
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 explanatory 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 explanatory text; passage length for narrative text in Grades 3, 6, and 9; and explanatory 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 explanatory 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.