

What Englwrit 112 Teachers Look For in Unit I Papers: A Case Study of Reader-Based Evaluation of Student Essays

David Fleming, with Dan Biegelson, Willie Bordwine, Amanda Carr, Shehryar Ikram, Ari Feld, John Gallagher, Ruth Hariu, Emma Howes, Haesong Kwon, Emily Madsen, Kyle McCord, Greg Sargent, and Penni Vaget

Evaluating student writing is one of the most complicated things that writing teachers do. It's so complicated, in fact, that the Writing Program often brackets the topic in both its curricula and teacher training. This reticence is understandable: *Basic* and *College Writing* are courses focused on *process* rather than product, on *helping* students rather than ranking them. So, as a program, we tend to concentrate our efforts on building supportive classroom communities; sponsoring rich reading and writing activities; providing students with useful feedback for revision; etc. As teachers, however, we are required to *evaluate* our students' work at least once every semester: in the assignment of final grades, which must come from the traditional system of ranked letters (A, B, C, D, & F) (http://www.umass.edu/registrar/gen_info/grading_system.htm). Unfortunately, we sometimes treat that process as our exclusive and solitary purview, to be conducted in private and after-the-fact. Or we reduce it to behavior monitoring, with grades awarded solely on the basis of effort. The fact is: we don't talk very much about how, as *readers*, we *value* our students' writing and how we communicate that *valuation* to others.

Even without talking about it very much, of course, teachers in our program usually end up developing thoughtful and responsible ways of assigning value to their students' work. But in the process, they are often isolated from other teachers at the very moment in their professional lives when they most need support and community. As for our students, they are often simply bewildered by our evaluation practices; even when they accept our commitment to process and follow conscientiously our instructions to revise, it's clear that they're not always sure what all that process is *for*, where it's supposed to lead, *why* they should keep revising besides the fact that we've asked them to. We sometimes lack, in other words, a shared language for talking effectively, both with our students and each other, about what we really mean by "good" writing, even when that is seen from a radically contextualized point of view.

There are, of course, many ways that a program like ours might work with teachers on the evaluation of student writing. Take, for example, the four methods of talking with students about "good" writing that George Hillocks discusses in *Research on Written Composition*: 1) the study of models, 2) the use of scales, 3) teacher comment, and 4) feedback and revision. It could be argued that we do a pretty good job with the third and fourth of these: helping TOs learn to effectively comment on students' writing, both to let them know how well they're doing but also to prompt revisions that could improve their work. We also do a pretty good job providing students with models of

good writing (through, for ex., the *Student Anthology* and the *Text Wrestling Book*), though we could probably do more to help our students *use* those models.

That leaves **scales** – for Hillocks the pedagogical method associated with the most impressive effects on students' writing, at least according to his analysis of the research. Hillocks shows, for example, how *specific* scales (or "rubrics") indexed to short models of effective performance can be profitably taught to students as guides for revision in particular writing contexts.

General criteria for what teachers mean by "good" writing can also, of course, be developed. Probably the best known general criteria-based system for evaluating student writing is the set of five factors (ideas, form, flavor, mechanics, and wording) that Paul Diederich and his colleagues developed for the Educational Testing Service in the early 1960s. Those criteria became the basis for the Northwest Regional Educational Laboratory's "6+1" scale, now used extensively across the country in K-12 education (see <http://www.nwrel.org/assessment/pdfRubrics/6plus1traits.PDF>). Its factors are:

- quality of ideas
- organization
- voice
- word choice
- sentence fluency
- conventions, and
- presentation.

From such criteria, grading rubrics can be devised that link student performance to a numbered or lettered scale (e.g., 6-1 or A-F), which can then be learned by teachers and students alike to guide revision and evaluate finished products.

These kinds of systems are not confined to elementary and secondary education, of course. Many *university* writing programs use rubrics or scales as a way to guide evaluation. Teachers of Penn State's English 15, for example, use a rubric for grading student writing (http://www.courses.psu.edu/engl/engl015_cjg6/#UnivPolicies), as do teachers of business communication at UMass' School of Management (see http://www.isenberg.umass.edu/businesscommunication/Jr_Year_Writing_Course_Syllabus/Grading/.) As for the Writing Program, we use a 6-point scale for scoring our entering students' placement exams. There is no official rubric for grading student papers in *Basic* or *College Writing* courses, though some teachers have developed such scales on their own: Prof. Haivan Hoang's rubric for grading Unit II essays in Englwrit 112, for example, is featured in the Writing Program's *Sourcebook for Instructors of Junior Year Writing Courses* (see <http://www.umass.edu/writingprogram/jy/index.html>). Thoughtfully designed and used, such rubrics can be important tools in our work: they're often admirably simple, clear, and brief; and both teachers and students can be taught to use them to achieve greater uniformity and transparency in evaluation.

But rubrics come with a cost. They tend to be developed top-down on the basis of abstract curricular goals rather than bottom-up on the basis of close attention to actual student work; they tend to perpetuate traditional "teacher talk" about writing (clarity, organization, voice, etc.), talk that has, in some instances, become largely meaningless; they imply a kind of simplicity in and consensus about the reading and evaluation of

student work that can be misleading and even dangerous; and they sometimes bear little or no relationship to what teachers actually do in the classroom and how they actually read students' work.

With those problems in mind, some of us began last year an experiment, inspired by Bob Broad's *What We Really Value: Beyond Rubrics in Teaching and Assessing Writing*, to develop alternative ways of thinking and talking *as a program* about the evaluation of student writing. The idea was to have actual teachers read actual student essays, selected at random with no predetermined sorting in mind, and talk together about them as honestly, comprehensively, and specifically as possible, not with the goal of coming to agreement about them but to "discover, negotiate, and publish the truth about the evaluative topography" of our curriculum. We wanted, that is, not "to turn away from complexity and dissent in judgments" (128) but to actually begin with such complexity and see where it could lead us as writing teachers.

So, last spring (2007), we compiled a collection of eight anonymous student papers from *College Writing's* Unit I assignment, elicited from diverse sections of the course and arranged randomly in a small booklet. In the fall of 2007, we then used that collection with both new and veteran TOs as a way to generate talk about what, as readers and teachers, we valued in Unit I papers. CD groups were then encouraged to continue those discussions in smaller settings, by, for example, sponsoring "trios" of TOs who could grade together "live" papers from their own classes. And at least one group of veteran TOs (English 698G) took the process a step further and, using Broad's book, tried to develop a list of criteria for evaluating Unit I essays based on their discussion of sample papers.

That list of criteria follows on the next page. We present it here as a work-in-progress, the tentative results of one group's semester-long struggle to negotiate together what we mean by a "good" Unit I final draft. We ended up with neither a typical rubric nor a Bob Broad-like "dynamic criteria map" but rather a one-page statement of shared values for reading essays in this particular context. We don't claim in any way that this solves the problem of grading student writing; but we think that it could provide some teachers and some students with a shared language for talking about the work of this one assignment in this one course.

The next questions to ask, of course, are these: how should this statement be used? to help teachers in the program better articulate to themselves and others what they're looking for in these papers? to help new teachers actually evaluate their first set of final student drafts? or to help students themselves, perhaps as a handout to be discussed at some point in Unit I? And finally, the biggest question of all: will such a statement of criteria actually improve student writing? The answers to those questions, as well as the development of statements for other units in the course, remain to be seen.

What Teachers of *College Writing* Value in Students' Unit I Essays

During the fall of 2007, the 14 veteran *College Writing* instructors enrolled in English 698G (*Composition Theories & Pedagogies*) read and talked about sample student papers from Unit I, recording what, as readers, we liked in those papers, what we noticed and remembered, what impressed, pleased, entertained, taught, and moved us; and what, in addition, we found flat, ineffective, unsatisfying, or frustrating. We paid close attention to what issues came up most frequently, what generated the most discussion among us, what language we resorted to again and again to talk about reading these essays. In the end, we identified four broad *qualities* of writing that seemed to drive our estimation of Unit I papers.

We present these qualities here as neither a formula for writing nor a recipe for grading. At best, they're *signposts* for how this one group of teachers read these papers, what *we* noticed and remarked on. Certainly, they could be used as *prompts* for revision or *criteria* for evaluation; but they can't do those things mechanically. For one thing, these qualities are neither fully discrete nor always compatible. For another, they are largely meaningless outside the context of this assignment and this curriculum. And, finally, they probably only make sense in conjunction with actual student papers, the qualities guiding discussion of the essays, the essays helping readers understand the qualities. One additional note: for each quality, we have supplied a headword that we believe encapsulates it, but we didn't always agree on the headwords, and we recognize that emphasizing them too much is risky. In any case, here's our statement:

Readers in our group especially prized Unit I papers that exhibited:

- **detail:** we valued essays that contained vivid imagery and concrete detail: that didn't just glide over the surface of things but found telling particulars, that showed rather than told, that propelled the reader forward, evoking something real in the world: actual people, places, and objects. For us, depth of exploration, therefore, was preferred over quantity of sub-topics: one deeply remembered and vividly recounted narrative, or an extended description of something stretched over several paragraphs, worked better for us than the superficial treatment of multiple items or vague generalizations.
- **surprise:** we valued essays that stuck in the mind, essays with a distinctive voice or unique style, essays which took risks (artistic, intellectual, etc.), even when they didn't fully succeed. We exhibited as readers, in other words, a wariness about cliché and easy sentiment. We liked reading uncommon, unexpected, memorable things, in which the extraordinary peeked out from behind the ordinary.
- **reach:** we valued essays written out of intense personal engagement with the subject matter, that seemed genuinely important to the writer and not just a response to an assignment, that made it clear that something was at stake in writing the essay. Such writing usually revolved around and revealed growth of some sort: self-discovery, change, or at least an awareness of their need. We appreciated essays, in other words, that *pushed* towards understanding, that connected the author's experiences to something bigger than him- or herself. But we weren't looking for simple life lessons or mottos: we liked essays that dealt with complicated issues, that faced complexity and contradiction and did so unflinchingly. In fact, confronting difficulty was often valued more than neat resolutions, happy endings, or simplistic moralizing.
- **care:** finally, we valued essays that showed conscious and considered care for and control over language: appropriate tone, diction, paragraphing, sentences, etc.