

WHAT WE KNOW AND DON'T KNOW ABOUT USING MULTIPLE-CHOICE TESTS TO ASSESS WRITING

Since Godshalk's classic study,¹ the amount of substantive research on the relationship between essay tests and multiple-choice tests to assess writing ability has been scarce. We continue to accept or reject the use of multiple-choice tests for reasons other than informed judgment based on solid research data. Indeed, if we were to attempt to decide whether we wanted to use either a multiple-choice test or a writing sample, we would not have all that much research data on which to base our decision. Too often, those selecting tests to assess writing base their selection on factors other than the best method of assessment consistent with the purpose of the testing program. Factors which influence the selection of a writing test include cost (essays cost a lot of money to read), reliability (multiple-choice tests offer high reliability), validity (essay tests offer high face validity while multiple-choice tests offer mixed face validity), time (essays take longer to read while multiple-choice tests can be quickly, even machine, scored), and politics (which group has the most power to influence the selection of the test). Rarely is a decision to select a test based upon the best and most recent research in test theory.

If we were to select a method of assessing writing based on research data, we would be hard pressed to make our selection. Of course, everyone knows that multiple-choice tests cannot assess writing as well as a writing sample, but then everyone also knows it is almost

impossible to assess writing samples reliably. And everyone knows that a writing sample must be at least forty-five minutes long because twenty minutes simply is not enough time for a student to produce enough writing for proper assessment. Then, too, everyone knows that weaker students do poorly on multiple-choice tests because they are bad test takers, and besides, reading ability interferes with their performance. I would point out that all of these preceding statements on testing have little if any basis in research data.

Progress has been made in some ways since Godshalk's study. If nothing else, we have clarified the issues involved in multiple-choice testing. A key finding of the Godshalk study was that different item types produced significantly different results, thus his caution to use "well-designed" multiple-choice tests. We also know that multiple-choice tests can be designed to tap similar skills as those tapped by essay tests, instead of simply testing editing or proofreading skills. We know also that multiple-choice tests tend to assess more reliably the skills of weak writers while not assessing as reliably the skills of better writers. We know, too, that testing must have a purpose, and that the purpose of the test will mandate the design of the test. Thus, a test to identify students who need developmental assistance in writing will be significantly different from a test designed to identify those students exempt from a required writing

(Continued on page 14)

LUTZ (continued)

course. We need to know as precisely as possible what multiple-choice testing can contribute to specific kinds of tests. I am presently conducting research to determine whether carefully designed multiple-choice tests will identify students needing developmental writing assistance as effectively as a writing sample. Preliminary results indicate that certain kinds of item types will identify such students as accurately as a writing sample. The preliminary results also indicate that the same multiple-choice item types do not identify as well as the writing sample students who write at an acceptable level of competence.

We need to know much more about the use of multiple-choice tests to assess writing. Can multiple-choice tests assess the full range of student writing ability? That is, can multiple-choice tests identify the student in need of developmental assistance as well as the student who should receive course credit? What kinds of item types do and do not work? To what extent does reading ability affect multiple-choice test results? Do certain item types work better with poor writers, and, conversely, do certain item types work better with good writers? Can item types be developed which would give some picture of the student's writing process? What new or different kinds of item types can be developed to assess the writing process as we now understand it?

Most of the problems with the design and use of multiple-choice tests can be traced to the lack of involvement in the design and development of such tests by writing teachers. We need now to become actively involved in research that will lead to a better understanding of multiple-choice tests, as well as lead to better tests and better use of test results.

¹ See Fred Godshalk, et al. *The Measurement of Writing Ability*, (New York: College Entrance Examination Board, 1966.)

William Lutz is Chair of the English Department at Rutgers University-Camden and former Director of The New Jersey Basic Skills Assessment Program, a state-wide program designed to assess the reading, writing, and mathematics skills of all freshmen entering public colleges in New Jersey.