

DIRECT VERSUS INDIRECT WRITING ASSESSMENT

Speakers: Evans Alloway, *Educational Testing Service*
Gertrude Conlan, *Educational Testing Service*

Introducer/Recorder: Jean English, *Tallahassee Community College*

Evans Alloway opened the session with the following statements:

1. There is no *good* way to evaluate writing, but we must use the *best* available way to find out how people can write.
2. Writing is a social and political event. We can't separate writing from thinking.
3. Assessing writing is expensive in terms of time, money, energy, and expertise.
4. Those who do not share your biases are neither stupid nor evil.

Alloway went on to discuss the two types of writing assessment: direct and indirect. Direct assessment means scoring a piece that a student has written, for example, an essay. Indirect assessment means scoring the choice that a student has selected from multiple options. In a multiple choice format, it is possible to have many sophisticated questions.

There are three elements of measurement to be kept in mind at all times when selecting a test: reliability, validity, and correlation. Reliability is the accuracy of whatever it is that is produced by the measure. Validity is truthfulness: Can the test provide the information that it was supposed to assess? Correlation is the relationship of performance in one area to performance in another area. Dr. Alloway noted that performance on any good verbal or writing measure will correlate with any other good verbal or writing measure.

Gertrude Conlan advised evaluators to abandon the use of the word "objective" and refer to that type of test as a "multiple-choice" test. The word "objective" makes a pejorative statement. Objectively scored tests are not designed by disinterested persons. Contending that all testing is contrived and artificial, Conlan stated that one can only estimate a student's total performance from one sample. Merely including an essay question in a writing test does not make that test valid. Evaluators must give students the opportunity to demonstrate how well they can perform. In addition, evaluators must increase the reliability of their tests by increasing the number of subtests and the number of raters.

Conlan also stated that multiple-choice tests can give information only at the sentence level. They cannot tell whether a student can express thoughts in his own words.

Conlan's "Steps in Building a Test" are as follows:

- I. Decide preliminary issues
 - A. Money available for building the text
 - B. Purpose of test (scholarship awards, college admissions)

- C. Subject to be tested (listening skills, American history, *Hamlet*)
 - D. Population (age, ability, number)
 - E. Method of testing (paper and pencil, oral, computer console)
 - F. Testing time available
 - G. Types of questions to be used (free-response, multiple-choice)
 - H. Test development time available
 - I. Personnel available to work on test
 - J. Reliability necessary for purpose (the number of items to be given in the time allotted)
- II. Set test development schedule and assign responsibilities for each task
 - A. General tasks in test development
 1. Set specifications
 2. Write items to meet specifications
 3. Review items
 - a. Appropriate content and difficulty
 - b. Appropriate (inoffensive) subject matter
 4. Prepare pretests
 5. Analyze results of pretests
 6. Prepare final form
 7. Review final form
 - B. Related tasks for both final form and pretests
 1. Invite faculty to serve on committee
 2. Schedule meetings
 3. Arrange for pretesting population
 4. Arrange for pretesting population
 4. Arrange for item writers
 5. Arrange for clerical services
 6. Arrange for printing
 7. Arrange for printing of answer sheets
 8. Arrange for special reviewers
 9. Arrange for distribution
 10. Prepare manual for administrators
 11. Arrange for scoring
 12. Arrange for item analysis
 13. Arrange for test analysis
 14. Prepare candidate bulletin
 15. Prepare other bulletins (interpreting scores, explaining test to faculty)
 16. Set schedule for final form administration
- III. Write the specifications
 - A. Purpose of test
 - B. Subject tested
 - C. Timing
 - D. Number of items
 - E. Population to be tested
 - F. Statistical specifications
 1. Desired reliability
 2. Desired mean difficulty
 3. Desired *r* biserial or other index of discrimination
 4. Desired standard deviation (range of difficulty in the items included)
 5. Other special characteristics (no items above middle difficulty, inclusion of items with low *r* biserials for special reason)
 6. Equating needs
- IV. Set content specifications
 - A. Knowledge to be tested and its weighting in the test
 - B. Skills to be tested and their weighting in the test
 - C. Types of items to be used
 1. Number of each type
 2. Skills and knowledge measure by each type
 - D. Other considerations
 1. Level of reading difficulty
 2. Special characteristics of items or passages
 - a. Context suitable to population
 - b. Context from particular kinds of sources
 - c. Subjects to be avoided