

Large-scale testing

Editor's note. The following two articles by Robert Lyons and Allan Brick are the second set in the *WPA* series on large-scale testing of writing. The first set, on the California State University English Placement Test by Edward M. White and Alice Brekke, appeared in spring, 1980. In each set in this series, one article discusses a system-wide test from the perspective of the system as a whole, and is written, when possible, by the system's current or past director of testing. The second article in each set explores the effects and implications of system-wide testing at a local campus within the system and is written by a current or past writing program administrator at that campus. Every article in the series attempts both to describe the test and to discuss some of the issues that the test has raised. *WPA* is most grateful to these busy people for undertaking to explore this important topic with us.

III. The City University of New York Writing Assessment Test: A faculty-generated model

Robert Lyons

In November, 1977, the City University of New York asked 10 members of its faculty to accept responsibility for creating a test to measure the writing skills of every student entering the university's 17 two- and four-year colleges. The university's Board of Trustees and administration had already determined that students should be tested in three skills areas—mathematics, reading, and writing. Students who did not pass these tests by the beginning of their junior year would not be permitted to enroll in additional college-level courses until they had satisfied the requirement. Testing was to begin in spring, 1978, for the September freshman class. Students who did not satisfy university standards in math, reading, or writing, would receive skills instruction in that area and would be retested until they met the standard. Because of this retest provision, the university anticipated administering approximately 70,000 writing tests in the first year alone.

The group responsible for the writing test, officially called the CUNY Task Force on Writing, was very much a cross section of the university: six four-year colleges and four two-year colleges were represented. More important, every member of the task force was very much committed to teaching writing: nine of its 10 members were currently teaching freshman-level writing courses; the tenth was an expert in educational testing and measurement. The task force hoped to use its practical experience to create a test that would set appropriate standards of

competency suited to the needs of the university and, at the same time, given the serious consequences of the tests for the students' academic careers, to create a test that would be fair, consistent, and reliable.

Testing writing with writing. The task force made two major recommendations. The first was that the test consist of a single writing sample, produced within 50 minutes, in response to one of two topics. The task force made this recommendation in order to accommodate a university-wide testing program, a complicated schedule of tests and retests on many individual campuses, and a very large number of students. The decision to provide a choice of topic was a way of acknowledging the extraordinary diversity of background and experience that entering students bring to CUNY. The topics themselves were to consist of a short passage on some current subject likely to interest students followed by a set of instructions asking students to agree or disagree with the passage. Students would be asked to develop their essays with personal experience, observation, or information gained from reading. The task force believed that such an assignment, requiring students to assert a position and support it with some evidence, gave the best indication of how ready they were to cope with college writing requirements.

The decision to use a writing sample, and only a writing sample, in the CUNY Writing Assessment Test reflected the unanimous conviction of the task force that proficiency in writing should be measured directly by examining student writing. The task force rejected the alternative of having students demonstrate a more limited range of skills by selecting answers to multiple-choice questions on a standardized test. In addition, the task force expected that an essay test would have a significant influence on the priorities of students, faculty, and administrators of CUNY colleges and New York City high schools. Everyone would become aware that the ability to produce coherent writings of some length was essential to academic progress in CUNY. High school teachers concerned with the improvement of student writing could use the CUNY requirement as further evidence that school administrations should devote additional resources to composition and that faculty should reemphasize the importance of writing as part of the school curriculum.

The second major recommendation of the task force was that all CUNY essays be read centrally by a centrally trained staff of readers. While the university administration accepted the kind of test proposed by the task force, it rejected this second recommendation on both financial and logistical grounds. The task force, therefore, had to develop an alternative plan for scoring the essays, while continuing to urge the adoption of centralized reading. Under the alternative plan, readings were to take place at each CUNY college, and faculty readers were to be trained and supervised by someone at the college who had been trained in the methods and standards recommended by the task force. The degree to which the college readings remained consistent with CUNY standards was to be monitored through an annual university audit of essays obtained by random sample from all the colleges. Readers were to judge the student essays holistically but be guided by explicit criteria reflecting university standards of proficiency.

Evaluation criteria. Far from having completed its work when it created the test, the CUNY Task Force on Writing then had to establish a scale for scoring

the test and characterize the writing abilities each score could be expected to reflect. The group agreed to a six-point scale and required two readers for each essay. Each essay would then receive a combined score of from two to 12. A combined score of eight or better (two readings of at least four on the scale) met the CUNY standard. When two readers differed near the cutoff, a third reader would resolve the conflict. The task force chose this extended scale, more elaborate than needed for a yes-or-no decision determining proficiency, in order to make the test more adaptable to further use by the colleges for placement.

The most difficult and time-consuming responsibility of the task force was to describe and distinguish the levels of writing associated with the different points on the evaluation scale. The group used the term "minimal readiness" to describe the qualities of writing that indicated students were prepared to benefit from the kind of instruction customarily offered in a freshman composition course at CUNY. To define minimal readiness, the task force worked inductively, reading a great many sample essays contributed by the CUNY colleges and pooling their own teaching experience.

The criteria that emerged from these discussions were intentionally kept fairly short and general so that they could be quickly assimilated by readers but would not interfere with the readers' general purpose of rating essays on the basis of an initial holistic impression.' To supplement the scale and to give college faculty some additional preparation for their role as readers, the task force prepared a booklet with sample essays illustrating each point on the evaluation scale; commentary explaining the score accompanied each essay.' Currently, the task force is engaged in developing a writing curriculum model based on successive stages in writing development implicit in the test's evaluation scale. The purpose of this curriculum model is to stimulate discussion of greater integration between curriculum planning and definitions of acceptable writing skills.

Aims, effects, and unsolved problems. The work of the CUNY Task Force on Writing is significant because it suggests that, properly done, the large-scale-testing of writing may have educational value beyond its primary aim of establishing a minimum level of writing preparedness in a large university system. I have pointed out that the CUNY task force was composed mainly of writing teachers, not professional testers. From the beginning, these writing teachers strove to create a test that would lead to wider discussion of writing evaluation, writing standards, and the writing curriculum throughout the university. These efforts have shown some signs of success. Decentralized reading, while making it more difficult to ensure consistency among readers, has brought faculty together for training in holistic reading and has, in many instances, led to a review of local standards of evaluation. The use of the test by some colleges for placement purposes has led to discussions of the relationship among courses in a writing sequence and the skills properly emphasized in each course. Since virtually every CUNY college already offers basic skills instruction in writing, the test is unlikely to prompt additional course requirements. But it may, in some instances, increase the number of students that colleges recognize as needing the existing skills courses. In short, the test has successfully raised the issue of what "minimal readiness" really is in writing for college-level work. In addition, the test has

Notes

served as a focus for contact between CUNY English faculty and New York City high school English department heads who are already preparing their students for a new and demanding writing test recently introduced as a high school graduation requirement by the New York State Board of Regents.

While most CUNY colleges are finding the CUNY Writing Assessment Test useful in a variety of ways, a substantial number of problems remain unsolved. The financial tribulations of New York City and the state have left the testing program markedly underfunded since its inception. The decentralized reading plan has placed a significant burden on English Department and writing program faculty throughout the CUNY system. The differing schedules of the various CUNY colleges have led to a highly complicated calendar for testing and retesting and have produced corresponding difficulties in assuring the rapid and accurate transmission of test data. As a result, some of the needed research on the test results and on the impact of the test has been unduly delayed.

These problems make it difficult to speak with assurance about the long-range influence of the writing test or even to predict the performance of the first group of students facing the proficiency requirement as a condition for continued matriculation this fall. What has undeniably been achieved is a greater consciousness at every campus of the importance of evaluation in establishing writing proficiency, a greater sense of the need for shared standards, and a greater congruence in the way these standards find expression in the writing curriculum.

¹ The CUNY Writing Assessment Test Evaluation Scale:

6: The essay is competently organized and the ideas are expressed in appropriate language. A sense of pattern or development is present *from* beginning to end. The writer supports assertions with explanation or illustration.

Sentences reflect a command of syntax within the ordinary range of standard written English. Grammar, punctuation, and spelling are generally correct

5-4: The writer introduces some point or idea and demonstrates an awareness that development or illustration is called for.

The essay presents a discernible pattern of organization, even if there are occasional digressions.

The essay demonstrates sufficient command of vocabulary to convey, without serious distortion or excessive simplification, the range of the writer's ideas.

Sentences reflect a sufficient command of syntax to ensure reasonable clarity of expression. The writer generally avoids both the monotony of rudimentary syntax and the incoherence created by tangled syntax.

The writer demonstrates through punctuation an understanding of the boundaries of the sentence.

The writer spells the common words of the language with a reasonable degree of accuracy. Exceptions can be made for the so-called spelling "demons" which frequently trouble an advanced writer.

The writer shows the ability to use regularly, but not necessarily faultlessly, the common forms of agreement and of grammatical inflection in standard written English.

3-2: An idea or point is suggested, but it is undeveloped or presented in a purely repetitious way.

The pattern of the essay is somewhat random and relationships between sentences and paragraphs are rarely signaled.

The essay is restricted to a very narrow range of language, so that the vocabulary chosen frequently does not serve the needs of the writer.

The syntax of the essay is not sufficiently stable to ensure reasonable clarity of expression. The syntax often is rudimentary or tangled.

The writer frequently commits errors of punctuation which obscure sentence boundaries.

The writer spells the common words of the language with only intermittent accuracy.

The essay reveals recurrent grammatical problems; if there are only occasional problems, this may be due to the extremely narrow range of syntactical choices the writer has used.

1: The essay suffers from general incoherence and has no discernible pattern of organization. It displays a high frequency of error in the regular features of standard written English. Lapses in punctuation, spelling, and grammar often frustrate the reader. OR. the essay is so brief that any reasonably accurate judgment of the writer's competence is impossible.

² The CUNY Writing Task Force booklet with sample essays and commentary illustrating these criteria is available from the CUNY Instructional Resource Center, 535 East 80th Street, New York, New York 10021.