CORE CURRICULUM AND THE WRITING PROFICIENCY EXAM RECONSIDERED

In 1979, when the College of Arts and Sciences at U Mass/Boston began giving its required Writing Proficiency Exam, we knew we would not have to ask students their opinions about the exam. No matter how we explained the exam, its aims, its design, they argued against it and denounced it. Even though the students did not have to pass the exam before the beginning of the junior year, and although the Core Curriculum provided instruction in writing in most of the Core courses and in freshman English, students thought the exam unfair and resisted it. They presented the penalties: the fact that first-term juniors who had not passed the exam could not take a full load of courses and had to enroll in a special non-credit writing course. Even more, they resisted the fact that after the first half of the junior year they could not take any courses until they passed the exam. Student sentiment most emphatically showed itself a year after the exam was first given, when students occupied a University building. The students at that first presented three demands: reversal of two faculty tenure decisions, rehiring of a Student Affairs staff member, and a 51% student vote on all University committees. But when the protesters needed wider support, they got it by expanding their list of demands. The fourth was "End the Proficiency Exam," well ahead of No. 9, "Repeal of the 43% tuition increase."

The exam is designed to test the students' ability to write about a specific topic discussed in a set of readings. Four or five times a year the examiners distribute packets of readings: each packet contains three or four essays about a specific topic, like "Immigration Policy, The Recombinant DNA Controversy," or "The Films of Alfred Hitchcock." A month before the exam students pick up the packets and look them over, choose to study one set of readings carefully, and in the exam room ask for the questions on that set. In June, 1981, after three years of assembling packets of readings on challenging yet manageable topics, the examiners fell back on the topic of Liberal Education to round out a packet that already contained sets of readings on The Brain and Social Science fiction. The Liberal Education set had essays by J.M. Cameron, Mickey Friedman, Peter Engel and Russell Baker. The examiners gave the students a choice of two questions. The first asked them to evaluate the extent to which various college curricula described in the essays met the goal of liberal education which one of the essayists said was "to awaken innocent minds to a suspicion of information." The second question, the one that provided some insight into student opinion of the exam itself, was:

You are now taking the UM/B College of Arts and Sciences Writing Proficiency Examination. To what extent is this examination consistent with the views of education contained in the assigned readings? (You may take any position you wish, as long as your answer is supported by specific references to the readings.)

Fifteen students wrote on the "suspicion of information" question, twenty on the educational value of the exam itself. Out of the 150 students at that sitting of the exam, twenty (8%) hardly constitute a sample; but their bluebooks offer some interesting comments and confirm a general college view of the exam that students have come to accept the exam. First of all, the bluebooks show that the college catalogue and brochures, the orientation meetings, the advising system, and classroom instruction have combined to establish the Proficiency Exam in the students' minds as a coherent part of the college's response to student problems in writing. Like the rest of us, students have failed to teach writing. Hence, whatever their appraisal of the exam, most bluebooks focused attention on how effectively the college helped improve student writing rather than on whether we should have an exam at all. And it was gratifying to find so many explicit comments that the faculty in introductory courses (Continued on page 21)

plan, and write (tasks that are, after all, inseparable in the experience of most working writers), extend the time for completing the test so that the students can proceed as a working writer would.

My second and third suggestions are amply discussed in the literature on the making of writing assignments. But the first suggestion needs some elaboration here.

Researchers on testing insist that first suggestion by asserting the need for precise comparability between the tasks set on different versions of a test. If the tasks differ, the argument goes, the scores on the different tests may not mean the same thing, and attorneys in legal disputes over test results may argue that the test is discriminatory. Possibly, but the usual inference from these assertions is that from version to version the "mode" of writing asked for must be the same. Despite their current status as stereotyped categories for use in discussing writing, the "modes" mislead us as teachers and test-makers. Writers do not write in modes; they write to reach audiences on subjects of concern, employing whatever speech acts (defining, restating, inferring, conceding, and so on) will enable them to accomplish their purposes. Instead of worrying about "modes," why cannot we, as teachers and test-makers, place before students in our assignments and in our tests the following specific elements that make up almost any imaginable situation in which writers write: a reason or impulse for writing, a subject, a body of data, a reader or group of readers, and a sense of the action or response desired from those readers? As test-makers...
BRODERICK (continued) were in fact teaching writing.

In saying this, I should also add that these students said nothing about some materials specially designed to help them. When students pick up the packets of reading a month before the exam, they also get a Fact Sheet, a Sample Place-and-Examine and a twenty-two page manual, Nine Elements of Proficiency in Writing, which explains the criteria used in grading the exam. I remember that in 1979 the manual was highly praised by students preparing for the first exam even though the exam was, as I have said, generally resented. Ironically, now that the exam itself has become almost routine and the manual is cited in most Core courses, no student even mentioned it. More about that later.

In any case, the twenty bluebooks showed something else, that students did not view it as part of an attempt to test a special skill or talent. "The exam requires thought," said one who was thankful for the month allowed for "study and consideration" of the readings. Another found that a main theme in the exams and grades in Liberal Arts was a complaint about students' inability to "think critically," their "inability to reason;" she decided that the exam provided a valid way of testing that ability to think critically. Just so, the manual stresses thought; the first five of its admonitions are: 1. Answer the question directly; 2. Support your answer with clear, developed arguments; 3. Show that you have read the source materials accurately; 4. Define key terms where necessary; 5. Focus on the intellectual issues in the questions and avoid irrelevant digression into personal reminiscences and anecdotes. Only the last four of the manual's admonitions deal with language, paragraphs, sentences, grammar, and spelling. All of the bluebooks showed that the students had learned that the exam was designed to test their ability to think in writing.

The few students who expressed their disapproval of the exam also trimmed their opinions. One who frankly said, "I am not crazy about it; so I don't want it," also said that "the impulse to get rid of it" might be an unthinking, "bureaucratic" response to a bureaucratic requirement. Another student saw the exam as part of an attempt to make the University "selective and elitist," for him and for other students the exam was a "psychological obstacle" for "our pursuit of a liberal education." He wanted to eliminate all classes and grades in favor of tutorials and senior exams; but he expressed, too, his worry about students' need to compete in an increasingly technological society. All in all, the criticism of the exam in the bluebooks was limited and, allowing for the fact that bluebooks inhibit protest, comments were thoughtful—certainly a long way from the formulacit sit-in claim that it was discriminatory, elitist, and punitive.

Thus, while the June bluebooks showed some skepticism about the exam, students overwhelmingly accepted it as part of a sensible and desirable college-wide emphasis on writing. Of that elaborate, overlapping, and, to use my son's basketball term, "massive" effort to provide instruction in writing, the students were conscious and properly appreciative. But they were indifferent, naturally enough I suppose, to the design of the testing program. They ignored its commitment to a "subject matter" that minimized cultural and social disadvantages of some students, that is, the test's commitment to a topic adequately presented in readings distributed to everyone who planned to take the test. The

(Continued on page 22)

KEECH (continued) resources they gained from the writing program.

Evaluators can simulate conditions for multi-draft writing without sacrificing test controls by providing two- or three-day test sequences, with discussion and exploratory writing on day one; sharing and drafting on day two; and completing and revising on day three. At the beginning of the term this sequence provides for diagnosis based not only on a finished product but on the whole composing process. At the end, the test is able to reveal which students have learned to evaluate their own composing time to good advantage, in addition to how well students at each ability level have learned to write under conditions more like those allowed for school assignments or self-initiated writing.

Evaluation of portfolios containing both impromptu and multi-draft samples as well as a variety of discourse types provides the ideal portrait of the individual writer and the best data for assessing effects of instruction. But work is only just beginning on quantifying the levels of performance or amount of growth revealed in these rich and complex data collections.

Flaws that Signal Advances

Holistic scoring has been criticized for failing to reveal particular strengths and weaknesses of the writer. Some evaluators have attempted to get the information about their students from mechanical error counts, but these are grossly misleading if errors are not analyzed for their effects on readability or for the level of attainment they signal. Writers are actually uninterested in making some errors and relatively high levels of performance; dangling modifiers cannot appear until the student begins to use participial phrases; failure to mark non-restrictive clauses cannot appear in the papers of students who have not extended clauses at all. The absence of error common in papers of students who merely play it safe, using simple syntax to avoid run-ons, writing only words that are easy to spell, cannot be regarded as signalling advanced competency.

Holistic scores are superior to error counts in assessing how much the balance of strengths and weaknesses affects a reader, but they respond erratically to the uneven nature of development. A pre-test sample, produced when a student is easily able to do all that he attempts, may score high, having few errors and a fluent, easy style, while a post-test sample, produced when he is applying new skills not fully assimilated or is attempting a more complex intellectual task, may score low, showing a fresh rash of errors or awkwardness resulting from a shift in the area of concentration during writing. 4

Post-Holistic analysis of the papers can break out performance on any number of features which may have affected the holistic score, providing important information for diagnosing individual writers or understanding program effects, in some cases revealing changes that were masked by the holistic scores. In addition, work is underway to measure increases in the cognitive complexity of what a student undertakes which may cause unexpected drops in holistic scores. 5 Research examining student texts at the micro-level may help identify which characteristics of student writing commonly regarded as rhetorical flaws may be associated with advances to higher levels of communicative competence.

Meantime, program evaluators need radical new techniques for recognizing and measuring positive

(Continued on page 22)
SCORING THE ESSAYS

For the most part, participants are adapting standard procedures for scoring essays to local situations. For instance, combining holistic scoring with clearly defined evaluation criteria is a common method.

Much discussion focused on the problems posed by attempts to score mechanics in large-scale assessments. Where holistic scoring is used, a general impression of a student's ability to use mechanics correctly is considered when assigning a score. Where analytic scoring is used, a separate score for mechanics is assigned. The National Assessment of Educational Progress scores mechanics by determining a percentage of errors. Scoring mechanics programs problems because a writing task does not necessarily stimulate all students to attempt the same mechanical conventions. In fact, research indicates that better writers will attempt more complex constructions and therefore might make more errors than less competent writers. Therefore, applying a uniform system of scoring mechanics may elicit misleading results.

The logistics involved in scoring large numbers of essays was also discussed. One state agency brought in teachers from all over the state to evaluate essays while in others outside contractors were responsible for scoring. Participants seemed to agree, however, that directly involving the classroom teachers in the scoring process had a favorable impact on instruction.

Time and cost of scoring varied to such a degree that no generalizations can be drawn. Criteria seemed to be the number of students being tested and the method of scoring employed.

KEECH (continued)

program impact on developing writers, being careful not to condemn programs simply because they do not produce large gain scores on short impromptu themes holistically scored.


5 Catharine Keech and Kate Bicklfrman, Competency Testing in Writing: Beyond the Holistic Score, Bay Area Writing Project, Tech. Rpt. No. 3. (Berkeley: University of California, School of Education, 1982.)


Catharine L. Keech teaches composition at San Francisco State University; she served on the Carnegie evaluation of the Bay Area Writing Project and was Research Coordinator for the NIE-funded Writing Assessment Project at the University of California, Berkeley.

SOME GENERAL CONCLUSIONS

A number of questions, both technical and met-physical, remain unanswered. The nature of writing tests which are designed to improve writing, presents a major area of needed research. Very little is known about how carefully norms are determined, how stable the results are, how the results might advantage one group over another, or how accurate their predictive value is. Because competency testing programs are being used to make important decisions about people's lives, it is imperative that administrators of testing programs, as well as the users of the information yielded, scrutinize their programs with care.

The NIE Assessment project with Louisiana Tech was one effort in this direction. A great deal of data about specific programs was gathered in the course of the two-year project which, in the interest of saving space for this publication, we have omitted here. Readers are welcome to write to us at Louisiana Tech for more information.

Michael A. McCreedy is Professor of Education at Louisiana Tech University in Ruston; he has been involved in large-scale writing assessment at both the state and national levels and has recently completed research at the national level for NIE.

Virginia S. Melton is Associate Professor of Education at Louisiana Tech where she teaches graduate and undergraduate courses in elementary language arts; she has served as a consultant, trainer of readers, and analyzer of data for the Louisiana Writing Assessment since its inception.

BRODERICK (continued)

students did not mention, as I have said, the manual's explicit criteria for judging proficiency in writing, nor the easy access to Fact Sheets and Sample Exams. Only a few noted the provision of a month to study the readings and none commented on the provision of three hours to write an essay of at least seven paragraphs. Only implicitly did students show awareness of the exam's emphasis on argument, analysis, and criticism. In short, we would have liked more students to acknowledge what we had wrought.

Indifferent or not to some aspects of the exam, about a thousand students a year have taken the exam, some for the fourth time, since there is no limit on retakes. At any sitting about 30 to 35% fail, but in three years only 71 students have been academically dismissed, and of these, 14 have won reinstatement by passing the exam. In this context, the June 1981 exam books gave us encouraging signs that after three years our Core Curriculum and its Writing Proficiency Exam, which were adopted by a narrow margin and opposed by students, are now accepted, and are working. If we cannot yet say how well they are working, how much the massive effort at teaching writing and administering the Proficiency Exam have in fact improved student writing, we look forward with some confidence to our first full-scale evaluation of the Core Curriculum, the Proficiency Exam, and student writing in the coming year.

James H. Broderick is Professor of English at U Mass/Amherst, where he has served in a number of administrative posts, most recently as Deputy Chancellor; in an earlier post, he instituted a Core Curriculum and a Writing Proficiency Exam.