## ESTABLISHING AND MAINTAINING SCORE SCALE STABILITY AND READING RELIABILITY

**Speakers:** *Wayne Patience*, GED Testing Service
*Joan Auchter*, GED Testing Service

**Introducer/ Recorder:** *Anne Aronson*, University of Minnesota

Wayne Patience and Joan Auchter presented the procedures used by the General Education Development Testing Service (GEDTS) to evaluate essay exams required as part of the GED Test for individuals seeking high school equivalency diplomas. They described and illustrated the methods employed by GEDTS to establish and maintain stability or consistency of scoring, and reliability among readers, despite the decentralized nature of their evaluation program.

Patience explained that the notion of equivalency derives from: (1) defining the content of the GED Tests so as to reflect the community expected outcomes of completing a traditional high school program of study and

(2) defining passing scores relative to the actual demonstrated performance of contemporary graduating seniors. Only those examinees who receive scores that are better than 30% of high school seniors are awarded the diploma. The job of the GED staff is rather to <u>describe</u> the skills and content knowledge that characterize the work of high school seniors than to <u>prescribe</u> levels of achievement.

The recent addition of an essay exam to the Writing Skills Test created questions about how to establish and maintain reliability. The GEDTS's first activity was to develop a scoring scale that would have the same criteria regardless of time or place. By administering an essay exam to thousands of high school seniors, sorting those essays into six stacks, and describing the characteristics of each stack, the Writing Committee of GEDTS was able to develop a holistic scoring scale that has been used successfully in hundreds of sites nationwide.

Auchter then reported on how GEDTS insures stability and reliability in the use of the scoring guide. A permanent GEDTS Writing Committee, consisting of practicing language arts professionals, selects the topics and the papers that are used in training, certifying, and monitoring site trainers and readers. The Writing Committee chooses and tests "expository" topics that do not require students to have any special knowledge or experience. The next step is for GEDTS to train and certify Chief Readers who are responsible for insuring that the GED scoring standards are applied uniformly. During the 2 1/2 day training, Chief Readers learn to overcome personal biases (e.g., responses to handwriting) that may influence scoring, and to use the language of the scoring guide alone to describe and evaluate papers. Sets of training papers contain a <u>range</u> of papers for each point, to illustrate the fact that there is no "perfect" paper for each point, but that there is typically a distribution of high, medium, and low papers. Training packets also include problematic papers (e.g., a paper written in the form of a rap song). Since the national average for high school essays scores is 3.25 and for GED scores is 2.7, training sets contain a disproportional number of 2, 3, and 4 papers. After working with training papers, GEDTS trainees are required to evaluate several sets of papers to determine whether or not they are currently certifiable as Chief Readers.

The same training and certifying procedure is carried out at the various decentralized testing sites, with the Chief Readers responsible for training and certifying readers. Auchter noted that language arts teachers trained through this process feel better about teaching writing and about using holistic grading in the classroom.

Further steps to insure score scale stability and reliability are site certification and monitoring. Each scoring site must demonstrate the ability to score essays in accord with the standards defined by the GED Testing Service. Essays used for site certification must receive at least 80 percent agreement in scoring among Writing