

The shadow of testing

Carolyn Matalene

No doubt we have all seen students denigrated for making surface errors and ignored when offering creative responses. And we have also seen the fear and cowardice and mindlessness and hatred of writing (and of reading far that matter) which mindless, begrudging, pinched and prescriptive composition (or literature) teaching has caused in our students. Between the idea we have of language teaching and the reality of classroom practices, between what Alan Hollingsworth calls "a rich model of language instruction" and what students actually get, falls the shadow. The more language instruction I observe and the more students I hear from, the more inevitable seems T.S. Eliot's line: "between the idea and the reality falls the shadow."

Our current cultural response to this educational--if not human condition--seems to be: into the shadows insert a test. A test, we easily assume, will remove the shadow, will make the ideal equal the reality. Now at last, that lumpen reality will have to shape up--like Tennessee 'round Wallace Stevens' jar--though the unstated assumption that a test will in fact cause learning is seldom examined.

Ever since the media decided to make copy out of students' language deficiencies, testing has become chic. It sells. Some of us have even figured out that getting into testing will also get our pictures in the paper, our names on the governor's committee, and our pay raises (and travel requests) past the dean. Others have been ordered into testing by administrators at their institutions. At year before last's CCCC workshop which I conducted (and which got me to San Francisco), participants from all over the country had been so directed: produce a test.

Testing language skills these days is as common a practice as cutting the budget, and sometimes the two are linked. As Oedipus discovered, however, obvious solutions sometimes yield unexpected results, a phenomenon we are all too well acquainted with in the twentieth century. Therefore, we should be aware of the unexpected, bizarre, and ironic effects that tests can cause.

As language teachers, we are usually interested in testing students to gain information for placement, to design appropriate courses, and to evaluate the effectiveness of our instruction. At my institution, for example, the students who score lowest on our Writing Proficiency Test are placed in Basic Writing; at the end of the semester we administer the same test again to see what, if anything, we have accomplished, and what major problems remain. But our colleagues in other departments or in the administration are more often interested in tests for exclusionary purposes. Thus, when we implemented this proficiency test for purposes of placement--to be scored one, two, three, or four--several people suggested that we add a score of zero and use it to deny students admission to the university. So, ironically, the placement tests which for many of us were inspired by responses to the Open Door policy at CCNY would now be used by many to Shut the Door.

Tests designed as instruments for measuring can easily become instruments for eliminating. When exclusion is the issue, few seem to care about the nature of the instrument. (Dr. Guillotin was widely lauded as a humanitarian in his day.) This results in an effect I had never expected, the "Mind if I borrow your test?" syndrome. Someone is always willing and anxious to use whatever test is around for something else, anything else. It was recently suggested to me that students' scores on that same Writing Proficiency Test, when high enough, could also be used to exempt honors students from a sophomore literature course. I found it difficult to retain my genteel voice as I explained that a student who can swim across the pool is not necessarily proficient at hang-gliding as well. Test-borrowing now seems to be a national phenomenon. After arguing in print that an objective test of particular writing skills could perhaps be useful but if and only if it were carefully designed according to local standards, devised by local faculty, and related to the specific content of a specific course, I was surprised by the calls and letters I received from all over the country asking, "Mind if we borrow your test? We've been told we've got to have one." Clearly, any test in a storm.

Those of us from states with minimal competency testing legislated into the public schools now know how many ways these test results can be used: to pressure teachers, to worry administrators, to heap abuse on colleges of education, to despair of today's youth, or to denigrate low-brow parents. But I hadn't expected them to be used for selling real estate. In my town test scores--and percentage increases--are reported on the front pages of both papers, school district by school district; the scores then are quoted by real estate agents as they push particular neighborhoods. No doubt as test scores rise and fall, so do house prices. But then, who would have expected the competition for high SAT Verbal scores to save the teaching of Latin?

We probably cannot control the peripheral effects of test scores, but we can admit that much of our own dependence on test scores is that tests make the best flak-catchers. As Eliot also told us, humankind cannot stand very much reality, and the unpleasant reality of education is that some will do very poorly, some will fail utterly. This reality is not pleasant, either for the person who failed or for the person who must make the announcement. The reason, after all, that educators hang on to SAT and GRE scores is not because we believe in them--they are disproved before our very eyes daily--but because numbers are the best flakcatchers. When lines have to be drawn somewhere, it seems so much kinder to show a student a number than to reveal an evaluation by another human which finds the student sorely wanting. Anyone who has ever had to be the flak-catcher, to hear the appeals and arguments, and to say to the student "this isn't very good" (always heard as "I'm not very good") knows that. Flak-catching could surely be added to the list of stressful situations which the life insurance companies circulate. And they also know why the numbers spat out by batteries of objective tests are so tempting. Thus, the testing phenomenon betrays a natural tendency to move toward numerical and objective evaluations, a tendency not lessened by the sublime and ubiquitous efficiency of the computer. And in fact, the more I try to analyze the shadow which always falls between the goals of our tests and the actual results or between our idea of instruction and the reality we have to settle for, the more it seems that a number of ineluctable Natural Laws lurk here. We might as well admit some of them.

Law I: We never teach what we think we're teaching.

This may sound preposterous. but it is truer than we dare to admit. It was brought home to me one day after I had taught a particularly fine class on Flaubert's great story "A Simple Heart." A rather spacey young woman approached me with unusual enthusiasm, "May I ask you a question, may I ask you a question?" At last, I thought, I have interested her. "Tell me," she said, "where's your diamond ring? I always watch your diamond ring during class, and you weren't wearing it today]" When students tell us, years hence, what they got from a particular class, it is seldom what we had in mind. Who knows what in fact I might have been teaching students about women academics who wear diamond rings--all the while that I thought I was teaching point of view, or Flaubert, or cumulative sentences? And who knows what the student was ready for or interested in receiving? In any classroom, the variables are innumerable: the levels of ability and achievement, the mixtures of interests, and motivation are so complex that what any one student is getting from any one teacher is not only unique, but also unpredictable.

Law II: We never test what we think we're testing.

This was revealed to me when a student, who had done poorly on the Writing Proficiency Test and who, with her daddy, had given me considerable flak, came by a year later to tell me of her progress through our course sequence. She said, "You know, I found out why I did bad on that test. I know what I did wrong. I only had three paragraphs. Now I know you have to have five."

Tests tend to measure the test taker's ability to take tests. Test-taking is a ritualized activity which, like praying, some are more comfortable with than others. There really are smart students who "test poorly" and lazy ones "who test well." (They often turn up in the bosoms of academic families.) Furthermore, tests will not make test takers learn--at least probably not what we want them to learn--see Law I--or at any rate the test takers who need to learn probably won't. Unless at least 12 to 15% of the takers fail or do poorly on any given test, it isn't worth the energy and expense to administer it. The 85% who pass then prove that the test was too easy and therefore unnecessary, and the 15% who fail the test--those who were supposed to be motivated by the test to pass--prove that the test has failed in motivating them. Naturally, in the population of failures, there will always be some who should not have failed and of those who passed, some who should not have passed. Thus, a test introduced to bring the real closer to the ideal will probably only introduce into the world more injustice and, as an accompaniment, more flak.

Law III: The effects of a test are never quite what we had in mind.

Not only do students draw crude conclusions about the ontological existence of the five-paragraph theme from the tests they take, they also draw conclusions about

what constitutes an education, what they need to know, and why they need to know it.

Teachers then easily follow the line of least resistance and teach to the test. When an extremely competent young teacher explained to me how he taught his students to analyze the test writer's mentality and then second guess it, I realized that the test he was teaching to had indeed affected instruction, but in an unexpected way. In my state, the Basic Skills Assessment Program, as the minimal competency tests are officially designated, has been renamed "Be Sap" by the teachers; they and their principals generally agree that although this carefully designed testing program has so far produced multitudinous ring binders of regulations, it has not yet improved instruction. Some cynics even argue that the time spent on form-filling both diminishes teaching time and further demoralizes teachers.

These three laws about teaching and testing are but corollaries of a fourth law which accounts for why there must always be a shadow.

Law IV: Nobody ever gets anything right.

Some people, upon reading or hearing about a plane hijacking think the news item is not a report but a direction to do likewise. And when it comes to testing, points of view also tend toward the personal. Administrators and legislators assume that tests will cause both improved instruction and increased learning; teachers see the same tests as harassments which interrupt instruction. Parents believe that test scores should be directly proportional to the money they have spent and the study hours their children have invested; students know that tests are cruel and mysterious determiners which often reward the dilatory and punish the diligent. Thus, students who do poorly on proficiency tests or who fail exit exams or qualifying exams or final exams seldom in my experience blame themselves or the education they have thus far received. Instead, they blame the test or the Director of Freshman English or the Dean or the institution, or, like the Greeks, the messenger. From the test taker's point of view, we are never testing what the test taker knows, but what he or she doesn't know--and we're back to Law II, we never test what we think we are testing.

So what's a conscientious WPA to do?

To be serious (but only mildly and briefly so), I should perhaps say that I do believe in tests. They undoubtedly have a legitimate place in our educational system. They can provide essential information; they can fulfill positive and useful functions. They can be made to do what we want them to do, and we can learn a great deal from designing and administering them--though usually not what we expected to learn. We can also effect change--sometimes even for the better.

The essential issue, too easily and too often overlooked, is what might be called the rhetorical context of testing. Tests are not simple cures--they are accompanied by many side effects--nor are they single events. Like writing, testing occurs in a

social context and is as complicated as is discourse itself. If a rich model of writing should inform our teaching, so should a complex model of testing inform our educational policies.

That perpetual *vade mecum* of writing teachers, Jakobson's diagram of the elements of discourse, might well serve as a model for testing programs as a heuristic for anticipating the unanticipated. This familiar six-part model specifies that any piece of discourse between an addresser and an addressee is transmitted in a Code, through a medium (Contact), about a subject (Context) and with a Message. Similarly, a test is a piece of discourse between the test giver and the test taker. It is transmitted in a Code (the test format), through a Contact (the actual administration of the test), about a Subject (the information given or received), and with an inevitable Message.

These components are easy enough to analyze; the catch is to remember Law IV, "nobody ever gets anything right." That is, the test giver and the test taker tend to have two profoundly different points of view. That the interpretation of any text can never be stabilized has, after all, kept us in work for these many years. Quite naturally, the test-giver's point of view is institutional and educational, while the test taker's is profoundly personal--maybe even paranoid. Consider, for example, the differences of interpretation that may result from a now common procedure in English departments, the placement test. Using the testing heuristic, let us look at such a test from these two points of view.

From the test-giver's point of view, *Context* is what matters. The information received from the student, revealing the student's level of competence, is the only reason for bothering with placement tests. Thus, the central debate in testing circles is over which kind of test, essay or objective, gives the most reliable and valid information. Not that test-givers seldom care *what* students say in placement essays, only how well or badly they say it. This kind of writing is not seen as communication, but as information. Therefore, the test format (Code) is determined according to how easily or cheaply that information can be obtained. (Mind if I borrow your test?) The ideal format, of course, is inexpensive, trouble-free, and easy to score. Test-givers commonly pay little attention to *Contact*, that is, the pens, paper, rooms, and proctors needed to make the test happen. Whatever rooms are available, preferably far from the department, and whatever proctor can be summoned, preferably not us, will do. The time for administering the test--days and hours--is often a problem, but test givers with political influence usually manage to borrow time from an orientation program rather than from scheduled classes.

The interaction of Context, Contact, and Code yield the Message sent by a particular test. But when it comes to Message, writer and reader or test-giver and test-taker commonly disagree. Certainly, test-givers seldom think carefully or deeply enough about the Message that they ought to send or that a particular combination of Context, Contact, Code will in fact send. A failure of self-consciousness is all too common, and test-givers often assume with relish that students are hearing smug threats, "Now we've got you. You can't fool us anymore. Off to the basics for re-education."

But the students are all Jamesian characters; their points of view are forever personal. For them, the Context of a placement essay is not the level of their writing skills, but the subject matter which the topic demands. They must invent reasonable responses, on the spot, to whatever topic or topics stare up at them. To

be given impossible, unsuitable, or dumb topics naturally confirms their worst suspicions about English teachers. Of course, one person's dumb topic might be another's special interest; but one group's current interests may not be another's. One of our topics which began, "Your high school guidance counselor..." incensed older students. The test-takers in almost any testing situation are not a homogeneous audience of eighteen-year-old middle class Americans; age-bound or culture-bound topics hardly allow a fair chance to all. Test-takers tend to view Context as content--not as information--and thus, they frequently assume that a low score means the grader disagreed with their opinions, even though the grader probably didn't even notice them.

The testing situation, Contact, is much more important to test-takers than test-givers assume. Most of us remember vividly the rooms we sat in to take important exams, and the distractions we had to endure. Hot, overcrowded rooms, rude proctors who won't stop talking, and seats that squeak all affect test-takers, though test-givers tend to ignore them. The test's format, however, is surely the most important element in the Message that students receive from a test. The format of a test used to place students in writing courses actually provides a constitutive definition of writing itself. Tests that require students to fill in bubbles on computer-scored answer forms assert that writing is "bubbling." Tests that ask true/false or multiple choice questions define writing as a matter of right or wrong decisions, of knowing rules and spotting errors. Essay questions about such old chestnuts as the value of an education remind students that any collection of clichés will do so long as it is an orderly and carefully punctuated collection.

The Context, the Contact, and the Code then of any placement test do indeed add up to the test's Message, a message about what writing is, what our philosophy of writing instruction is, what our definition of writing competence is, as well as how we value writing and how we respond to it. Thus, we had better be careful. Furthermore, we reveal our response to writing by how we score tests and by how rigidly we maintain the validity of these scores. Second chances and appeal procedures send the message that a writer's skill can vary according to the task and according to the day; they also send the message that the test-givers and the test-scorers are human.

By now it should be clear that as far as I am concerned, the Message a test sends is vastly more important than any information we gather, and that therefore, the debate over whether objective tests or essay tests gather more accurate information is useless. Objective tests send such bad messages that it hardly matters what placement information we get from them. The information we get from essay tests is, as we all know, very crude and very messy and often unreliable. Few of us are ready to defend our graders' decisions in a court case. But the messages they send are much closer to our goals as writing teachers. Moreover, these messages are explicit, and they reverberate.

A second set of receivers is listening quite carefully to the message we send: the high school teachers. Students do go back and do complain and praise and blame and question. Here is a tremendous force for change--in teaching methods as well as in curriculum. As more students carry the message back, more writing is taught and more writing courses are offered. A test that sends the right message, then, can provide high school teachers with a political lever in their own institutions and can also provide a much-needed channel for communication between high school and

college English teachers. Students and teachers both assert the final Law of Testing: The message matters more than the information gathered. (And that is why we never test what we think we are testing.)

Thus, a complex set of factors is involved in the design and administration of any test. The best we can do is anticipate as many problems as possible and imagine as many of the messages as possible. The urge to test is likely to be around for a while, so testing should be seen for what it is. It is not the solution for making the reality equal the idea; it is rather another shadow, another distorting lens falling between the idea and the reality, one that should be used with great circumspection, lest our best energies only serve to deepen the shadows.