MAINTAINING SCORE-SCALE STABILITY AND READING RELIABILITY IN DECENTRALIZED LARGE-SCALE ESSAY SCORING PROGRAMS

Wayne Patience and Joan Auchter, GED Testing Service

A central aim in any writing assessment program is to ensure score-scale stability and reading reliability across essay scoring sessions. In large-scale writing assessment programs, the essay scoring is often necessarily decentralized; that is, the scoring sites vary in location, personnel, frequency of scoring sessions, etc. In such large-scale essay scoring programs, it is essential to construct training, certifying, and monitoring procedures that are designed to establish scoring standards and ensure the consistent and reliable application of these scoring standards across scoring sites and scoring sessions.

The revised Tests of General Educational Development (GED) that were introduced in 1988 include a direct measure of writing as part of the Writing Skills Test. Because the Writing Skills Test scores, along with the scores on the other four subject-area tests in the battery, determine whether an examinee receives a high school equivalency credential, the reliability of the weighted composite score based on a combination of the essay and multiple-choice parts of the Writing Skills Test is of major importance.

This presentation described the methods of establishing and monitoring score-scale stability and reading reliability and reported results showing how well these methods succeeded in the first year of the GED Testing Program's decentralized essay scoring operation. Several GED Testing Service scoring sites participated in monitoring two types of essay-scoring. Random monitoring was conducted by randomly sampling and reading essays that are scored at the decentralized sites and routinely sent to the GED Testing Service. Systematic monitoring was conducted by requiring sites to read and return previously scored and selected essay sets provided by the GED Testing Service.

Results showed that the procedure for establishing and maintaining score-scale stability are effective. In the minority of sites wherein some scale drift or inconsistency among readers is present, monitoring reports sent to site chief readers provide diagnostic information that shows the nature of the problem and specifies corrective action.

WRITING ASSESSMENT IN QUEBEC: THE STATE OF THE (LANGUAGE) ARTS

Tom McKendy, Marianopolis College, Montreal Allan Patenaude and Beverly Steele, Quebec Ministry of Education

This session began with a description of the integrated secondary Language Arts Curriculum prescribed in 1982 by the Quebec Ministry of Education (M.E.Q.). For the first time, teachers throughout the province faced a rationale and methodology of teaching English predicated on the principles that language serves as an agent of communication and conceptualization. In 1990, the common provincial exam for graduating students will be based on this new curriculum.

This common exam has served as a catalyst to the province-wide discussion of several major issues: Should examination questions be linked directly to specific works? Should talk occur during the examination? Should rough drafts and outlines be graded as part of the examination? Each of these issues is currently being addressed by the M.E.Q. through in-service training and the development of evaluation models.

Over the past several years we have used both college English grades and midterm ratings of writing ability by college English teachers as standards against which to measure the effectiveness of various methods of placement in writing courses. Although the research was originally intended to compare the relative reliability of a holistically scored writing sample and a multiple-choice test, statistics on secondary school records were also gathered. Recent data analyzes indicate that the correlations between the secondary school grades and college standards have risen dramatically.

Although it is premature to conclude that the improved correlations are the result of the new secondary school curriculum, the data encourage further testing of this hypothesis.
However, it seems that secondary school English grades (based in part on the evolving provincial exam and in part on the teacher’s assessment of work throughout the year), are more valid measures of writing ability than current one-shot placement tests.

**FIRST- AND SECOND-LANGUAGE WRITING OF FRENCH IMMERSION STUDENTS**

*Marie-Josée Vignola and Marjorie Bingham Wieche, University of Ottawa*

Our presentation provided a review of the research literature on L1 and L2 writing of French immersion students, and reported the results of a pilot project with such students in first-year university studies. Over the past two decades, various forms of "immersion" have become a prominent approach to teaching second languages, particularly French, in Canadian schools. In immersion programs, children whose home background is English (L1) are taught some or all of the regular school curriculum through the medium of French (L2). Popular starting points for these programs are kindergarten, the middle elementary grades, and grades 6/7. Many students continue their high school studies in bilingual programs. Research has consistently shown that students completing immersion programs as a group catch up to, or sometimes surpass, English program students in English language arts and subject matter courses, and that they achieve high, if not native-like, levels of functional French skills.

Most research to evaluate the impact of immersion education on students’ writing skills has focused on French (L2) writing skills, including the influence of L1 on L2 writing. Some research has also been done on L1 writing skills in early immersion students. Several studies have demonstrated that students after 3 to 4 years in early French immersion not only catch up in L1 skills but may surpass English language control groups in certain areas (e.g. vocabulary, grammar, punctuation, sensitivity to grammatical structures) despite the fact that the immersion students have had more limited training in their L1 than students enrolled in the regular English program.

To our knowledge, all research to date on immersion students’ writing skills has compared immersion students with English instruction or French L1 control groups in order to evaluate their L1 and L2 language skills. We have found no previous research which compares the L1 and L2 writing skills of the same subjects. As ever-increasing numbers of former immersion students university age, the question of the effects of immersion education on high-level first language writing skills is frequently posed.

**HOLISTIC AND PERFORMATIVE ASSESSMENT OF ESL WRITING**

*Ulla Connor, Purdue University*  
*Elaine Fischer-Kohn, San Francisco State University*

This session focused on two types of assessment of ESL writing: (1) large-scale holistic scoring for pre- and posttesting for program assessment purposes and (2) "performative assessment" for assessing specific writing skills as a diagnostic tool for classroom instruction. First, we described a new tool for assessing specific writing skills: performative assessment, developed by Faigley et al. at the University of Texas at Austin. We explained what a performative assessment instrument is (a set of rubrics, or evaluation guidelines, describing specific levels of proficiency for several general skills required by the writing task); how it differs from other criterion-based assessment tools such as primary-trait scoring; and how it is developed. We discussed the practical applications of performative assessment, both in diagnostic testing and as a tool for classroom instruction.

Next, we explained how holistic scoring, used effectively in native-speaker composition testing, is currently being adopted by schools and testing companies for non-native speaker (ESL/EFL) composition testing. Holistic scoring enables evaluators to identify high- and low-level writing efficiently and reliably. However, holistic scoring fails to differentiate between different qualities of middle-range ESL writing.

Our research was conducted on immigrant university freshmen in a year-long ESL program.