

MATCHING ASSESSMENT PURPOSES AND INSTRUMENTS

Speakers: Gretchen Glick, McGraw-Hill, Inc.

Barbara Cole, McGraw-Hill, Inc.

Introducer/Recorder: Ellen Kiba Summy, University of Akron, Ohio

Gretchen Glick discussed four interrelated options that testers and educators must consider in order to match a writing test to the testers' purposes: formats, prompts, scoring, and reporting. When choosing a test, testers must first consider the information they want to obtain, the purpose of the testing, and their resources--both financial and temporal. Indirect assessment of writing, done through an objective instrument, may be adequate and desirable in some circumstances, but most programs use direct measures (a writing sample) for purposes of placement and admissions, research, and proficiency. The ultimate purpose of direct writing assessment is, of course, to improve instruction and to benefit students.

Once the testing purpose is established, testers must develop an assignment or prompt designed to produce the kind of information that they want. Mechanical skills, such as punctuation, spelling, and vocabulary, can be evaluated from a diversity of prompts. However, if the purpose is to evaluate content or organization, all students must be given one prompt, and that prompt must have directions about the topic and the type of writing, whether descriptive, narrative, expository, or persuasive. Then, all papers written in response to a prompt can be compared to a standard--either a model or a set of guidelines. When all students write to the same prompt, their papers can be compared to each other and ranked. In choosing the prompt to be used, testers must consider their program, its texts, and its writing instruction. They should also consider what type of writing is appropriate for the students' ages and abilities. Finally, the topic used should interest students, and the prompt should be general enough to give them something to think about, but specific enough to focus the content of their writing. Sometimes, more than one prompt will be used to provide more information to evaluators; however, the testers' rule is to do the minimum amount of testing that will provide the information needed.

Next, testers must consider scoring options, including the method, the scale, and the location. Although there is no nationwide consensus on the best method for scoring writing samples, the three used most are holistic, analytic, and primary-trait (but even these labels have different meanings to different groups of educators). In choosing a method of scoring, testers must focus on two issues: the testing purpose and the kind of information desired. For instance, holistic scores can be used to evaluate programs and to rank students; analytic scoring can give teachers diagnostic information on individuals' weaknesses so that they can plan classroom instruction. The point scale for rating is also based on the testing purpose. To separate writers into ability levels, a four-point scale might be used: good, acceptable, below average, unacceptable. But to give a more detailed evaluation, an eight- or ten-point scale might be used. The choice of a location for scoring involves the choice of evaluators. With purpose as well as resources in mind, testers could choose professional evaluation with experienced evaluators and computer-monitored, bias-free scoring.

Or they could choose local scoring that brings teachers together to discuss their writing programs and their expectations; these local evaluators would, then, have to be trained as readers.

In reporting testing results, testers must tailor the form of the report to its audience. For example, a summary report will go to administrators at different levels who want to know how groups of students, or schools, or districts compared to each other. Classroom teachers need information on individuals: perhaps holistic data for grouping by general writing ability and analytic data for determining specific instruction. And parents will want an individual report on their child's ability, strengths, and weaknesses.

Because many educators attended this session hoping that McGraw-Hill was developing a direct writing assessment instrument to meet the needs of their school systems, Glick briefly discussed the publisher's new CAT Writing Assessment System, which offers interrelated options in the four areas of testing, prompts, scoring, and reports. For testing, it offers indirect assessment, direct assessment, or a combination. It offers two or three prompts per grade level. For scoring, the assessment system offers both professional evaluation in their Composition Evaluation Center in Monterey, California, and an Administration and Scoring Manual that can be used

by local teachers to score papers. For reports, CAT can produce a variety of forms, including individual records, class roster reports, and frequency distributions.

The second speaker, Barbara Cole, Director of McGraw-Hill's Composition Evaluation Center, began her part of the session by echoing Glick's point that testers should always be concerned with the purpose of a test and by commenting on the fact that tests, scoring methods, and scores can all be used in inappropriate ways. Cole then described how the Composition Evaluation Center recruits, selects, and trains evaluators for direct assessment of writing tests. The trainee program includes theory and training in holistic, analytic, and primary-trait scoring, as well as practice in the appropriate scoring of "validation packets" before trainees are allowed to score "live papers." Finally, table leaders constantly evaluate readers and their scoring to maintain the integrity of the program.

Cole noted that all prompts in the CAT system are two to four sentences long, with only one writing task specified in each. Tests take twenty-five to forty minutes, depending on the grade level being tested. She admitted the possibility of problems if prompts remain the same over the years, and she sees the need to develop and field-test equivalent prompts for the testing system. ■