consequences, we have made and are making decisions which involve ethics and responsibility: to test or not to test; to use multiple choice or writing sample tests; entrance versus graduation requirements; how much writing across the curriculum emphasis; how much remediation; and how to establish effective dialogue with the lower schools.

The process is heuristic—and we are still in it. We who profess the value of heuristic learning need to lobby hard for shared responsibility rather than monolithic decision-making. Writing problems cannot be adequately addressed by the English Department or the Writing Centre alone; only a recognition that all faculty at all levels share the responsibility for improving student writing will enable that to occur.

AN EXAMINATION OF HOLISTIC SCORING: RESEARCH AND THEORY

Brian Huot and Sydney Smith, Syracuse University
Penny Pence, University of Pittsburgh
Michael M. Williamson, Indiana University of Pennsylvania
Kathleen Hunter, College of Saint Elizabeth

Because faculty view holistic scoring as the most economic, efficient, and flexible of all direct writing assessment procedures, it has become the most popular method for assessing student writing ability. Proponents of holistic scoring have claimed that most of the basic reliability problems inherent in direct writing assessment have been solved. However, increasing concern about the validity of holistic scoring procedures has been emerging in the last couple of years.

The panel presented a review of theory and research on holistic scoring and reported an ongoing research study of the validity of holistic scoring. The review aimed to place holistic scoring in a theoretical framework of educational measurement. The goal of the research study is to provide some validation of holistic scoring by examining the differential responses of expert and novice raters through the use of verbal protocols. Interestingly, expert raters appear to achieve a more personal reading and to be able to read from a much wider range of stances than do novice raters.

LARGE-SCALE PORTFOLIO ASSESSMENT IN THE RESEARCH UNIVERSITY: STORIES OF PROBLEMS AND SUCCESS

Chris M. Anson and Robert L. Brown, Jr., University of Minnesota

We reported on the three-year progress of an institutionally-mandated assessment program at Minnesota. The first two years of the program seemed unremittingly positive. In response to the university administration's call for a mandatory admissions testing program—as a means to cull and limit the student population—the Composition Program had effectively redirected efforts into a principled portfolio assessment for entrance and graduation. More significant than the program itself was the change in faculty attitudes brought about by a massive reeducation effort on the part of composition professionals acting as change agents. But major financial, administrative, and programmatic confusion in the University in 1988-89 virtually canceled the portfolio project.

That these changes were simply announced without consulting the Composition Program indicated that the ideological forces initiating the original regressive testing mandate remain undiminished. We detailed the ethnographically oriented process through which we reexamined the place of writing and assessment in their institutions. Working by a collaborative series of stories and analyses, we showed the complex of beliefs about writing which characterize many research universities and which must be well understood by program designers. Using theoretical systems from ethnography and credit-cycle economics, we suggested that faculty members' interests drive educational and curricular policy. To succeed, writing policy must not conflict with those interests. Since a major university is a conglomerate of small working units, policy makers must proceed locally and specifically—from the bottom up. The labor- and time-intensive program of portfolio assessment directly threatened the reward system of the university at large. But at more local levels,
fewer conflicts emerged. Recent history of the Minnesota program suggests that assessment and writing across the curriculum projects with a local focus have an excellent chance, where massive programs mandated from the top down are deeply problematic.

---

ASPECTS OF HOLISTIC SCORING VALIDITY

Victor Froese, University of British Columbia

This session focused on a study of the validity of holistic scoring conducted by researchers at the University of British Columbia. Three related observations prompted this study: 1) insufficient research has been done as to whether readers trained in holistic rating base their judgments on substantive or superficial characteristics; 2) compositions can be scored more quickly by computer than by trained markers; and, 3) the validity of holistic scoring has not been satisfactorily demonstrated.

This study compared holistic scores with composition length, number of spelling errors, and sentence length—the first two being distractors frequently associated with holistic scoring. For each of grades 3, 6, and 9, teachers scored forty randomly selected narrative and forty randomly selected expository passages holistically and via the Writer’s Workbench software (which scores many other features as well). A graduate research assistant entered the same compositions for computer analysis, being careful to encode the text and spellings exactly as in the originals.

Findings from statistical analyses indicated that grade-by-type interactions existed for holistic scores, sentence length, and passage length but not for spelling errors. Subsequent analyses indicated that holistic scores predicted sentence length for Grade 3 narrative and expository text; passage length for narrative text in Grades 3, 6, and 9; and expository text in Grades 3 and 9. Spelling errors could not be predicted from holistic scores. When sentence length, passage length, and spelling errors were combined, narrative text scores could be predicted at Grades 6 and 9 but expository text only at the Grade 9 level. This implies that some of the mechanically counted features (sentence length, passage length) of the Writer’s Workbench predict holistic scores in a statistically significant manner (except for expository text at the Grade 6 level).

The researcher concluded that holistic scoring is apparently sensitive to the sometimes irrelevant factors such as sentence length and passage length, but that these operate somewhat differentially at different grade levels and for different types of writing. Also, the parsimony of computer-based composition scoring should not be overlooked since it is accurate, less costly, and faster than human-based holistic scoring. Finally, because the trained observers’ scores were not related to misspellings but to sentence length and passage length, the validity of holistic scoring still needs further investigation in order to explain this relation.

---

COLLECTING EVIDENCE IN SUPPORT OF CONSTRUCT-RELATED VALIDITY

Sybil B. Carlson, Consultant, Educational Assessment and Research

One category of evidence for validity, that of construct-related evidence, is particularly critical to tests of writing skills in which readers assign scores to papers written by examinees. Because the direct assessment of writing relies on human judgments, its quality greatly depends on the validity of the perceptions invoked by individuals when making these judgments. Thus it is essential to focus on the process evaluation—a creative act of interpretation that is influenced not only by the perceptions of readers but also by the perceptions of the various audiences involved in the assessment.

The construct of writing competence must be described and defined within the context of a specific assessment. Furthermore, since readers are expected to apply that construct objectively and systematically in a testing situation, evidence must be collected to support inferences that the intended construct actually is being applied—by readers as well as persons who interpret and use test results.

Attention must be directed, not only to the mechanisms necessary for the implementation of the program, but also to the construct that is being