critique of both the entire agenda of holistic scoring and of the specifics of assessment design. But it also allowed Murphy and Ruth format in which to report some of the findings from their work.

The six questions treat variously the syntactico-semantic structure of the items, the discourse structures suggested, the power relationships established between test(er) and writer, and the cultural knowledge presupposed. The six questions and comments from the presenters and audience are as follows:

1. How much information should be provided about the subject?

Murphy and Ruth’s findings suggest that a simple referring phrase (NP) elicited less rich responses than a full proposition. When a predicate was provided, writer responses were more “reasonable and responsible.”

2. How does specification of a subject constrain response?

Discussion demonstrated the range of possible constraints: discourse type, qualification, quantification, text structure, style, and—always—ideology, explicit and implied.

3. How does knowledge of the subject affect performance?

The session members soon raised the meta-question of whether any topic could not require "specialized knowledge," and therefore whether holistic essay testing could be free from political bias. Generally, Murphy and Ruth and the session members agreed that knowing a lot about the topic was a great advantage, and the “knowledge” extended well beyond simple propositional knowledge to familiarity with cultural discourses.

4. Should students be given options in selecting topics?

Generally, options invite confusion. Items may not be equally difficult. Students may not be wise in selecting, picking complex topics and writing complex, bad essays. Confusion over the selection process may penalize.

5. How do rhetorical specifications affect performance?

Students did not seem to be helped by suggestions of rhetorical type. Typically, they ignored them or found that the problem of executing the rhetorical command interfered with their writing in general.
(2,4,6) unspecified in order to give the raters greater flexibility. The essays were also scored analytically according to ten items in three categories: (1) development, (2) voice/speaker/persona, and (2) readability.

Although the original plan had been gather samples from extreme student populations (high- and low-ability), differences between institutions in the average quality of student writing were noticeable: many "low-ability" students wrote as well as or better than students ranked as "high-ability." As a result, the sample fell into a bell-curve distribution. The research concluded that there is no evidence from either the holistic-scale scores or the analytic-scale scores that even gross variations in phrasing affect either the quality of student responses or the nature of student-topic interaction. Other conclusions: the appearance of first-person voice is significantly higher in essays written in response to topics calling for accounts of personal experience, but it is unaffected by the degree of rhetorical specification.

In a discussion following the presentation of the research, Brosell and Hootker mentioned plans for future work that include a study to evaluate the effect of content variation in essay topics when wording and rhetorical specification are held constant. They also plan to develop their analytic score further, based on additional essays written at greater leisure and revised, and representing average and high-ability students as well as low-ability students. With revision and development to make the scale reliable and "transportable," the analytic scale might, according to Brosell and Hootker, have the potential to become an alternative to the single-digit holistic score.

WHAT SHOULD BE A TOPIC?

Speakers: Sandra Murphy, San Francisco State University, Leo Ruth, University of California, Berkeley

Introducer/Recorder: Robert L. Brown, Jr., University of Minnesota

Taking a cue from the Bay Area Writing Project's collective spirit, Sandra Murphy and Leo Ruth rejected the usual panel format by opening the session to audience discussion of issues influencing subject-selection for holistic scoring. They directed the session with six questions (treated at greater length in their recent ABLEX book Describing Writing Tasks for the Assessment of Writing). Their questions examined the dual problem facing assessment designers: naming a subject and providing the writers with instructions about what to do with it. In part, the session provided a forum for a critique of both the entire agenda of holistic scoring and of the specifics of assessment design. But it also allowed Murphy and Ruth format in which to report some of the findings from their work.

The six questions treat variously the syntactic-semantic structure of the items, the discourse structures suggested, the power relationships established between test(er) and writer, and the cultural knowledge presupposed. The six questions and comments from the presenters and audience are as follows:

1. How much information should be provided about the subject?
   Murphy and Ruth's findings suggest that a simple referring phrase (NP) elicits less rich responses than a full proposition. When a predicate was provided, writer responses were more "reasonable and responsible."

2. How does specification of a subject constrain response?
   Discussion demonstrated the range of possible constraints: discourse type, qualification, quantification, text structure, style, and--always--ideology, explicit and implied.

3. How does knowledge of the subject affect performance?
   The session members soon raised the meta-question of whether any topic could not require "specialized knowledge," and therefore whether holistic essay testing could be free from political bias. Generally, Murphy and Ruth and the session members agreed that knowing a lot about the topic was a great advantage, and the "knowledge" extended well beyond simple propositional knowledge to familiarity with cultural discourses.

4. Should students be given options in selecting topics?
   Generally, options invite confusion. Items may not be equally difficult. Students may not be wise in selecting, picking complex topics and writing complex, bad essays. Confusion over the selection process may penalize.

5. How do rhetorical specifications affect performance?
   Students did not seem to be helped by suggestions of rhetorical type. Typically, they ignored them or found that the problem of executing the rhetorical command interfered with their writing in general.
6. To what extent should admonitions about the writing task be mentioned? Time limits, pitfalls, and so on?

Again, the political demands of the writing assessment as an institution overwhelm the testers' attempts to help: students write the essay they have in mind, ignoring the instructions or finding themselves confounded by them.

The session eloquently expressed reservations about the ideology of holistic scoring and mass assessment in general. The conferees reacted to the inherent artificiality of pretending to write authentic prose while authentically demonstrating familiarity with academic conventions. They agreed that students who know the conventions of testing will, predicably, do best.

CLASSROOM RESEARCH AND WRITING ASSESSMENT

Speaker: Myles Meyers, California Federation of Teachers
Introducer/ Recorder: Deborah Appleman, Carleton College, Minnesota

Myles Meyers addressed the issue of large scale assessment from the perspectives of the K-12 administrator and classroom teacher. From these perspectives he finds large scale assessment to be problematic and often ill-advised. The enormous diversity of schools makes it difficult to capture the current "state of the art." Myers also contended that state assessments such as California's CTBS work against teaching as well as against the professionalization of teachers.

Meyers discussed at length the seemingly reductionist quality of large scale assessment. Although recent research on writing maintains that writing is a multiple construct, time and financial constraints limit the constructs that can be examined. The construct that is employed to define writing thus becomes the primary focus for a particular grade (for example, autobiography in grade 10). In our effort to handle the assessment task by limiting constructs, our definition of writing, as well as its instruction, therefore becomes uni-dimensional. Moreover, because of the inevitable prescriptive quality of

the interpretation of assessment results as well as teachers' lack of involvement and consequently lack of ownership in the entire assessment process, Meyers claimed that statewide assessments can destroy teaching-as-inquiry and harm student learning.

Meyers then presented several suggestions for involving teachers in the assessment process. He emphasized the importance of having teachers participate significantly through summer institutes at university settings. He also underscored the importance of viewing assessment as a process of inquiry, one in which disagreement is as important as agreement. To illustrate the value of assessment as inquiry, Meyers handed out three sample student papers and asked the audience to rank them as high, middle, and low. The resulting scoring was quite discrepant, as were the reasons offered for the rankings. Meyers then discussed the value of discrepancy in our aim to improve literacy for all children. Rather than considering agreement as the ultimate goal in assessment, discrepancy can lead to a fruitful dialogue about our underlying assumptions about teaching good writing as well as about its evaluation.

Meyers pointed out that dialogues or debates such as those generated by the conferees when they were asked to rank the papers were a critical aspect of the assessment process. He stressed the importance of having classroom teachers as active participants in an on-going debate on assessment, rather than as recipients of an administrative decision to employ a particular large scale assessment instrument. He then handed out six additional student papers, and asked conferees to rank them and then to discuss the rankings in pairs. As with the first exercise, the rankings were widely discrepant. Meyers illustrated how this kind of exercise can be used to encourage teachers to think explicitly about their pedagogy and also described several ways in which the ranking of student writing can be employed to generate discussion among teachers. For example, he has asked teachers to devise sample lessons for students whose papers they have ranked.

Meyers ended his provocative discussion by suggesting several ways in which writing can be viewed as a speech act and as a collaborative social event. He discussed the differences and similarities between conversation and written presentation. Meyers concluded her talk with the following thought: "When you teach people how to write, you teach them a new definition of themselves."