

RANKING, EVALUATING, AND LIKING:
SORTING OUT THREE FORMS OF
JUDGMENT

Peter Elbow

This essay is my attempt to sort out different acts we call assessment—some different ways in which we express or frame our judgments of value. I have been working on this tangle not just because it is interesting and important in itself but because assessment tends so much to drive and control *teaching*. Much of what we do in the classroom is determined by the assessment structures we work under.

Assessment is a large and technical area and I'm not a professional. But my main premise or subtext in this essay is that we nonprofessionals can and should work on it because professionals have not reached definitive conclusions about the problem of how to assess writing (or anything else, I'd say). Also, decisions about assessment are often made by people even less professional than we, namely legislators. Pat Belanoff and I realized that the field of assessment was open when we saw the harmful effects of a writing proficiency exam at Stony Brook and worked out a collaborative portfolio assessment system in its place (Belanoff and Elbow; Elbow and Belanoff). Professionals keep changing their minds about large-scale testing and assessment. And as for classroom grading, psychometricians provide little support or defense of it.

THE PROBLEMS WITH RANKING AND THE
BENEFITS OF EVALUATING

By ranking I mean the act of summing up one's judgment of a performance or person into a single, holistic number or score. We rank every time we give a grade or holistic score. Ranking implies a single scale or continuum or dimension along which all performances are hung.

Peter Elbow is Professor of English at the University of Massachusetts at Amherst. He has written books about writing and *Oppositions in Chaucer* and *What Is English?* He has taught at diverse institutions and directed the writing program at SUNY-Stony Brook.

By evaluating I mean the act of expressing one's judgment of a performance or person by pointing out the strengths and weaknesses of different features or dimensions. We evaluate every time we write a comment on a paper or have an conversation about its value. Evaluation implies the recognition of different criteria or dimensions—and by implication different contexts and audiences for the same performance. Evaluation requires going *beyond* a first response that may be nothing but a kind of ranking (“I like it” or “This is better than that”), and instead looking carefully enough at the performance or person to make distinctions between parts or features or criteria.

It's obvious, thus, that I am troubled by ranking. But I will resist any temptation to argue that we can get rid of all ranking—or even should. Instead I will try to show how we can have *less* ranking and *more* evaluation in its place.

I see three distinct problems with ranking: it is inaccurate or unreliable; it gives no substantive feedback; and it is harmful to the atmosphere for teaching and learning.

(1) First the unreliability. To rank reliably means to give a *fair* number, to find the single quantitative score that readers will agree on. But readers don't agree.

This is not news—this unavailability of agreement. We have long seen it on many fronts. For example, research in evaluation has shown many times that if we give a paper to a set of readers, those readers tend to give it the full range of grades (Diederich). I've recently come across new research to this effect—new to me because it was published in 1912. The investigators carefully showed how high school English teachers gave different grades to the same paper. In response to criticism that this was a local problem in English, they went on the next year to discover an even greater variation among grades given by high school geometry teachers and history teachers to papers in their subjects. (See the summary of Daniel Starch and Edward Elliott's 1913 *School Review* articles in Kirschenbaum, Simon, and Napier 258–59.)

We know the same thing from literary criticism and theory. If the best critics can't agree about what a text means, how can we be surprised that they disagree even more about the quality or value of texts? And we know that nothing in literary or philosophical theory gives us any agreed-upon rules for settling such disputes.

Students have shown us the same inconsistency with their own controlled experiments of handing the same paper to different teachers and getting different grades. This helps explain why we hate it so when students ask us their favorite question, “What do you want for an A?”: it rubs our noses in the unreliability of our grades.

Of course champions of holistic scoring argue that they get *can* get agreement among readers—and they often do (White). But they get that agreement by “training” the readers before and during the scoring sessions. What “training”

means is getting those scorers to stop reading the way they normally read—getting them to stop using the conflicting criteria and standards they normally use outside the scoring sessions. (In an impressive and powerful book, Barbara Herrnstein Smith argues that whenever we have widespread inter-reader reliability, we have reason to suspect that difference has been suppressed and homogeneity imposed—almost always at the expense of certain groups.) In short, the reliability in holistic scoring is not a measure of how texts are valued by real readers in natural settings, but only of how they are valued in artificial settings with imposed agreements.

Defenders of holistic scoring might reply (as one anonymous reviewer did), that holistic scores are not perfect or absolutely objective readings but just “judgments that most readers will agree are the appropriate ones given the purpose of the assessment and the system of communication.” But I have been in and even conducted enough holistic scoring sessions to know that even that degree of agreement doesn’t occur unless “purpose” and “appropriateness” are defined to mean acceptance of the single set of standards imposed on that session. We know too much about the differences among readers and the highly variable nature of the reading process. Supposing we get readings only from academics, or only from people in English, or only from respected critics, or only from respected writing programs, or only from feminists, or only from sound readers of my tribe (white, male, middle-class, full professors between the ages of fifty and sixty). *We still* don’t get agreement. We can sometimes get agreement among readers from some subset, a particular community that has developed a strong set of common values, perhaps *one* English department or *one* writing program. But what is the value of such a rare agreement? It tells us nothing about how readers from other English departments or writing programs will judge—much less how readers from other domains will judge.

(From the opposite ideological direction, some skeptics might object to my skeptical train of thought: “So what else is new?” they might reply. “Of *course* my grades are biased, ‘interested’ or ‘situated’—always partial to my interests or the values of my community or culture. There’s no other possibility.” But how can people consent to give grades if they feel that way? A single teacher’s grade for a student is liable to have substantial consequences—for example on eligibility for a scholarship or a job or entrance into professional school. In grading, surely we must not take anything less than genuine fairness as our goal.)

It won’t be long before we see these issues argued in a court of law, when a student who has been disqualified from playing on a team or rejected from a professional school sues, charging that the basis for his plight—teacher grades—is not reliable. I wonder if lawyers will be able to make our grades stick.

(2) Ranking or grading is woefully uncommunicative. Grades and holistic scores are nothing but points on a continuum from “yea” to “boo”—with no

information or clues about the criteria behind these noises. They are 100 percent evaluation and 0 percent description or information. They quantify the degree of approval or disapproval in readers but tell nothing at all about what the readers actually approve or disapprove of. They say nothing that couldn't be said with gold stars or black marks or smiley-faces. Of course our first reactions are often nothing but global holistic feelings of approval or disapproval, but we need a system for communicating our judgments that nudges us to move beyond these holistic feelings and to articulate the basis of our feeling—a process that often leads us to change our feeling. (Holistic scoring sessions sometimes use rubrics that explain the criteria—though these are rarely passed along to students—and even in these situations, the rubrics fail to fit many papers.) As C. S. Lewis says, “People are obviously far more anxious to express their approval and disapproval of things than to describe them” (7).

(3) Ranking leads students to get so hung up on these oversimple quantitative verdicts that they care more about scores than about learning—more about the grade we put on the paper than about the comment we have written on it. Have you noticed how grading often forces us to write comments to justify our grades?—and how these are often *not* the comment we would make if we were just trying to help the student write better? (“Just try writing several favorable comments on a paper and then giving it a grade of D” [Diederich 21].)

Grades and holistic scores give too much encouragement to those students who score high—making them too apt to think they are already fine—and too little encouragement to those students who do badly. Unsuccessful students often come to doubt their intelligence. But oddly enough, many “A” students also end up doubting their true ability and feeling like frauds—because they have sold out on their own judgment and simply given teachers whatever yields an A. They have too often been rewarded for what they don't really believe in. (Notice that there's more cheating by students who get high grades than by those who get low ones. There would be less incentive to cheat if there were no ranking.)

We might be tempted to put up with the inaccuracy or unfairness of grades if they gave good diagnostic feedback or helped the learning climate; or we might put up with the damage they do to the learning climate if they gave a fair or reliable measure of how skilled or knowledgeable students are. But since they fail dismally on both counts, we are faced with the striking question of why grading has persisted so long.

There must be many reasons. It is obviously easier and quicker to express a global feeling with a single number than to figure out what the strengths and weaknesses are and what one's criteria are. (Though I'm heartened to discover, as I pursue this issue, how troubled teachers are by grading and how difficult they find it.) But perhaps more important, we see around us a deep *hunger to rank*—to create pecking orders: to see who we can look down on and who we must look up

to, or in the military metaphor, who we can kick and who we must salute. Psychologists tell us that this taste for pecking orders or ranking is associated with the authoritarian personality. We see this hunger graphically in the case of IQ scores. It is plain that IQ scoring does not represent a commitment to looking carefully at people's intelligence; when we do that, we see different and frequently uncorrelated *kinds* or *dimensions* of intelligence (Gardner). The persistent use of IQ scores represents the hunger to have a number so that everyone can have a rank. ("Ten!" mutter the guys when they see a pretty woman.)

Because ranking or grading has caused so much discomfort to so many students and teachers, I think we see a lot of confusion about the process. It is hard to think clearly about something that has given so many of us such anxiety and distress. The most notable confusion I notice is the tendency to think that if we renounce ranking or grading, we are renouncing the very possibility of judgment and discrimination—that we are embracing the idea that there is no way to distinguish or talk about the difference between what works well and what works badly.

So the most important point, then, is that *I am not arguing against judgment or evaluation*. I'm just arguing against that crude, oversimple way of *representing* judgment—distorting it, really—into a single number, which means ranking people and performances along a single continuum.

In fact I am arguing *for evaluation*. Evaluation means looking hard and thoughtfully at a piece of writing in order to make distinctions as to the quality of different features or dimensions. For example, the process of evaluation permits us to make the following kinds of statements about a piece of writing:

- The thinking and ideas seem interesting and creative.
- The overall structure or sequence seems confusing.
- The writing is perfectly clear at the level of individual sentences and even paragraphs.
- There is an odd, angry tone of voice that seems unrelated or inappropriate to what the writer is saying.
- Yet this same voice is strong and memorable and makes one listen even if one is irritated.
- There are a fair number of mistakes in grammar or spelling: more than "a sprinkling" but less than "riddled with."

To rank, on the other hand, is to be forced to translate those discriminations into a single number. What grade or holistic score do these judgments add up to? It's likely, by the way, that more readers would agree with those separate, "analytic" statements than would agree on a holistic score.

I've conducted many assessment sessions where we were not trying to impose a set of standards but rather to find out how experienced teachers read and evaluate, and I've had many opportunities to see that good readers give grades or

scores right down through the range of possibilities. Of course good readers sometimes agree—especially on papers that are strikingly good or bad or conventional, but I think I see difference more frequently than agreement when readers really speak up.

The process of evaluation, because it invites us to articulate our criteria and to make distinctions among parts or features or dimensions of a performance, thereby invites us further to acknowledge the main fact about evaluation: that different readers have different priorities, values, and standards.

The conclusion I am drawing, then, in this first train of thought is that we should do less ranking and more evaluation. Instead of using grades or holistic scores—single number verdicts that try to sum up complex performances along only one scale—we should give some kind of written or spoken evaluation that discriminates among criteria and dimensions of the writing—and if possible that takes account of the complex context for writing: who the writer is, what the writer's audience and goals are, who we are as readers and how we read, and how we might differ in our reading from other readers the writer might be addressing.

But how can we put this principle into practice? The pressure for ranking seems implacable. Evaluation takes more time, effort, and money. It seems as though we couldn't get along without scores on writing exams. Most teachers are obliged to give grades at the end of each course. And many students—given that they have become conditioned or even addicted to ranking over the years and must continue to inhabit a ranking culture in most of their courses—will object if we don't put grades on papers. Some students, in the absence of that crude gold star or black mark, may not try hard enough (though how hard is "enough"—and is it really our job to stimulate motivation artificially with grades—and is grading the best source of motivation?).

It is important to note that there are certain schools and colleges that do *not* use single-number grades or scores, and they function successfully. I taught for nine years at Evergreen State College, which uses only written evaluations. This system works fine, even down to getting students accepted into high quality graduate and professional schools.

Nevertheless we have an intractable dilemma: that grading is unfair and counterproductive but that students and institutions tend to want grades. In the face of this dilemma there is a need for creativity and pragmatism. Here are some ways in which I and others use *less ranking* and *more evaluation* in teaching—and they suggest some adjustments in how we score large-scale assessments. What follows is an assortment of experimental compromises—sometimes crude, seldom ideal or utopian—but they help.

(a) Portfolios. Just because conventional institutions oblige us to turn in a single quantitative course grade at the end of every marking period, it doesn't

follow that we need to grade individual papers. Course grades are more trustworthy and less damaging because they are based on so many performances over so many weeks. By avoiding frequent ranking or grading, we make it *somewhat* less likely for students to become addicted to oversimple numerical rankings—to think that evaluation always translates into a simple number—in short, to mistake ranking for evaluation. (I’m not trying to defend conventional course grades since they are still uncommunicative and they still feed the hunger for ranking.) Portfolios permit me to refrain from grading individual papers and limit myself to writerly evaluative comments—and help students see this as a positive rather than a negative thing, a chance to be graded on a body of their best work that can be judged more fairly. Portfolios have many other advantages as well. They are particularly valuable as occasions for asking students to write extensive and thoughtful explorations of their own strengths and weaknesses.

A midsemester portfolio is usually an informal affair, but it is a good occasion for giving anxious students a ballpark estimate of how well they are doing in the course so far. I find it helpful to tell students that I’m perfectly willing to tell them my best estimate of their course grade—but only if they come to me in conference and only during the second half of the semester. This serves somewhat to quiet their anxiety while they go through seven weeks of drying out from grades. By midsemester, most of them have come to enjoy not getting those numbers and thus being able to think better about more writerly comments from me and their classmates.

Portfolios are now used extensively and productively in larger assessments, and there is constant experimentation with new applications (Belanoff and Dickson; *Portfolio Assessment Newsletter*; *Portfolio News*).

(b) Another useful option is to make a strategic retreat from a wholly negative position. That is, I sometimes do a *bit* of ranking even on individual papers, using two “bottom-line” grades: H and U for “Honors” and “Unsatisfactory.” I tell students that these translate to about A or A- and D or F. This practice may seem theoretically inconsistent with all the arguments I’ve just made, but (at the moment, anyway) I justify it for the following reasons.

First, I sympathize with a *part* of the students’ anxiety about not getting grades: their fear that they might be failing and not know about it—or doing an excellent job and not get any recognition. Second, I’m not giving *many* grades; only a small proportion of papers get these H’s or U’s. The system creates a “non-bottom-line” or “non-quantified” atmosphere. Third, these holistic judgments about best and worst do not seem as arbitrary and questionable as most grades. There is usually a *bit* more agreement among readers about the best and worst papers. What seems most dubious is the process of trying to rank that whole middle range of papers—papers that have a mixture of better and worse

qualities so that the numerical grade depends enormously on a reader's priorities or mood or temperament. My willingness to give these few grades goes a long way toward helping my students forgo most bottom-line grading.

I'm not trying to pretend that these minimal "grades" are truly reliable. But they represent a very small amount of ranking. Yes, someone could insist that I'm really ranking every single paper (and indeed if it seemed politically necessary, I could put an OK or S [for satisfactory] on all those middle range papers and brag, "Yes, I grade everything.") But the fact is that I am doing *much less sorting* since I don't have to sort them into five or even twelve piles. Thus there is a huge reduction in the total amount of unreliability I produce.

(It might seem that if I use only these few minimal grades I have no good way for figuring out a final grade for the course—since that requires a more fine-grained set of ranks. But I don't find that to be the case. For I also give these same minimal grades to the many other important parts of my course such as attendance, meeting deadlines, peer responding, and journal writing. If I want a mathematically computed grade on a scale of six or A through E, I can easily compute it when I have such a large number of grades to work from—even though they are only along a three-point scale.)

This same practice of crude or minimal ranking is a big help on larger assessments outside classrooms, and needs to be applied to the process of assessment in general. There are two important principles to emphasize. On the one hand we must be prudent or accommodating enough to admit that despite all the arguments against ranking, there *are* situations when we need that bottom-line verdict along one scale: which student has not done satisfactory work and should be denied credit for the course? which student gets the scholarship? which candidate to hire or fire? We often operate with scarce resources. But on the other hand we must be bold enough to insist that we do far more ranking than is really needed. We can get along not only with fewer occasions for assessment but also with fewer gradations in scoring. If we decide what the *real* bottom-line is on a given occasion—perhaps just "failing" or perhaps "honors" too—then the reading of papers or portfolios is enormously quick and cheap. It leaves time and money for evaluation—perhaps for analytic scoring or some comment.

At Stony Brook we worked out a portfolio system where multiple readers had only to make a binary decision: acceptable or not. Then individual teachers could decide the actual course grade and give comments for their own students—so long as those students passed in the eyes of an independent rater (Elbow and Belanoff; Belanoff and Elbow). The best way to begin to wean our society from its addiction to ranking may be to permit a tiny bit of it (which also means less unreliability)—rather than trying to go "cold turkey."

(c) Sometimes I use an analytic grid for evaluating and commenting on student papers. An example is given in Figure 1.

Strong OK Weak

			CONTENT, INSIGHTS, THINKING, GRAPPLING WITH TOPIC
			GENUINE REVISION, SUBSTANTIVE CHANGES, NOT JUST EDITING
			ORGANIZATION, STRUCTURE, GUIDING THE READER
			LANGUAGE: SYNTAX, SENTENCES, WORDING, VOICE
			MECHANICS: SPELLING, GRAMMAR, PUNCTUATION, PROOFREADING
			OVERALL [Note: this is not a sum of the other scores.]

Figure 1.

I often vary the criteria in my grid (e.g. “connecting with readers” or “investment”) depending on the assignment or the point in the semester.

Grids are a way I can satisfy the students’ hunger for ranking but still not give in to conventional grades on individual papers. Sometimes I provide nothing but a grid (especially on final drafts), and this is a very quick way to provide a response. Or on midprocess drafts I sometimes use a grid in addition to a comment: a more readerly comment that often doesn’t so much tell them what’s wrong or right or how to improve things but rather tries to give them an account of what is *happening to me* as I read their words. I think this kind of comment is really the most useful thing of all for students, but it frustrates some students for a while. The grid can help these students feel less anxious and thus pay better attention to my comment.

I find grids extremely helpful at the end of the semester for telling students their strengths and weaknesses in the course—or what they’ve done well and not so well. Besides categories like the ones above, I use categories like these: “skill in giving feedback to others,” “ability to meet deadlines,” “effort,” and “improvement.” This practice makes my final grade much more communicative.

(d) I also help make up for the absence of ranking—gold stars and black marks—by having students share their writing with each other a great deal both orally and through frequent publication in class magazines. Also, where possible, I try to get students to give or send writing to audiences outside the class. At the University of Massachusetts at Amherst, freshmen pay a ten dollar lab fee for the writing course, and every teacher publishes four or five class magazines of final drafts a semester. The effects are striking. Sharing, peer feedback, and publication give the best reward and motivation for writing, namely, getting your words out to many readers.

(e) I sometimes use a kind of modified *contract grading*. That is, at the start of the course I pass out a long list of all the things that I most want students to do—the concrete activities that I think most lead to learning—and I promise

students that if they do them *all* they are guaranteed a certain final grade. Currently, I say it's a B—it could be lower or higher. My list includes these items: not missing more than a week's worth of classes; not having more than one late major assignment; *substantive* revising on all major revisions; good copy editing on all final revisions; good effort on peer feedback work; keeping up the journal; and substantial effort and investment on each draft.

I like the way this system changes the “bottom-line” for a course: the intersection where my authority crosses their self-interest. I can tell them, “You have to work very hard in this course, but you can stop worrying about grades.” The crux is no longer that commodity I've always hated and never trusted: a numerical ranking of the quality of their writing along a single continuum. Instead the crux becomes what I care about most: the *concrete behaviors* that I most want students to engage in because they produce more learning and help me teach better. Admittedly, effort and investment are not concrete observable behaviors, but they are no harder to judge than overall quality of writing. And since I care about effort and investment, I don't mind the few arguments I get into about them; they seem fruitful. (“Let's try and figure out why it looked to me as though you didn't put any effort in here.”) In contrast, I hate discussions about grades on a paper and find such arguments fruitless. Besides, I'm not making fine distinctions about effort and investment—just letting a bell go off when they fall palpably low.

It's crucial to note that I am *not* fighting evaluation with this system. I am just fighting ranking or grading. I still write evaluative comments and often use an evaluative grid to tell my students what I see as strengths and weaknesses in their papers. My goal is not to get rid of evaluation but in fact to emphasize it, enhance it. I'm trying to get students to listen *better* to my evaluations—by uncoupling them from a grade. In effect, I'm doing this because I'm so fed up with students *following* or *obeying* my evaluations too blindly—making whatever changes my comments suggest but doing it for the sake of a grade; not really taking the time to make up their own minds about whether they think my judgments or suggestions really make sense to them. The worst part of grades is that they make students obey us without carefully thinking about the merits of what we say. I love the situation this system so often puts students in: I make a criticism or suggestion about their paper, but it doesn't matter to their grade whether they go along with me or not (so long as they genuinely revise in some fashion). They have to think; to decide.

Admittedly this system is crude and impure. Some of the really skilled students who are used to getting A's and desperate to get one in this course remain unhelpfully hung up about getting those H's on their papers. But a good number of these students discover that they can't get them, and they soon settle down to accepting a B and having less anxiety and more of a learning voyage.

THE LIMITATIONS OF EVALUATION AND THE BENEFITS
OF EVALUATION-FREE ZONES

Everything I've said so far has been in praise of evaluation as a substitute for ranking. But I need to turn a corner here and speak about the *limits* or *problems* of evaluation. Evaluating may be better than ranking, but it still carries some of the same problems. That is, even though I've praised evaluation for inviting us to acknowledge that readers and contexts are different, nevertheless the very word *evaluation* tends to imply fairness or reliability or getting beyond personal or subjective preferences. Also, of course, evaluation takes a lot more time and work. To rank you just have to put down a number; holistic scoring of exams is cheaper than analytic scoring.

Most important of all, evaluation harms the climate for learning and teaching—or rather *too much* evaluation has this effect. That is, if we evaluate *everything* students write, they tend to remain tangled up in the assumption that their whole job in school is to give teachers “what they want.” Constant evaluation makes students worry more about psyching out the teacher than about what they are really learning. Students fall into a kind of defensive or on-guard stance toward the teacher: a desire to hide what they don't understand and try to impress. This stance gets in the way of learning. (Think of the patient trying to hide symptoms from the doctor.) Most of all, constant evaluation by someone in authority makes students reluctant to take the risks that are needed for good learning—to try out hunches and trust their own judgment. Face it: if our goal is to get students to exercise their own judgment, that means exercising an immature and undeveloped judgment and making choices that are obviously wrong to us.

We see around us a widespread hunger to be evaluated that is often just as strong as the hunger to rank. Countless conditions make many of us walk around in the world wanting to ask others (especially those in authority), “How am I doing, did I do OK?” I don't think the hunger to be evaluated is as harmful as the hunger to rank, but it can get in the way of learning. For I find that the greatest and most powerful breakthroughs in learning occur when I can get myself and others to *put aside* this nagging, self-doubting question (“How am I doing? How am I doing?”)—and instead to take some chances, trust our instincts or hungers. When everything is evaluated, everything counts. Often the most powerful arena for deep learning is a kind of “time out” zone from the pressures of normal evaluated reality: make-believe, play, dreams—in effect, the Shakespearian forest.

In my attempts to get away from too much evaluation (not from all evaluation, just from too much of it), I have drifted into a set of teaching practices which now feel to me like the *best* part of my teaching. I realize now what I've been unconsciously doing for a number of years: creating “evaluation-free zones.”

(a) The paradigm evaluation-free zone is the ten minute, nonstop freewrite. When I get students to freewrite, I am using my authority to create unusual

conditions in order to contradict or interrupt our pervasive habit of always evaluating our writing. What is essential here are the two central features of freewriting: that it be private (thus I don't collect it or have students share it with anyone else); and that it be nonstop (thus there isn't time for planning, and control is usually diminished). Students quickly catch on and enter into the spirit. At the end of the course, they often tell me that freewriting is the most useful thing I've taught them (see Belanoff, Elbow, and Fontaine).

(b) A larger evaluation-free zone is the single unevaluated assignment—what people sometimes call the “quickwrite” or sketch. This is a piece of writing that I ask students to do—either in class or for homework—without any or much revising. It is meant to be low stakes writing. There is a bit of pressure, nevertheless, since I usually ask them to share it with others and *I* usually collect it and read it. But I don't write any comments at all—except perhaps to put straight lines along some passages I like or to write a phrase of appreciation at the end. And I ask students to refrain from giving evaluative feedback to each other—and instead just to say “thank you” or mention a couple of phrases or ideas that stick in mind. (However, this writing-without-feedback can be a good occasion for students to discuss the *topic* they have written about—and thus serve as an excellent kick-off for discussions of what I am teaching.)

(c) These experiments have led me to my next and largest evaluation-free zone—what I sometimes call a “jump start” for my whole course. For the last few semesters I've been devoting the first three weeks *entirely* to the two evaluation-free activities I've just described: freewriting (and also more leisurely private writing in a journal) and quickwrites or sketches. Since the stakes are low and I'm not asking for much revising, I ask for *much more* writing homework per week than usual. And every day we write in class: various exercises or games. The emphasis is on getting rolling, getting fluent, taking risks. And every day all students read out loud something they've written—sometimes a short passage even to the whole class. So despite the absence of feedback, it is a very audience-filled and sociable three weeks.

At first I only dared do this for two weeks, but when I discovered how fast the writing improves, how good it is for building community, and what a pleasure this period is for me, I went to three weeks. I'm curious to try an experiment with teaching a whole course this way. I wonder, that is, whether all that evaluation we work so hard to give really does any more good than the constant writing and sharing (Zak).

I need to pause here to address an obvious rejoinder: “But withholding evaluation is not normal!” Indeed, it is *not* normal—certainly not normal in school. We normally tend to emphasize evaluations—even bottom-line ranking kinds of evaluations. But I resist the argument that if it's not normal we shouldn't do it.

The best argument for evaluation-free zones is from experience. If you try them, I suspect you'll discover that they are satisfying and bring out good writing. Students have a better time writing these unevaluated pieces; they enjoy hearing and appreciating these pieces when they don't have to evaluate. And *I* have a much better time when I engage in this astonishing activity: reading student work when I don't have to evaluate and respond. And yet the writing improves. I see students investing and risking more, writing more fluently, and using livelier, more interesting voices. This writing gives me and them a higher standard of clarity and voice for when we move on to more careful and revised writing tasks that involve more intellectual pushing—tasks that sometimes make their writing go tangled or sodden.

THE BENEFITS AND FEASIBILITY OF LIKING

Liking and disliking seem like unpromising topics in an exploration of assessment. They seem to represent the worst kind of subjectivity, the merest accident of personal taste. But I've recently come to think that the phenomenon of liking is perhaps the most important evaluative response for writers and teachers to think about. In effect, I'm turning another corner in my argument. In the first section I argued against ranking—with evaluating being the solution. Next I argued not *against* evaluating—but for no-evaluation zones in *addition* to evaluating. Now I will argue neither against evaluating nor against no-evaluation zones, but for something very different in addition, or perhaps underneath, as a foundation: liking.

Let me start with the germ story. I was in a workshop and we were going around the circle with everyone telling a piece of good news about their writing in the last six months. It got to Wendy Bishop, a good poet (who has also written two good books about the teaching of writing), and she said, "In the last six months, I've learned to *like* everything I write." Our jaws dropped; we were startled—in a way scandalized. But I've been chewing on her words ever since, and they have led me into a retelling of the story of how people learn to write better.

The old story goes like this: We write something. We read it over and we say, "This is terrible. I *bate* it. I've got to work on it and improve it." And we do, and it gets better, and this happens again and again, and before long we have become a wonderful writer. But that's not really what happens. Yes, we vow to work on it—but we don't. And next time we have the impulse to write, we're just a *bit* less likely to start.

What really happens when people learn to write better is more like this: We write something. We read it over and we say, "This is terrible. . . . But I *like* it. Damn it, I'm going to get it good enough so that others will like it too." And this time we don't just put it in a drawer, we actually work hard on it. And we try it

out on other people too—not just to get feedback and advice but, perhaps more important, to find someone else who will like it.

Notice the two stories here—two hypotheses. (a) “First you improve the faults and then you like it.” (b) “First you like it and then you improve faults.” The second story may sound odd when stated so baldly, but really it’s common sense. Only if we like something will we get involved enough to work and struggle with it. Only if we like what we write will we write again and again by choice—which is the only way we get better.

This hypothesis sheds light on the process of how people get to be published writers. Conventional wisdom assumes a Darwinian model: poor writers are unread; then they get better; as a result, they get a wider audience; finally they turn into Norman Mailer. But now I’d say the process is more complicated. People who get better and get published really tend to be driven by how much *they* care about their writing. Yes, they have a small audience at first—after all, they’re not very good. But they try reader after reader until finally they can find people who like and appreciate their writing. I certainly did this. If someone doesn’t like her writing enough to be pushy and hungry about finding a few people who also like it, she probably won’t get better.

It may sound so far as though all the effort and drive comes from the lonely driven writer—and sometimes it does (Norman Mailer is no joke). But, often enough, readers play the crucially active role in this story of how writers get better. That is, the way writers *learn* to like their writing is by the grace of having a reader or two who likes it—even though it’s not good. Having at least a few appreciative readers is probably indispensable to getting better.

When I apply this story to our situation as teachers I come up with this interesting hypothesis: *good writing teachers like student writing* (and like students). I think I see this borne out—and it is really nothing but common sense. Teachers who hate student writing and hate students are grouchy all the time. How could we stand our work and do a decent job if we hated their writing? Good teachers see what is only *potentially* good, they get a kick out of mere possibility—and they encourage it. When I manage to do this, I teach well.

Thus, I’ve begun to notice a turning point in my courses—two or three weeks into the semester: “Am I going to like these folks or is this going to be a battle, a struggle?” When I like them everything seems to go better—and it seems to me they learn more by the end. When I don’t and we stay tangled up in struggle, we all suffer—and they seem to learn less.

So what am I saying? That we should like bad writing? How can we see all the weaknesses and criticize student writing if we just like it? But here’s the interesting point: if I *like* someone’s writing it’s *easier* to criticize it.

I first noticed this when I was trying to gather essays for the book on freewriting that Pat Belanoff and Sheryl Fontaine and I edited. I would read an

essay someone had written, I would want it for the book, but I had some serious criticism. I'd get excited and write, "I really like this, and I hope we can use it in our book, but you've got to get rid of this and change that, and I got really mad at this other thing." I usually find it hard to criticize, but I began to notice that I was a much more critical and pushy reader when I liked something. It's even fun to criticize in those conditions.

It's the same with student writing. If I like a piece, I don't have to pussyfoot around with my criticism. It's when I don't like their writing that I find myself tiptoeing: trying to soften my criticism, trying to find something nice to say—and usually sounding fake, often unclear. I see the same thing with my own writing. If I like it, I can criticize it better. I have faith that there'll still be something good left, even if I train my full critical guns on it.

In short—and to highlight how this section relates to the other two sections of this essay—liking is not same as ranking or evaluating. Naturally, people get them mixed up: when they like something, they assume it's good; when they hate it, they assume it's bad. But it's helpful to uncouple the two domains and realize that it makes perfectly good sense to say, "This is terrible, but I like it." Or, "This is good, but I hate it." In short, I am not arguing here *against* criticizing or evaluating. I'm merely arguing *for* liking.

Let me sum up my clump of hypotheses so far:

- It's not improvement that leads to liking, but rather liking that leads to improvement.
- It's the mark of good writers to like their writing.
- Liking is not the same as evaluating. We can often criticize something better when we like it.
- We learn to like our writing when we have a respected reader who likes it.
- Therefore, it's the mark of good teachers to like students and their writing.

If this set of hypotheses is true, what practical consequences follow from it? How can we be better at liking? It feels as though we have no choice—as though liking and not-liking just happen to us. I don't really understand this business. I'd love to hear discussion about the mystery of liking—the phenomenology of liking. I sense it's some kind of putting oneself out—or holding oneself open—but I can't see it clearly. I have a hunch, however, that we're not so helpless about liking as we tend to feel.

For in fact I can suggest some practical concrete activities that I have found fairly reliable at increasing the chances of liking student writing:

(a) I ask for lots of private writing and merely shared writing, that is, writing that I don't read at all, and writing that I read but don't comment on. This makes me more cheerful because it's so much easier. Students get *better* without me. Having to evaluate writing—especially bad writing—makes me more likely to

hate it. This throws light on grading: it's hard to like something if we know we have to give it a D.

(b) I have students share lots of writing with each other—and after a while respond to each other. It's easier to like their writing when I don't feel myself as the only reader and judge. And so it helps to build community in general: it takes pressure off me. Thus I try to use peer groups not only for feedback, but for other activities too, such as collaborative writing, brainstorming, putting class magazines together, and working out other decisions.

(c) I increase the chances of my liking their writing when I get better at finding what *is* good—or *potentially* good—and learn to praise it. This is a skill. It requires a good eye, a good nose. We tend—especially in the academic world—to assume that a good eye or fine discrimination means *criticizing*. Academics are sometimes proud of their tendency to be bothered by what is bad. Thus I find I am sometimes looked down on as dumb and indiscriminating: “He likes bad writing. He must have no taste, no discrimination.” But I've finally become angry rather than defensive. It's an act of discrimination to see what's good in bad writing. Maybe, in fact, this is the secret of the mystery of liking: to be able to see potential goodness underneath badness.

Put it this way. We tend to stereotype liking as a “soft” and sentimental activity. Mr. Rogers is our model. Fine. There's nothing wrong with softness and sentiment—and I love Mr. Rogers. But liking can also be hard-assed. Let me suggest an alternative to Mr. Rogers: B. F. Skinner. Skinner taught pigeons to play ping-pong. How did he do it? Not by moaning, “Pigeon standards are falling. The pigeons they send us these days are no good. When I was a pigeon . . .” He did it by a careful, disciplined method that involved close analytic observation. He put pigeons on a ping-pong table with a ball, and every time a pigeon turned his head 30 degrees toward the ball, he gave a reward (see my “Danger of Softness”).

What would this approach require in the teaching of writing? It's very simple . . . but not easy. Imagine that we want to teach students an ability they badly lack, for example how to organize their writing or how to make their sentences clearer. Skinner's insight is that we get nowhere in this task by just telling them how much they lack this skill: “It's disorganized. Organize it!” “It's unclear. Make it clear!”

No, what we must learn to do is to read closely and carefully enough to show the student little bits of *proto*-organization or *sort of* clarity in what they've already written. We don't have to pretend the writing is wonderful. We could even say, “This is a terrible paper and the worst part about it is the lack of organization. But I will teach you how to organize. Look here at this little organizational move you made in this sentence. Read it out loud and try to feel how it pulls together this stuff here and distinguishes it from that stuff there. Try to remember what it

felt like writing that sentence—creating that piece of organization. Do it some more.” Notice how much more helpful it is if we can say, “Do *more* of what you’ve done here,” than if we say, “Do something *different* from anything you’ve done in the whole paper.”

When academics criticize behaviorism as crude it often means that they aren’t willing to do the close careful reading of student writing that is required. They’d rather give a cursory reading and turn up their nose and give a low grade and complain about falling standards. No one has undermined behaviorism’s main principle of learning: that reward produces learning more effectively than punishment.

(d) I improve my chances of liking student writing when I take steps to get to know them a bit as people. I do this partly through the assignments I give. That is, I always ask them to write a letter or two to me and to each other (for example about their history with writing). I base at least a couple of assignments on their own experiences, memories, or histories. And I make sure some of the assignments are free choice pieces—which also helps me know them.

In addition, I make sure to have at least three conferences with each student each semester—the first one very early. I often call off some classes in order to keep conferences from being too onerous (insisting nevertheless that students meet with their partner or small group when class is called off). Some teachers have mini-conferences with students during class—while students are engaged in writing or peer group meetings. I’ve found that when I deal only with my classes as a whole—as a large group—I sometimes experience them as a herd or lump—as stereotyped “adolescents”; I fail to experience them as individuals. For me, personally, this is disastrous since it often leads me to experience them as that scary tribe that I felt rejected by when *I* was an eighteen-year-old—and thus, at times, as “the enemy.” But when I sit down with them face to face, they are not so stereotyped or alien or threatening—they are just eighteen-year-olds.

Getting a glimpse of them as individual people is particularly helpful in cases where their writing is not just bad, but somehow offensive—perhaps violent or cruelly racist or homophobic or sexist—or frighteningly vacuous. When I know them just a bit I can often see behind their awful attitude to the person and the life situation that spawned it, and not hate their writing so much. When I know students I can see that they are smart behind that dumb behavior; they are doing the best they can behind that bad behavior. Conditions are keeping them from acting decently; something is holding them back.

(e) It’s odd, but the more I let myself show, the easier it is to like them and their writing. I need to share some of my own writing—show some of my own feelings. I need to write the letter to them that they write to me—about my past experiences and what I want and don’t want to happen.

(f) It helps to work on my own writing—and work on learning to *like* it. Teachers who are most critical and sour about student writing are often having trouble with their own writing. They are bitter or unforgiving or hurting toward their own work. (I think I've noticed that failed PhDs are often the most severe and difficult with students.) When we are stuck or sour in our own writing, what helps us most is to find spaces free from evaluation such as those provided by freewriting and journal writing. Also, activities like reading out loud and finding a supportive reader or two. I would insist, then, that if only for the sake of our teaching, we need to learn to be charitable and to like our own writing.

A final word. I fear that this sermon about liking might seem an invitation to guilt. There is enough pressure on us as teachers that we don't need someone coming along and calling us inadequate if we don't *like* our students and their writing. That is, even though I think I am right to make this foray into the realm of feeling, I also acknowledge that it is dangerous—and paradoxical. It strikes me that we also need to have permission to hate the dirty bastards and their stupid writing.

After all, the conditions under which they go to school bring out some awful behavior on their part, and the conditions under which we teach sometimes make it difficult for us to like them and their writing. Writing wasn't meant to be read in stacks of twenty-five, fifty, or seventy-five. And we are handicapped as teachers when students are in our classes against their will. (Thus high school teachers have the worst problem here, since their students tend to be the most sour and resentful about school.)

Indeed, one of the best aids to liking students and their writing is to be somewhat charitable toward ourselves about the opposite feelings that we inevitably have. I used to think it was terrible for teachers to tell those sarcastic stories and hostile jokes about their students: "teacher room talk." But now I've come to think that people who spend their lives teaching *need* an arena to let off this unhappy steam. And certainly it's better to vent this sarcasm and hostility with our buddies than on the students themselves. The question, then, becomes this: do we help this behavior function as a venting so that we can move past it and not be trapped in our inevitable resentment of students? Or do we tell these stories and jokes as a way of staying stuck in the hurt, hostile, or bitter feelings—year after year—as so many sad teachers do?

In short I'm not trying to invite guilt, I'm trying to invite hope. I'm trying to suggest that if we do a sophisticated analysis of the difference between liking and evaluating, we will see that it's possible (if not always easy) to like students and their writing—without having to give up our intelligence, sophistication, or judgment.

Let me sum up the points I'm trying to make about ranking, evaluating, and liking:

- Let's do as little ranking and grading as we can. They are never fair and they undermine learning and teaching.
- Let's use evaluation instead—a more careful, more discriminating, fairer mode of assessment.
- But because evaluating is harder than ranking, and because too much evaluating also undermines learning, let's establish small but important evaluation-free zones.
- And underneath it all—suffusing the whole evaluative enterprise—let's learn to be better likers: liking our own and our students' writing, and realizing that liking need not get in the way of clear-eyed evaluation.

WORKS CITED

- Diederich, Paul. *Measuring Growth in English*. Urbana: NCTE, 1974.
- Belanoff, Pat, and Peter Elbow. "Using Portfolios to Increase Collaboration and Community in a Writing Program." *WPA: Journal of Writing Program Administration* 9.3 (Spring 1986): 27–40. (Also in *Portfolios: Process and Product*. Ed. Pat Belanoff and Marcia Dickson. Portsmouth, NH: Boynton/Cook-Heinemann, 1991.)
- Belanoff, Pat, Peter Elbow, and Sheryl Fontaine, eds. *Nothing Begins with N: New Investigations of Freewriting*. Carbondale: Southern Illinois UP, 1991.
- Bishop, Wendy. *Something Old, Something New: College Writing Teachers and Classroom Change*. Carbondale: Southern Illinois UP, 1990.
- . *Released into Language: Options for Teaching Creative Writing*. Urbana: NCTE, 1990.
- Elbow, Peter. "The Danger of Softness." *What Is English?* New York: MLA, 1990. 197–210.
- Elbow, Peter, and Pat Belanoff. "State University of New York: Portfolio-Based Evaluation Program." *New Methods in College Writing Programs: Theory into Practice*. Ed. Paul Connolly and Teresa Vilardi. New York: MLA, 1986. 95–105. (Also in *Portfolios: Process and Product*. Ed. Pat Belanoff and Marcia Dickson. Portsmouth, NH: Boynton/Cook-Heinemann, 1991.)
- Gardner, Howard. *Frames of Mind: The Theory of Multiple Intelligences*. New York: Basic, 1983.
- Kirschenbaum, Howard, Simon Sidney, and Rodney Napier. *Wad-Ja-Get? The Grading Game in American Education*. New York: Hart Publishing, 1971.
- Lewis, C. S. *Studies in Words*. 2d ed. London: Cambridge UP, 1967.
- Portfolio Assessment Newsletter*. Five Centerpointe Drive, Suite 100, Lake Oswego, Oregon 97035.

Portfolio News. c/o San Dieguito Union High School District, 710 Encinitas Boulevard, Encinitas, CA 92024.

Smith, Barbara Herrnstein. *Contingencies of Value: Alternative Perspectives for Critical Theory*. Cambridge: Harvard UP, 1988.

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

Zak, Frances. "Exclusively Positive Responses to Student Writing." *Journal of Basic Writing* 9.2 (1990): 40–53.