TRAINING READERS
Facilitator: Linda Stanley, Queensborough Community College

Participants at this session represented a geographic range from Utah to Paris. Most of the participants came from institutions where they were trained or competency testing programs had been recently instituted and where they, as directors of those programs or as chief readers, were faced with the task of training and/or recruiting faculty for the program.

Chief among the topics of discussion was the problem of training new readers to differentiate between placement and competency measurement. Several of the participants admitted that in their own minds, the two criteria often tend to blend together. The trend of the discussion suggested that before testing programs get under way, a reading time ought to be devoted to developing rangelinders that accurately reflect the tests being used. Several participants also remarked about the importance of punctuating reading sessions with frequent reference to rangelinders, especially in those cases where experienced faculty are readers and where placement is not the only goal. It was thought that experienced faculty would have, as readers, a natural bent to read rather than judge.

Some of the participants new to testing inquired about procedures for training sessions. Some experienced test administrators provided information about the day-to-day tasks involved in training readers to work effectively in a testing program. Above all, readers need motivation, not only in the form of money or released time, but in the knowledge that they are viewed by their colleagues as competent professionals involved in a very serious process. To this end, training sessions ought to be held in an atmosphere fostering mutual respect. Getting college faculty to put their own judgments to the test in range-finding sessions can be difficult. It is very easy for someone who has made a judgment out of line with the consensus to feel humiliated, even when no one intended that result. Therefore, leaders of training sessions ought to be very sensitive to their colleagues’ sense of professional pride.

Training sessions should take place on the same day as a test. Morning sessions, preferably in the morning when readers are alert. Amenities such as coffee and snacks should be provided to keep readers alert and to foster a spirit of group cooperation. In addition to learning how to follow their institution’s technical procedures, readers need to learn to subordinate their personal standards to those determined by the rangelinders. The aim of the person in charge of the training should not be to get an individual to forever abandon his or her personal view of how writing ought to be evaluated, but to temporarily set aside that view in favor of larger common aims. This end, it seemed generally agreed, ought to be the central aim of training sessions.

The role of readers in the generation of rangelinders was also a topic of discussion. Some participants thought that rangelinders should be continually modified over the course of the academic year, both to set better standards and to give the readers a sense of involvement in the direction of the testing program. Other participants pointed to the “floating” standards likely to result from such a practice. Everyone seemed to agree, however, that readers should have a role in determining future topics. A thorny problem for the session, as for most testing programs, was the question of resolving pass/fail splits. Most testing programs represented at the session use the CUNY model where each essay receives two readings. Where there is a pass/fail difference, a third (Continued on page 16)

WRITING ASSESSMENT AND MINORITIES
Facilitator: Paul A. Ramsey, Educational Testing Service

Paul Ramsey opened the session with some general remarks about writing assessment bias. In general, many teachers believe that an essay test is less culturally biased than a multiple-choice test. We know that SAT results have revealed differences of 100 points between the mean scores of black students and of white students. What we do not know is the meaning of these differences. The fact that we know so little about cultural bias in test construction suggests that it has been ignored in terms of research; the real racism is that we have persisted in our ignorance.

While we may test essay tests, we need to be concerned about how they are scored. Most essay tests are read holistically, to get a general impression. Some critics of holistic scoring say it merely assesses “scribal fluency,” and if that is so, dialect features will influence the general impression left with the reader. An interesting area for research is whether primary trait analysis of holistically-scored essay tests would change the scores of minority students.

The group also discussed other issues concerning writing assessment and minorities. Participants wanted to know whether the expository/persuasive modes discriminate against minority students. Many minority students are products of inadequate secondary schools and, thus, may lack information we assume they have. It was suggested that students be allowed to choose among topics in different modes. In addition, testing should be linked to the goals of instruction: if students write expository prose in their courses, then it is fair to test them in this mode.

Various questions were also raised about linguistic minorities. There was no agreement as to whether the exit criteria should be the same for ESL students as it is for native speakers of English.

The group also discussed the problem of responding to Black English features in essays. Several people noted that dialect papers were scored as nonminority papers. The minority students were very sensitive to their colleagues’ perception of their writing.

Competence and tolerance became the key words. As teachers, we are responsible for teaching competence in the language accepted by the larger culture. Many believe that the dialect features that do not interfere with communication are acceptable, those that interfere are not. It is the moral responsibility of teachers to offer instruction in communicative competence, but it is also the moral responsibility of teachers not to make value judgments about students who do not want that instruction. An additional responsibility of teachers is to broaden the definition of what is acceptable language in the general culture and to stretch the limits of tolerance of language diversity. We must help make our students—minority and non-minority—sensitive to language variation and make them agents of change.

It was also expressed that dialect variation was not the key problem in teaching or assessing the writing of minority students. Minority students have the same problems as other students: lack of clarity, lack of organization. We must address these problems.

Finally, Ramsey reminded us that we do not determine what is or is not acceptable language. Furthermore, we must remember that while tests can be used to assess the best methods of instruction, they can also be used to exclude or segregate minority students.

Jane Paznik-Bondarin, Recorder
Borough of Manhattan Community College
Facilitators: Leo Ruth and Sandra Murphy
Bay Area Writing Project

Leo Ruth began the session by reminding us that all writing tasks involve reading problems which the test maker must attend to. Sandra Murphy seconded this by further reminding the group that classroom prompts are negotiable: students may question assignments and teachers, while working to make themselves understood, disclose goals and suggest strategies. Knowledge of the teacher and the context also helps. Writing assignment test prompts, however, must stand on their own.

With this introduction, we then got to criticize a sample real-life prompt. Some of the questions we raised involved content: Would most students know enough to write on this topic? Would they have to worry about offending the unknown reader? Most of the questions concerned the rhetorical mode and structure of the essays elicited by the prompt. The prompt was criticized for calling for three tasks: Is this too many for ease of students' performance or coherence of the product? What variations of emphasis or organization of the tasks are tolerable? If the prompt says, "Describe a problem," are we asking for a description? How will exam readers react to narrative writings which students find easier and do better, if the readers are expecting argument? These and other questions called for agreement between test makers and graders relating the nature of the prompt to the purpose of the test and an assessment of the abilities of the test-takers. As an example of the first consideration, do we want typical writing or good writing and, if the latter, what kind of good writing? As an example of the second, are our test-takers sophisticated enough to distinguish between "real" questions and "school" questions and to respond with appropriate behavior? One participant told of a returning adult student who refused to respond to her prompt on the ground that she believed the teacher knew the answer to her question.

Many of the participants, particularly those from CUNY, were interested in the discussion of "agree-disagree" prompts (CUNY uses this format for its Writing Assessment Test). Though research suggests the superiority of this type of prompt, participants pointed to problems. The phrasing of the question may give an edge to one side, since the question offers suggestions on what to say. For the insecure writer, agreeing generates more verbiage, while disagreeing, which permits and even enforces a narrower focus, may generate better essays from those for whom verbiage generation is no problem.

Besides sample prompt, participants were given two other handouts: "Guidelines for Developing Topics for Writing Assessments" and "An Evolving Model for Studying the Writing Assessment Episode," the latter detailing relationships among participants (test-maker, test-taker, and test-rate), processes and products. The handouts turned out to have anticipated many of the participants' questions and concerns. Additional handouts can be obtained from Professor Leo Ruth, Bay Area Writing Project, University of California, Berkeley.

Charles Pitch, Recorder
John Jay College

POLITICS OF TESTING, (continued)

the test. Have faculty generated the criteria for grading a test, then have them participate in grading. Realize that there is significant public support for testing. One participant described a junior-level reading/essay test which avoids the pass/fail grade and asks instead, "What will this student's program for the next two years look like?" Students at this institution also have a chance to revise their essay the next day, and faculty from all disciplines are involved in grading the tests and counseling students. Are we observing a natural cycle that begins with public alarm at poor writing and that leads to a demand for testing, new courses, refinement of the test, and ultimately an impetus toward programs in writing across the curriculum? Although some participants felt that the notion of a cycle accurately describes recent developments on their campuses, a participant from a SUNY campus argued that there is too much rigidity in the system to permit the cycle to occur. Another participant pointed out that if a movement toward writing across the curriculum is to develop, it must be actively promoted from the very top, preferably by the president of the college.

Hendrix remarked that the discussion confirmed his hypothesis that testing has been a divisive issue on campuses. On the one hand, assessment has been used to get students, especially adults, into education, to develop alternative routes for education, and to educate by defining competencies. On the other hand, public outcry for a return to basics, mandated state-wide or city-wide tests, and new research interests in testing has lead to more and more tests, some of which screen students out of education. Thus, he concluded, we stand at an interesting and complex moment in the politics of testing.

Nancy Black, Recorder
Brooklyn College

TRAINING READERS, STANLEY, (continued)

reading decides the issue. Some participants, mindful of budget constraints at their institutions, wanted to know how to keep third readings to a minimum. Others wanted to know who should do the third readings? Should chief readers have to shoulder this burden in addition to their other administrative tasks? Should special, experienced, "third readers" be designated (or elected)? What should chief readers do when they "know" that a third reading is incorrect? Should some of the third readings be discussed and reviewed at some point in the testing session? What procedures are used for dealing with consistently aberrant readers?

There was a general consensus that many of the problems associated with essay readings could best be resolved by developing good training sessions for readers. Such sessions should be repeated each time a reading is held, and readers should be required to attend them. The general aim of the test administrator should be to impart to the readers a feeling that they have an active input into the testing program. Testing sessions should not be too long; should take place in a pleasant, collegial environment, and should be oriented to reader participation. The ideal goal should be to make "training" a continuing and natural, organic part of every reading, for everyone involved.

It would seem from the geographic range of the session's participants that testing programs throughout the United States (and elsewhere) are likely to grow rapidly in the near future. This expansion will be especially interesting for us at CUNY as many of these programs are based on the CUNY model. This session indicated that in the future we will certainly be able to learn more from each other.

Harry Brent, Recorder
Baruch College