

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.