they experience greater control and ownership over their learning and achievement (Williams 267, 274)?

As many teachers know, one apparent quality of evaluating student papers is its uncertainty. Each of us struggles to come to fair measures regarding what constitutes successful or poor writing. What teacher hasn't seen a complex mixture of both qualities in a single paper? Each of us comes to our own conclusions about what we measure and the degree to which we measure it against other qualities. We had come to believe that our students could only benefit from this same experience in their attempts to justly evaluate their peers: "The practice of peer evaluation would give students more responsibility for their own successes and failures" (Williams 267). Their struggles to come to decisions would encourage them to look more closely at what constitutes good writing, how readers and writers can talk about it, how we already continually judge its quality, and the consequences of those judgments.

Since assessing and evaluating student writing is difficult, we thought that it might be better if the students did it in pairs because this way, they could each read a particular paper and discuss the merits or weaknesses of the piece and share the responsibility for evaluating it. Beyond assessment and evaluation, students can develop a sense of community as writers with their partners; feel more secure in their decisions, which will eventually be shared with the rest of the class; and become more aware of, through articulation, why they make the choices they do as readers and writers.

After surveying a variety of approaches, we decided to experiment with a variation of trait scoring. We noticed that when students worked in peer-response groups, they were overly concerned with surface features, often overlooking organization and audience. We assumed trait scoring would encourage students to examine particular

writing qualities, specifically thesis, organization, sense of audience and purpose, coherence, paragraphs, and surface structure. Next, we designed a generic six-point rubric in which we tersely described the qualities of each trait under each score (see Figure 1). Later efforts at composing guidelines for evaluation moved us away from a generic rubric and toward rubrics that are constructed according to the specific paper assignment. For example, students working on an interview paper created the criterion of "claim," which, in the context of the class, meant taking the interview material and using it for their own argumentative or contemplative purposes. In addition, students placed this criterion first on the rubric because they believed it was the most important one. In short, students combined our original criteria of thesis, audience, and purpose into one criterion because they saw these elements as integrally connected and central to a successful paper.

Understanding these problems, we knew that preparation was necessary before that day of evaluation. During the first week of classes, we had introduced students to the grading system we intended to use. We contemplated many possibilities, and although we saw many potentially successful methods for moving from scores to grades, we chose the following:

- The criteria to earn an A grade for the essay writing component of the course were that students needed to receive at least a 5 out of 6 on four of the six papers for the class, and no papers could receive below a 3.
- To earn a final grade of B, students needed to earn a score of 5 on two papers and at least a score of 4 on three other papers.
- For a grade of C, students needed to earn a score of 4 on four papers.
- In addition, final grades for the course included the evaluation and grading of participation, attendance, and journals. These were simply evaluated on a satisfactory/unsatisfactory scale.

Although we wanted students to be involved in the accountability of giving scores that would result in grades, we chose not to place specific grades on papers or equate scores with grades in any direct way so that they would not feel overwhelmed with their newfound responsibility. Each paper received scores from two different pairs of readers. If two or more groups gave a paper the same score, that score stood. In the event of a split score, we acted as final readers. Also, in the event that we felt a student paper received a particularly unjust score for a

---

Author's Code Name and Title:

Thesis	6 5 4 3 2 1
Organization	6 5 4 3 2 1
Audience/Purpose	6 5 4 3 2 1
Coherence	6 5 4 3 2 1
Paragraphs	6 5 4 3 2 1
Surface Structure	6 5 4 3 2 1
Overall Score	6 5 4 3 2 1

Comments (use back of sheet if necessary):

Evaluators' Code Names:

Figure 1. Generic six-point rubric for trait scoring.

number of reasons—including, for instance, the possibility of the evaluators' dislike of the topic—we would raise the score accordingly. We also told them we would never lower a score. We realized that by maintaining our role as final arbiters of grades, we could not fully give the class evaluative control, but we felt we needed to be able to have some direct influence on scores because students are inexperienced in the process, and the academic consequences could be significant. Thus, our right to raise scores operated as a kind of safety net against unjust scoring. Students could feel they had latitude in what they might give and receive, and we hoped students would still be willing to take risks in their writing as well as to evaluate honestly.

During the week prior to evaluation, we brought in a number of transparencies of examples of student essays from previous classes. We handed out our rubric and spent a short time reviewing the various descriptions. Students listened quietly, some nodded knowingly, and there were very few questions.

After having a student read the paper aloud, we walked through the assessment and evaluation process, attempting to model the kind of thinking and talking behavior we hoped our students would use. We continually moved from paper to rubric and back, keeping our evaluation focused and limited. Finally, we gave a number to a trait and then moved down to the next one, again talking through our process until we had completed our discussion of the paper.

Next, we placed a second paper on the overhead, asking a different student to read it aloud. At this point, we started a class discussion, asking our students questions about what they saw, making sure our questions were open-ended in order to encourage dialogue. We began with the general question "What is this paper about?" Students gave tentative answers, and we encouraged them to talk about those answers. As the discussion progressed, we guided students first to examine more global issues like organization, asking them questions about the specific arrangement of ideas, and then moved to more local concerns like the ordering of sentences in a paragraph, and finally concluded with talk about surface features. Our goal was to encourage conversations about papers, taking care not to rush them into making scoring decisions. We wanted our students to begin thinking through the scoring process, to practice talking about it, describing what they believed they saw, and what they felt worked and didn't work. Eventually, we did ask students to score each trait, knowing that asking them to commit to a score would increase their awareness of the complexities of scoring and the need for careful observation and deliberation.

It wasn't long before students began to see the difficulties in our rubric, and by extension, any rubric. First, they began to recognize that this method of scoring compartmentalizes writing qualities, denying the interconnectedness among these qualities and its inseparability from the content. Students persistently raised questions and made comments which addressed the inextricable nature of many of these qualities. For instance, they found decisions regarding organization difficult to evaluate without taking audience into consideration. Yet, the rubric asked them to do so. Of course, having them recognize this difficulty was one of our pedagogical goals. In spite of this problem, we still thought this method was most beneficial for the purpose of having students look closely at writing. One student claimed that he saw that his attention had mostly been on correctness, but that he gave little attention to an engaging thesis or sophisticated coherence. For this reason, we believe that trait scoring gave students specific directions for addressing the writing.

Second, students noticed that trait scoring with a generic rubric overgeneralizes the qualities it is asked to assess. Much class time was spent discussing, or "norming" (bringing the class toward agreement), what each quality might consist of for a particular paper. During these discussions, students reiterated the difficulty they had with the explicit definitions of the different traits. They had difficulty with the

varying definitions of coherence. At first, students said coherence meant "flow." Through further discussion, they narrowed their definition to the presence of transitional words within a paper, such as "however" and "yet." But we asked the students, if a paper lacked transitional words, should the paper be scored lower for the trait of coherence? They quickly recognized that coherence was much more complex than the presence of transitional words. So they struggled with developing a definition that was specific but at the same time not so abstract that it became useless. Over time, the students in the class, as a community, began to come to decisions about what these traits should mean, above and beyond our own short descriptions. Essentially, they began to take ownership of the criteria for evaluation, testing them against their own experience with the piece of writing. Again, we see this type of discussion as vital to students' development as writers. So, though there is difficulty in what each term means in relation to a specific writing task, the struggles students go through, like those instructors go through, are part of the process we believe necessary to becoming a good writer.

Third, students quickly began to concern themselves with the value placed on the traits we had had them score. They recognized the problems in privileging some qualities over others, and many of them quickly came to realize that they held what one student called his "lopsided" idea of what good writing should be. When we first began to use peer evaluation in our classes, we selected particular traits to be evaluated. But by choosing and naming the traits we do, we inevitably devalue other traits that could be scored. For instance, we had particular trouble using *development* as a trait unto itself. We hoped that issues of development would be included in how students assessed other traits, but we realized we could not be certain that development would in fact be assessed. One response to this problem which we have begun to employ is to have students develop their own list of traits and a rubric to accompany them. We chose to do this because in the process of developing criteria as a community, students engage directly with the purpose of and goals for the assignment and define which elements might make a successful response to that assignment. In our reading of the literature, we also learned that "students and teachers tended to differ in the criteria they employed for deciding what constituted a successful completion of the task and in the criteria they employed for ranking the essays according to scoring criteria" (Ruth and Murphy 202). So, to address this problem, we simply decided to let students help in developing the criteria. This way we

could, as both a class and community of writers, design what is valuable within the context of the assignment itself.

Fourth, our students often expressed dismay at being forced to score these traits on a linear scale. Students focusing on placing a paper somewhere along this scale have a tendency to move away from talking about specific issues in the writing—as they had been accustomed to in their peer-response groups. If students find that a paper exhibits strong coherence at the beginning but less at the end, in their effort to grade the paper they lose the possibility of acknowledging and discussing specific successes or weaknesses of that trait.

Finally, students had to take the scores they had given each trait and try to reduce them to one numerical score, and this required a relative value to be placed on each trait scored. This last step increased the frustration they had already experienced in their struggle to give each trait a score. In response, we invited them to develop a strategy for turning the trait scores into one value for the paper. Students attempted to find a mathematical solution, but they quickly found that no formula would fairly convert the trait scores into a single number. The students suffered from the same problem of subjectivity that instructors confront in their effort to evaluate. Through this process, students discovered that there were many possibilities for what readers value as good writing, and we think this is a useful realization for them to grasp as writers. And acting upon this realization, we assisted students in class discussion again, as a community, in using the trait scores to intuitively decide on an overall score for the paper.

After discussing the benefits and shortcomings with the class and modeling the kind of behavior we wanted them to use during the scoring sessions, we placed another student paper on the overhead. This time we asked students to talk with their partners about each score, recreating the dialogue modeled in class. We gave them two conditions: They couldn't give half scores, and they needed to come to an agreement on each. The chatter quickly rose as they struggled with making their evaluative decisions. Many students became animated, adamantly defending their choices, then shifting, reconsidering the evidence offered by their partner, finally making themselves decide upon a score. One student exclaimed, "I don't care if you don't like the topic. It's still well written. Find a place where it isn't." After many minutes of haggling, they reached an agreement. Encouraging this kind of dialogue, we gave them as much time as they needed. This didn't seem to be a process that could be rushed. When they were finished, we again opened up discussion to the class.

We wrote the name of each trait up on the board and collected tallies: Three pairs gave the paper's organization a 5; seven pairs gave it a 4; one gave it a 3. Pairs were asked to talk about their scoring. They were expected to give evidence, to point to specific examples of where they thought the paper succeeded or failed, and to explain why that success or failure deserved the score they gave it. We had begun to "norm" the readers, while simultaneously aiming them toward an even greater awareness of which qualities they were discussing, why they saw them the way they did, and how these qualities applied to successful writing. All throughout this discussion, the class returned to the inherent problems of scoring, but each time developing and expressing a greater sophistication about those problems and the pote.~+iʌl solutions to them. The discussions in our classes have been lively, informative, and fruitful.

Given that students would be responsible for grading each other's work, we found it important to spend a good portion of class time having them score and then discuss their scores. The students in the class could see some of the dangers easily enough. Some pairs were repeatedly overly critical, ready to punish the slightest mistake in a paper with a low score. Others were all too ready to give a string of sixes. But as instructors, we exerted a helpful influence in how our students decided what constituted successful or less successful writing. We found class discussions an opportune place to share our opinions, as readers, of what we value in writing. In many instances, this provided the guidance students needed.

Before the day students were to score their classmates' papers, we had prepared them in the following ways: (1) discussed the use of rubrics, their benefits, and drawbacks; (2) facilitated students' designing of an assignment-specific rubric; (3) modeled assessment and evaluation; and (4) had students practice evaluation as a class and then with their partners. Because we believe that conversation is "the ultimate context within which knowledge is to be understood" (Rorty 389), during this process we fostered a sense of community by negotiating the terms for evaluation, honoring the various voices that contributed to the discussions.

On scoring day, students came to class with two copies of their papers in a manila folder with the names removed and a number on the folder and on the top of each paper to help us with identification. In addition, we provided forms on which they could record their scores. We handed out the folders to each pair. Each member received

his or her own copy for easier reading. We asked them not to write on the papers.

And then they began scoring them, following through on the model we helped them to create through their practice in class. After they completed reading the papers, the chatter again rose up. This time students were much quicker about deciding what they believed the scores should be. As they finished, they came up to turn in folders and to take new ones. After all the hours of preparation the class had given to this task, the students approached the work with excitement and seriousness. Each time a student returned to pick up another folder to read and score, it became clear to us that they had made their decisions about their fellow students' papers in earnest. Rarely, it seemed to us, was a decision made lightly. With only a few folders left to be read, some of the students talked more openly about the difficulty of scoring fairly:

> Stephanie said, "This is hard. I've never had to read papers so closely and then argue about it."
> "Do you think this gives you a better idea of what good writing might look like?"
> Stephanie stopped for a moment, thinking about my question.
> "I think so. I think I see that I'll just have to be more careful when I do my writing. I can see I have a lot of stuff to think about."
> "Good, Stephanie," I smiled. "Here's your last folder."

Works Cited

Rorty, Richard. *Philosophy and the Mirror of Nature.* Princeton: Princeton UP, 1979.

Ruth, Leo, and Sandra Murphy. *Designing Writing Tasks for the Assessment of Writing.* Norwood: Ablex, 1988.

Williams, James D. *Preparing to Teach Writing.* Belmont: Wadsworth, 1989.

Interlude

In my collaborative/self-grading program, students "grade" only at semester's end— we don't do intermittent or paper grading. We *do* work at looking at how they view grades after midterm. Leading into the final essay, we talk about grades and how they see them, as well as how others see them. I want them to appreciate how difficult and complicated grades really are. We also spend time in class reviewing the evaluations and peer feedback they gave and received, looking for places where they captured in writing what they have learned about writing as well as where they have made improvements. The goal setting directly helps with this. . . . Then they have to establish what criteria they are using for arguing for a grade: improvement; meeting goals; understanding what writing is; the ability to experiment and try things with their writing which may not have succeeded, but from which they have learned; and others. After establishing the criteria, they have to show how they have met them. They must refer to their writing and must be able to talk about it. . . . After they argue on how well they have met the criteria—I call it "building value"—they then have to tell me what the value is worth in terms of a grade. They also expect or assume about them as writers on the basis of that grade. If they don't meet those traditional expectations, then they need to argue further why they should get the grade in spite of that. I read the papers and look at how well they've constructed the argument, noting especially the criteria and explication of how well they've met them. I prepare a response and we discuss it.

    —Nick Carbone
     University of Massachusetts at Amherst