THE CENTRALITY OF MEASUREMENT

Speakers: Edward White, California State University, San Bernardino

Edward White pointed out that many writing teachers not only are ignorant of measurement concepts but are proud of their ignorance. Since measurement is central, not peripheral, to the teaching of writing, a writing teacher must know something about measurement, according to White. By examining three of the most prominent problems in the teaching of writing today, one can see how an understanding of the principles and procedures of writing measurement is necessary in order to address these problems. Inattention to these problems represents inattention to the teaching of writing itself:

1. Unclear and inappropriate writing assignments.
2. Arbitrary and unfair grading of student writing.
3. Insufficient or no attention to the writing process in general and to revision in particular.

White said that writing teachers are legendary for their bad assignments. Students often do not know what their teachers want. Yet, a teacher who has participated in a test committee for the creation of an essay test will generally change her or his classroom assignments for the better. Test committees review, revise, and protest essay assignments. Those teachers who have worked on test committees understand the obligation seriously to consider test construction, that is, the kinds of writing assignments that are appropriate.

White went on to discuss the second problem—arbitrary and unfair grading of student writing. Our students generally believe, based on their experience, that grades for writing are unpredictable, arbitrary, inconsistent, and normally a matter of luck more than skill; and most research into teacher grading shows that they are right. Teachers involved in an essay testing program will soon realize that scoring reliability is necessary for the sake of fairness and achievable with some effort.

In regard to the third problem, insufficient or no attention to the writing process in general or to revision in particular, White commented that every experienced writer, particularly a professional writer, knows that revision is the substance of writing. The student writer needs to learn this. The most important obstacle to the teaching of writing is that most students write only a first draft. White believes that students will not revise their work until teachers teach them enough about evaluating writing to enable them to see what needs improvement in their own drafts. The half-hearted tinkering that most good students do to please teachers—often worsening rather than improving their work—has little to do with real revision. Revision means evaluation, and teachers cannot teach revision until they themselves understand evaluation and measurement.

PORTFOLIO ASSESSMENT AS AN ALTERNATIVE TO PROFICIENCY TESTING

Speaker: Peter Elbow, State University of New York, Stony Brook

Peter Elbow began the session by noting problems with current types of writing proficiency tests. He believes the major problem with most tests is that they include only a single sample of writing. Yet evaluators cannot, he maintains, get a trustworthy picture of a student’s writing proficiency unless they look at several samples produced on several days in several modes or genres. Another problem is that most proficiency tests undermine our teaching of writing by implying to students that they can demonstrate their proficiency in one hour, without extensive thinking, talking, composing, and revising. Finally, the topics on most proficiency tests are irrelevant to the students’ curriculum, unconnected to the study of any material, and cut off from ongoing intellectual discussions.

Thus, in 1982, five teachers at Stony Brook began experimenting with portfolio assessment: each student in an experimental section produced a portfolio consisting of four revised pieces (two arguments, one informal essay, and one piece of free-choice prose, in addition to an in-class essay). Only 55 percent of the students passed. The teachers then instituted a mid-semester evaluation of one of the compositions so that students and teachers would have a better understanding of the criteria for passing. In addition, the number of pieces on the portfolio was reduced to two (one argument, one analysis of an argument, and one piece of free-choice prose). In the fall of 1984, portfolio assessment was made an official procedure for all sections of the freshman composition course. Every student must get a C or higher in this course, and no student can get a C unless his or her portfolio has been judged worth a C by at least one other teacher who does not know him or her.

Currently, all composition teachers meet at mid-semester to discuss sample papers and to “calibrate” their standards. Scores on the essays are binary: C or above/below a C. If a teacher disagrees with the score of the second reader, the essay gets read and scored by a third reader. At the end of the semester, this evaluation process is repeated but with full portfolios. Each portfolio is treated more or less as a whole (that is, each essay within a portfolio does not receive a separate score). If two readers agree that a portfolio has failed because of one of the papers in it, the writer may revise that paper and resubmit the portfolio. About half of the mid-semester papers fail. At the end of the semester, about 10 percent of the portfolios fail, but that figure goes down to about 5 percent after some are rewritten.

Elbow stated that he and his colleagues see the portfolio as a way to ask students for better writing, a way that is proving successful. The portfolio process judges student writing in ways that better reflect the complexities of the writing process, and it does not insist that students be judged on a single sample. Moreover, the process makes
students are given their choice of six to nine topics, such as "my first car," the "opportunity of a lifetime," and "my most memorable person." This placement exam, like the 101 exit exam, is taken pass/fail, but students who fail have to enroll in English 101; they cannot transfer their equivalent class credit. About 23 percent of USA's transfer students fail this exam.

The major difference between this transfer placement test and the 101 exit exam is in the scoring. While the 101 exit exams are graded by the English Department, the scoring of the transfer placement tests is done by the faculty members who are teaching W courses across the curriculum. More than half of the faculty who have attended the writing workshops have participated as raters of these exams. The use of faculty members from other departments reflects a belief that evaluating the load of transfer-student placement tests cannot be absorbed wholly by the English Department. Moreover, this provides a vehicle for increasing the participation of non-English faculty in the writing program and serves to develop coherence between all the departments on campus and the writing standards required at USA. Using non-English faculty raters also allows faculty members teaching W courses to have some impact on who can take their courses.

Raters are given a training session in evaluation before each new placement exam to review the testing materials. The purpose of the test is reiterates during this session, and sample papers are discussed, using a range of papers from those that would clearly pass, to marginal cases, to those that would clearly fail. These sessions serve not only as preparation for new readers but also as review for evaluators who have not rated for a quarter or more. Furthermore, the confidence of the raters is bolstered when they see that they generally agree with their colleagues. To eliminate bias, raters are placed in heterogeneous groups so that no group is all one sex or all from one college. Matre has developed a computer program that assigns groups by controlling these variables. Exams are then assigned to the groups randomly. Faculty evaluators are paid $30 for rating a packet of twenty to thirty placement exams, which have been photocopied. After receiving the exams, raters have twenty-four hours to score them. Each paper is evaluated by at least two raters, and splits are rated by three or more readers. Because the pass/fail rating seemed too rigid, faculty members elected to use a confidence rating of 1-2-3-4-5 for passing papers, ranging from 1, meaning "pass with no confidence," to 5, meaning "passing with confidence." Papers that receive a split evaluation with the passing mark indicating low confidence may have more than just the third reader to insure the evaluation is as fair as possible to the student.

Two concerns of this cross-curriculum grading program were whether raters could be consistent and whether they could agree. Therefore, Matre developed another computer program to determine the most reliable and the most valid readers, in an attempt to establish an evaluator core from these readers. This proved to be unrealistic in practice, however, because of the diplomatic problem of dismissing people as poor readers. Matre also wanted to develop these reliability and validity scores to aid in pairing readers so that groups were varied not only by sex and college but also by reliability and validity ratings. For a reliability rating, he had to collect scoring data for the placement exam, which he computerized, comparing (1) the decision of the first rater, (2) the decision of the second rater, (3) how closely they agreed. The most reliable raters, of course, agreed the most often. To determine validity, he used computerized information on GPA's, available through other sections of the University, to compare each student's placement test rating to his GPA for courses taken at USA the semester following the test. Valid evaluators did not pass students who were falling nor fail students who were passing. Matre presented reliability and validity statistics for each of the raters in the winter quarter of 1984 and the fall quarter of 1985; the data indicated that there are not many unreliable or invalid readers. However, those identified invalid are not asked to rate again, though this is still difficult to uphold if someone really wants to be a rater.

Matre concluded by stating that he hopes in the future to draw a random sample of the holistically scored placement tests, have them analytically evaluated, and compare the results to determine the reliability of the holistic evaluation as well.

Portfolio Assessment continued

teachers allies of their students, allies who work with them to revise and to pass. And it also draws teachers together, encouraging discussion about standards and pedagogy, and, inevitably, making departmental standards more consistent. Elbow acknowledged that the process makes much more work for teachers and puts more pressure on them. Also, some teachers feel that the opportunity to revise failed portfolios coddles students too much and lets lazy students get by with help and nagging from their teachers and peers.

Elbow concluded, however, that the portfolio system improves the trustworthiness of evaluation because raters can base their judgment on several pieces of writing. Moreover, it sends students the message that writing is a rich and multifaceted process.