Pat Belanoff

THE MYTHS OF ASSESSMENT

ABSTRACT: *This article challenges several myths about writing assessment: that we know what we're grading, that we know what the results mean, that we can agree in practice on the relative weight of various criteria, and that it is possible to establish absolute standards and apply them uniformly. Despite these seeming difficulties, this article argues for the validity of assessments developed within particular environments for particular purposes agreed to by those teaching within those environments. And, finally, the article celebrates the lack of conformity in grading as a sign of a rich and nurturing environment for the development of writing skills.*

Back when I started to teach writing, my first students were mostly middle and upper middle class White kids. What I was learning at the time about the teaching of writing, the theories behind various approaches, and the supporting philosophies, I was applying to a fairly privileged group of students and was gratified by the results. When I moved from teaching that group and began to teach at the Borough of Manhattan Community College and became familiar with the work of Mina Shaughnessy, Marie Ponsot, Rosemary Dean, and others, I discovered that what I had learned about teaching writing continued to apply in classrooms of so-called basic writers and somewhat advanced ESL students. I didn't realize that immediately. I thought I needed to teach basic writers and ESL students lots of grammar and how to write sentences so someday

---

Pat Belanoff is director of Writing Programs at State University of New York (SUNY) at Stony Brook. She is coauthor (with Peter Elbow) of A Community of Writers *and coeditor (with Peter Elbow and Sheryl Fontaine) of the just published* Nothing Begins With N: New Investigations of Freewriting. *Currently, she is coediting (with Marcia Dickson) a collection of essays* Portfolio Grading: Process and Product *being published by Heinemann Boynton/Cook in June 1991. She has also published articles on the women of Old English poetry. The above paper was the keynote address presented at the CUNY ESL Conference in March, 1990.*

DOI: 10.37514/JBW-J.1991.10.1.06

they could write paragraphs, and then compositions, and maybe some day even discourses. I discovered how wrong I was. I often believe that the students of BMCC taught me more than I taught them.

When I moved to Stony Brook and began to teach less advanced ESL students and lower middle class and working class students, I discovered again that their needs were not so different from the needs of my previous students. Learning to write is learning to write—what works for advanced students also works for ESL students. Even in beginning language courses, students use language to think within restricted contexts and need to think in order to learn. To quote Janet K. Swaffar in *Profession 89*, "The notion that thinking and intentionality were integral to language use at any level made viable a claim heretofore rejected out of hand: that language learning need not be remedial learning. If taught in terms of creativity rather than replication, even beginners could find language learning an intellectually challenging activity, a bona fide academic enterprise."

All of us have been accused of doing remedial work, even those of us teaching advanced composition; a recent survey of faculty at Stony Brook makes that conception of our work painfully obvious. We need to argue that point constantly to our colleagues in other fields. Nor can we exclude our writing center colleagues and say they are in charge of grammar and mechanics, and classroom teachers deal with "ideas"—as though they were separate. This is a common dichotomy, but we're all teachers and we're all tutors—certainly the best classroom teachers I've known are tutors.

What we need to argue within our field and to each other is equally important: that all of us engaged in the teaching of writing—regardless of the names given to the courses we teach—are working within the same paradigm and have much to learn from each other once we recognize the commonality of our pursuits. We all need to talk to each other more often.

I've entitled my talk today the myths of assessment and plan to speak generally about four myths:

1) We know what we're testing for
2) We know what we're testing
3) Once we've agreed on criteria, we can agree on whether individual papers meet those criteria
4) And the strongest myth of all, that it's possible to have an absolute standard and apply it uniformly

First myth: we know what we're testing for. Let's think about the writing tests we're connected to in some way—tests we give in our

own classrooms and standardized tests administered apart from our classrooms.

What are those standardized tests testing for? Are they finding out, as the CUNY test supposedly does, that students have achieved a certain level and are ready to go on to another level—where they may or may not get more help with their writing? Does that mean that students in supposedly below-level classes cannot apply what skills they have to writing about economics or literature or whatever their other classes will ask them to write about? Our portfolio proficiency test at Stony Brook certifies that students have satisfied the first level of our writing requirement; what it really means is that students do not have to retake our basic composition course—in truth, what it does for far too many students is assure that they won't be asked to write again for a couple of years—or until they have to satisfy their upper-level writing requirement.

What does the National Association for Educational Progress' writing sample measure? How well students can write to demonstrate they can write? And what about New York State's minimum competency tests in writing? They demonstrate that students can or cannot reorganize a list of things and write up a report. I'm not saying those things aren't measurable or shouldn't be measured—but once they've been measured, what can we say about a student's skill as a writer? We overgeneralize about all these results.

Some standardized tests purport to say that students write well enough to be allowed to graduate from college. How well is that? How well *should* a college graduate write? And why do we need separate writing requirements? If a college degree doesn't certify literacy, what does it certify?

Well, perhaps we are testing to see if students are improving (I think now of pretests and posttests used to evaluate either programs or students or both), how much can students genuinely improve in one semester and can we measure the ways in which they improve? A lot of what we want to teach them is subsumed under attitudes and approaches and how do we test for that? We want them to take risks, to try harder things which may make their writing look as though it's deteriorating depending on when we decide to look at it. We don't want them to write what they already know how to write; we want them to write something that pulls and stretches their skills—and that pulling and stretching can result in some pretty messy stuff.

And what about the testing—formal and informal—in our classrooms? What are our purposes? To see if students have mastered a particular skill? To see if students write better than they did three weeks ago or three months ago? Do we need tests to know

that? What if students haven't mastered the skill or don't write better, have we failed? Have they failed? Is growth steady or does it come in spurts?

The assumption here is that we have some precise notion of what skills students need to master in order to be good or better writers and that we know in what order these should be learned: word forms before paragraphs, narrative before argument, etc., or vice versa. Unfortunately, the skills which are easiest to measure are the ones least important to the development of good writing. We can determine with some degree of accuracy and agreement from others whether or not a word form is correct or whether an essay has a topic sentence or whether all sentences end with the proper punctuation marks. But we can't agree so easily that the word or the topic sentence selected is effective stylistically and rhetorically and whether the groups of words ending with periods communicate some idea clearly and effectively and integrate that idea into what comes before and after. We won't agree about the latter to the degree we agree about the former. We can't agree on something as seemingly concrete as where commas go. Rules, after all, are abstractions, humanmade—they're not real. As abstractions, they do not reflect any reality exactly. Consequently, rules are only clear until we apply them—then they fuzzy up. But, more importantly, we cannot separate rhetorical issues and issues of correctness even though textbooks and handbooks purport to do it all the time.

So we just don't know whether what we test in class makes for good writing or not, and if students improve whether they become better writers. In fact, we really can't isolate skills and judge them separately from the entire act of communication because it is that act that sets the perimeters for us and for them, and it is that act against which we have to measure whatever students do.

Well, that's my first myth: that we know what we are testing for. My second myth is that we know what we are testing. What we're testing is the student's writing ability, correct? And how do we do that? By looking at some piece of writing the student has done in 20 minutes or an hour? To what degree does a particular piece of writing represent a student's total ability? Are we assessing the student's ability or the quality of the piece of writing? In fact the only thing it's really possible to find out is if the particular piece of writing before us does or does not accomplish some particular purpose. Could the student duplicate the piece, do something else like it just as well again? And even if so, can writing tasks be so much alike that we can be sure that if a student does one he can do the other? Or that someone will even ask him to do this thing again some day? But given the nature of most of our tests, I suspect no one

will ask the student to do quite this same thing again. In fact, are we grading a piece of writing in any meaningful sense at all? Under what circumstances would a student ever be asked to do this thing we've asked him to do on the test?

Other issues are relevant too. Did the student struggle to write this? Was it easy for her? Was she feeling well, poorly, hostile? What, in fact, does this piece of writing in front of us represent?

Listening to Muriel Harris[1] this morning as she spoke of the role of writing centers in relation to the increasing cultural diversity of colleges made me realize with even greater intensity how ludicrous it is to use a single instrument to measure writing competency. I would add another diversity to her provocative list of cultural diversities. What does it mean to write as a woman in a profession so long dominated by western male standards of performance derived from classical rhetoric?

And what is writing ability anyway? What does it mean to write well? Is a good writer someone who can write anything? Is a good writer someone who can fulfill a school assignment? Is being able to record one's thoughts in a diary, write a letter to a friend, write a poem—are these things a good writer can and should do; is one a good writer if one can do them? As Ed White points out in *Teaching and Assessing Writing* (and he's a proponent of assessment), our profession has no agreed upon definition of proficiency and certainly as a consequence, no agreed upon definitions for proficiencies at various levels of schooling.

So that's my second myth: that we know what we're testing. My third one is that even if we know what we're testing for and what the artifact is in front of us, we still don't agree on how well the student has achieved the goals. In truth we don't always agree on which characteristics of a good piece of writing are most significant in making us judge the piece positively.

I've often sat with groups of teachers and worked out what we could agree on as the traits of a good piece of writing—they'll come out something like clarity, effective organization, contextual awareness, coherence, correctness of language, and so on; probably the same set of traits any group of good teachers would come up with. In the abstract, they sound fine. The problem comes when we get around to applying them to actual papers. What I think is clear, someone else doesn't. What I see as well-organized, another doesn't. Or I value the work because it's well-organized and another reader agrees, but thinks the good organization is overshadowed by superficiality of content. Modern critical theory points to something we've always known—that people don't read in the same way—that, as a result, texts do not *embed* meaning, they *enable* meaning.

Subjects affect us; our acquaintance with a variety of forms affects us; the authority we're willing to grant to authors and to our own right to judge affects us—we can't really codify what goes into the interpretation of a particular text, we can't even be sure that we would assess the same text the same way a second time.

We can, of course, be trained by the Educational Testing Service (ETS), or through similar methods, to agree on texts—agree on numbers we would assign to particular texts. ETS is wise not to insist on expressions of why a grader awards a particular score to a particular paper. In the process of "training" (a form of brainwashing for sure), a grader learns the community standards and learns to apply them quite well, but never questions their validity for the task they set themselves up to do. They're not asked to.

But most of us simply don't want our students to be subjected to such an inhumane process. We rightly insist that writing is not genuinely writing if it degenerates into a performance whose content is irrelevant. We need to beware of valuing some scheme simply because it produces interreader reliability. Reliability is high, but what does a 3 or 4 really mean in any context outside the room where the scoring is occurring? No question that it means quite a lot to the students who have taken the test—it places them in a level of college composition or it increases or decreases their scores on tests such as the National Teacher's Examination.

But how well should a beginning teacher be able to write? And what does the NTE test itself suggest to new teachers about the role of writing in their own classrooms? What kinds of things will they ask their students to do as a result? As we are tested, so we will test others. Frankly, I'd rather test a teacher's ability to get students enthused about writing—that, of course, includes getting the teacher enthused about her own writing. I'd also like that new teacher to know something of current theories about the teaching of writing if only as an indication that all methods of teaching writing assume certain things about language and about learning in general; all methods of teaching writing, that is, are philosophically based, whether we recognize the basis or not. But, of course, the writing test she has just taken invalidates those theories I want her to know.

In addition, this sort of brainwashing, holistic testing, and grading separates the graders from the testmakers and often separates the latter from those who devise the standards for admittance into a particular profession. Graders are protected from the consequences of their grading, and teachers are isolated from judgments of students they have taught. Furthermore, new teachers are pragmatically taught something quite undesirable about writing.

So, this is my third myth: that we know what good writing is and

that, in meaningful contexts, we can agree when we apply those standards to pieces of writing. Students have always known we don't agree. They tell us over and over again that a former teacher or their roommate's teacher would have given them a different grade (usually higher of course), although in their more honest exchanges, they'd also admit that some prior teacher would have given them a lower grade. Though they may be exaggerating the size of the differences, they're not wrong in principle. Such disagreements exist all over. I've had the same article (revised each time of course) rejected by *College Composition and Communication* three times. One reader has been fairly consistent in his or her comments; I'd love to sit with that reader and discuss the issues I want to raise. But the other readers tell me disparate things. One thinks my subject is strongly significant within the profession; another considers it only somewhat significant. One thinks the personal references enrich my piece; another thinks they make the style rough and uneven. The truth is that for all sorts of reasons, readers don't agree on texts. We may be judging at different levels (unskilled, skilled, professional) but there's no more agreement at one level than at the others. It's no easier to determine a student's readiness for regular composition than it is to certify graduate level competency or a paper's suitability for publication.

This brings me to the last and most harmful of my myths: the myth that there is some Platonic image out there of "good writing" and that there is as a result a Platonic standard of writing which we can all learn to apply uniformly. Within this myth, the problem is only that we haven't yet discovered this absolute standard, but if we keep working at it, we will find it some day.

But there is no such Platonic ideal—there are only lots and lots of real texts around us in our world, some of which we have to judge because they're written by our students within an educational system which says we have to judge them. But, in real-world reading, we always judge for a reason, within a context, according to the purposes a writer sets up. Thus, the only decisions we *can* make are contextual. Over and over at ETS grading sessions, I've heard graders say that they know that some paper they've scored gets a 3 by the standards we've been asked to adopt, but that they'd never "in the real world" give it that high a rating. By "real world" I assume they mean the usual context in which they grade.

We all judge holistically, despite the fact that we can then find reasons for our judgments. We judge first and then articulate our reasons. The rhetoricians Perelman and Olbrechts-Tyteca believe this to be true of all our decisions. Thus we react to discourses as a whole and not to parts of discourses in isolation. And because we

judge on the basis of whole discourses, we inevitably take into consideration at the conscious and unconscious level an integration of all the traits of a piece of writing. We don't judge on the basis of one or two of these—we judge on the basis of the whole which is always greater than its parts. This is not to deny that within my holistic judgments I don't value one trait more than someone else might.

So, what am I saying—that we can't judge at all and should just give it up altogether? Well, in one very real sense, I'd love to say that. Many of you here have been talking about writing centers and what goes on and can go on in them. One of the wonderful things about being a writing center tutor is that one doesn't have to give grades: one's function is simply to help students become better writers—usually through talk and revision and feedback and such, not through grades. "This is what you've done well, do more of it." "This is what doesn't work well for me because I don't see its relation to your main point; can you do something to help me with this?" These are the sorts of things we can say and do when we're writing center tutors. And, most importantly, we can through our talk and feedback begin to direct students toward becoming evaluators of their own texts—at least to the point of understanding where they may need to think about doing some more work. Evaluation and feedback merge. Almost everyone I know who has moved from the classroom into a writing center loves the emancipation from grading and finds it stimulates whole new ways of looking at and commenting on students' texts. We don't like grading.

Think about it. Have you ever noticed that you can find lots of articles on assessment and evaluation, but how many articles have you read or seen published on grading—on the actual giving of grades? Not very many. Most of us would just rather not talk about it at all; it's the dirty thing we have to do in the dark of our own offices. We can spend lots of time talking about teaching writing and encouraging students to like writing—to find subjects they can relate to, to find ways of dealing with subjects they have trouble relating to, to give and receive feedback, to work on revision, and so forth. We love to talk about those things to each other; we don't love to talk about grading and we do very little of such talking, though we're likely to moan and groan about it.

But, modern society and the structure of modern educational institutions are simply not going to let us not deal with the issue. We are stuck giving grades and administering standardized tests. But are we?

There is a movement afoot in elementary and secondary schools

to give teachers more say in the running of schools and in the make-up of curricula. Finally, there seems to be developing some institutional awareness of the value of a classroom teacher's knowledge. That movement needs to move into college writing classrooms so that what we know will be given as much credit as almost all other college faculty's knowledge. Who's checking up on their standards? I know there are lots of bad teachers out there—I've had them; you've had them; my kids have had them; I have some in my department. But I see little reason to build systems as ways to subvert bad teachers; we need to build systems that release the strengths of good teachers. We need to take more of a hand in our own fates. What are some ideas we can build on?

First, we need to realize that our inability to agree on standards and their applications is not something we need to be ashamed of—something to hang our heads and wring our hands about in the presence of our colleagues in the sciences or other disciplines (even including our literature colleagues at times) who have "content" to test. Our inability is no sign of weakness—far from it, it is a sign of strength, of the life and vitality of words and the exchange of words. For, if we agreed, we could set up hierarchies and fit ourselves and others into them and then all could dictate to those below them and follow the orders of those above them. And in fact, in such a set up there would have to be an autocrat at the top who knows what's best for us and who knows what texts are best. Then someone would know what sort of texts to write and to teach and the variety would leave our profession and along with the variety, the richness.

Texts reflect life and the multitude of tastes and standards in real life; it is for that reason that we're motivated to create them, as expressions of our place in a multifaceted world. We've learned that texts have a peculiar strength, a peculiar ability to make us feel ourselves and the uniqueness of those selves.

Colleagues in other disciplines can tsk-tsk-tsk at our subjectivity because theirs is so well hidden. Do introductory biology teachers agree on what should be taught, what should be tested, and how tests should be balanced and averaged into the final grade? I doubt it. Have all introductory biology teachers in CUNY gotten together recently to discuss these issues? I doubt that too. When I'm not teaching composition, I'm teaching introductory Old English. It will come as no surprise to you to know that no one else is teaching it at the same time I am. I determine what to teach and when, what to test and when, and what elements to figure into my final grade. If there were 39 other sections of introductory Old English, I'd bet we'd be called to a meeting one day to talk about how to measure

competency in Old English and how to determine when students should move on to the next semester. And someone would come up with a standardized test just to make certain I was indeed teaching my students what they needed to know about Old English language and literature.

Let's not apologize for our lack of agreement—let's make it work for us. How can we do that?

Well, I've certainly cast much doubt on our ability to agree on standards, but I've never cast doubt on our ability to *have* them. Each of us does have his or her standards: we read a text and we judge it almost as a reflex action, the judgment usually growing out of whether or not we like the text. Each of us also has the ability, enhanced when we talk with others, of figuring out the basis of the judgments we make. We can learn to articulate that basis for ourselves, for our students, and for other teachers. Frankly if we can't, we shouldn't be teachers of writing. Our judgments are the result of a number of factors—what we've read, what our values are, what our philosophy is, who our colleagues are, what our own education has been, and for many of us, years and years of reading and responding to student papers. Whatever those factors are, they feed into our judgments. Thus, there is a kind of individual validity of judgment which arises from our well-trained and experienced response to all sorts of texts, including student texts. In a very real sense, no one else can "disprove" my response and judgment of a text.

But there is another kind of validity of judgment which can come from the pooling of individual judgments in the process of discussion of specific papers about which decisions need to be made for reasons we all know within a context we all share, a validity quite different from ETS readings. The more we participate in such collaborative decision making, the more we become a community—a community which exists in a very specific time and place and for a very specific purpose within that time and place.

This is in fact what we do in our portfolio system at Stony Brook. A passing portfolio is what students need in order to satisfy the first level of the writing requirement at Stony Brook. A portfolio passes if at least two teachers agree that it is passing. The judgment is holistic in terms of the whole portfolio.

So, what I am saying is that there are two sorts of valid judgments—the totally personal and the communal—but it has to be a community which is engaged in conversation about teaching and standards all the time, not just during grading sessions and not in the abstract. These discussions always have to be tied to actual student papers, and they need to include the student's teacher and be based on a range of work.

Additionally, and perhaps paradoxically, I want to argue for the importance and benefit of evaluation. The more we talk about evaluation with our colleagues, the better we'll become at giving feedback to our students on their writing and the better we'll be able to guide our students into making their own evaluations of all sorts of texts, including their own. James Moffett wisely reminds us that the more talk we hear, the more our own voices are likely to be individualized, and yet remain solidly embedded in the language which provides the basis for communication. The same is true of evaluation. The more we engage in talk with students and colleagues about our reactions to texts, the more we're able to construct individual evaluations firmly embedded in our communities. Ultimately our students also need to learn that, to understand the variety of ways a particular community will respond to their texts. This understanding will open the doors to the revision and improvement of texts based on context and purpose and personal intentions. Without some internalization of our voices and through our voices an internalization of the voices of our community, students will not be able to become good editors and revisers of their own writing.

And so, outsider as I am, I'd like to propose something fairly radical to you, all the while recognizing that any evaluation system needs to grow from the strengths and initiation of individual teachers; it cannot be imposed from above—the standards must come from within the group and be constantly open to alteration and transmutation. My suggestions are meant to start a conversation.

Here's my suggestion: Conduct your classes as you always do, getting students to collect all their work, formal and informal writings, graded and ungraded, journals, whatever you ask them to write, but including I hope some writing about their own writing. Many of you undoubtedly already do this. Two weeks before the end of the term ask your students to look through their own folders and write a letter summarizing the contents, the sorts of processes involved in producing those contents, including also some analysis of the strengths and weaknesses of the writing, concluding with their estimates of whether they should pass on to another level. You will then meet with each student (or perhaps only with those selected because their status is problematic) and discuss their evaluations of themselves and your evaluations of them. In the final week (or during the time normally spent scoring standardized tests) meet with a group of your colleagues and discuss the following specific folders:

1) A folder you are certain passes
2) A folder you are certain fails
3) All folders on the borderline

Whatever decisions the group makes would be final. Even if teachers, particularly adjuncts, were paid extra for these group sessions, the overall cost would have to be less than all it costs to prepare, administer, and store thousands of standardized tests every year.

The two questions I'm most often asked about portfolio grading are 1) doesn't it take loads of time and 2) how reliable is it? No, it really doesn't take loads of time because it usually demonstrates that something else we're spending lots of time on doesn't warrant that time. As for the second question, how valid (not reliable) is what's currently being done? Because you and I agree on a score doesn't mean that the student whose paper we are reading is necessarily the writer we say she or he is. What's more, Roberta Camp of ETS has a delightful little table that I love to show skeptics; it's a statistical study which demonstrates that the more people who read a particular set of papers, the more genres or modes there are in the set, and the more examples of each genre or mode there are, the higher the reliability—that is, the more likely it will be that evaluators will agree on their evaluations. This is the closest we can come to making judgments about a writer; everyone's running about trying to make a difficult job easy. Was the Nobel Prize for Literature ever given to a writer who produced just one book?

You cannot, of course, adopt my plan because it's *my* plan, not yours. But you can come together with like-minded colleagues and begin to try some things—things that don't bastardize what you teach in your classrooms. Through trial and error, you'll find a way if you continually remind yourselves that evaluation of writing cannot and should not be removed from those contexts which alone provide the possibility for meaningful and useful evaluation. We cannot continue to allow others to tell us how to do the job we know best how to do. But if we don't step in, speak up, develop strategies, others, including state legislators, will gain greater and greater influence over our classrooms. They will be making those decisions which it is our responsibility to make in ways consonant with what we have learned and are continuing to learn about language and the teaching of writing. If you work together, I'm confident you can find ways to evaluate your students' writing fairly for whatever purposes you need, and thus do your job better. If you do work together and pool your knowledge, experience, and commitment to your students and your work, you will come up with something better. Then I can

hope with a great deal of confidence that by the year 2000 the CUNY Writing Assessment Test, as we know it (and its clones throughout the country) will no longer exist.

Good luck.

Note

[1] Muriel Harris spoke at The City University of New York's ESL conference in March, 1990.

Works Cited

Camp, Roberta. "The Writing Folder in Post-Secondary Assessment." *Directions and Misdirections in English Evaluation.* Ed. Peter J. A. Evans. Canadian Council of Teachers of English, 1985.

Perelman, Chaim, and L. Olbrechts-Tyteca. *The New Rhetoric: A Treatise on Argumentation.* Trans. John Wilkinson and Purcell Weaver. Notre Dame, IN: U of Notre Dame P, 1969.

Ponsot, Marie and Rosemary Dean. *Beat Not the Poor Desk.* Upper Montclair, NJ: Boynton/Cook, 1982.

Shaughnessy, Mina P. *Errors and Expectations.* New York: Oxford UP, 1977.

Swaffar, Janet K. "Curricular Issues and Language Research: The Shifting Interaction." *Profession 89:* 32–38.

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