RESEARCH FINDINGS: A PANEL DISCUSSION

Moderator: Harvey Weiner, Co-Director, National Testing Network in Writing
Panelists: Gertrude Conlan, Educational Testing Service
           Rosemary Hake, California State College, Los Angeles
           Richard Larson, Lehman College

Gertrude Conlan spoke of her past research and the pragmatic experience of ETS, from which a number of clear observations had emerged:

- Every essay measures "an almost infinite universe of knowledge and skills", topics and questions do not equal each other
- There is no formula to produce effective test questions
- Any change in the topic—including what was on TV the night before the test—may change the task, often in unpredictable ways
- A question must be able to be scored. It cannot be too emotion-charged, or too boring for the readers

Finally, Conlan noted, there was typically no clear personal voice in the writing of younger college students; it was present in older writers. James Hoetker pointed out that elementary school students typically wrote with a voice, but that it was suppressed in high school.

Rosemary Hake's presentation focused on her present research on the way teachers read essay exams. She noted that narration is hard to grade because reader sympathy interferes with the evaluation. Hake also discussed the following research findings:

- More abstract topics elicit more consistent readings
- The reader tends to make an aesthetic, intellectual whole of a piece of writing, and otherwise to interpret and add to it

Hake advised that the grading leader should determine "standard grading patterns", individual readers' approaches and the group, and adjust scores to maintain consistency.

In closing, Hake re-emphasized that grading is "safer" (more consistent) when the topic and the writing are intellectual rather than direct or personal.

Richard Larson's presentation concerned future research—specific areas that need examination. He rapidly surveyed recent texts and articles in testing, with special mention of the work of Barbara Ash, Karen Greenberg, James Hoetker, Gordon Brossell, and Leo Ruth. He also called for an assessment of writing done outside of composition courses. And finally, Larson reminded us what the test says to the student-writer, and to the classroom teacher. What does the test writer "hear" when he reads instructions and questions? What does a test tell teachers about what they should be doing in class?

Allan Danzig, Recorder
The City College, CUNY

HOW TO BEGIN A TESTING PROGRAM

Facilitator: Edward M. White, California State College, San Bernadino

Edward White opened the session by outlining five problems or stages which institutions and organizations must move through to develop and implement a testing program. He used these stages, listed below, as his framework:

- Discovering or demonstrating the need for a testing program
- Developing the necessary political strengths and goals
- Choosing or developing a test
- Evaluating the program
- Reporting the scores or results of the testing program

White intended to deal with the five phases sequentially. However, in response to the needs and interests of the participants, he devoted most of the session to responding to questions. As a consequence, the phases were not explored sequentially or in the same amount of detail. Particular attention was paid to choosing or developing a test, and specifically the importance of thinking through what it is you want to test.

Demonstrating a need is the first essential step in beginning a testing program regardless of its purpose or focus. If there is no need to test, do not bother. The need for a testing program can be established in a number of ways, including: statistics on the use of services, the performance of students in courses, and the amount of shifting between classes which takes place.

Choosing a test is extremely complicated; test goals must be clear and the test must be compatible with these goals. Tests are not only devices for ranking people—they send messages about what is considered important and as a consequence will have an impact on curriculum. The message of a test is the most important reason for using a test which requires writing, rather than relying solely on a multiple-choice test.

To be able to defend a competency test at a later date, one must show that it measures what it purports to measure. For example, when designing an essay test, the questions should be designed to test for a particular purpose. Moreover, for a question to work, it must be valid and able to be reliably scored. To do this properly, it is essential that essay questions be pilot tested. The test should exhibit high validity and reliability. It should be fair, proper, and reasonable. Second, it is important that the test be administered properly—students must take the test under the same conditions and within the same constraints. Finally, it is essential that the test be scored properly (controlled readings, with at least two readers scoring each essay).

Although this is not generally the case, the evaluation of a program should start when the testing program is begun. However, in designing the evaluation model, it is important not to fall into the trap of evaluating placement tests in terms of predictive validity (the relationship between scores on the placement test and performance in the course). While White feels that this approach to evaluation is flawed because better instruction can result in lower predictive validity, he closed the discussion by reemphasizing the importance of thinking through what it is you want to test and why because the answers to these questions must be the foundation of any testing program.

Susan Remmer Ryzewicz, Recorder
Office of Academic Affairs, CUNY