The two studies described in this session deal with the questions "How does students' writing change over time?" and "How can writers establish coherence for their readers?". Leo Rigsby of Temple's Sociology Department spoke first. For his study, he analyzed writing samples from three points in the undergraduate careers of 382 students: a placement essay written before the beginning of the first year, an essay written as part of the final exam in Freshman Composition, and a sophomore level proficiency exam. Graders were carefully trained to score essays, and training occurred in several sessions over a period of a week to establish criteria for marking errors and to ensure uniformity in evaluation.

Any study of this kind, Rigsby noted, is beset by great difficulties. In his, for example, only 92 of the original 382 students actually completed all three writing tasks, thus reducing the sample size and opening the possibility of bias in the results. Moreover, although the first and third writing samples were controlled by a writing program administrator and thus were homogeneous, the second was controlled by course instructors and was much more heterogeneous. The three writing tasks were, therefore, not fully comparable. Finally, because evaluative criteria had to be subjective in the extreme—for example, whether the thesis was developed by appropriate logical methods, whether needed definitions were omitted, whether unnecessarily blunt or indirect statements were made—the opportunity for inconsistency in evaluation was always present, especially because the same readers were not available to read all three sets of essays.

Rigsby's analysis showed some evidence that fewer errors appeared in the proficiency essays than in the placement essays, suggesting that writing improves over time. To avoid the potential bias caused by the fact that writers become more "selected" over time (those who wrote the second and third samples had already survived in the university for a year or more), he looked at the fixed set of students (i.e., those who took part in both the first and the last writing tasks). Comparison of writing for these students revealed that error rates did not decrease but instead actually increased slightly. When these results were examined separately for men and women, they showed a reversal in gender differences over time. Whereas women made fewer errors than men on the placement essays, men actually made fewer errors on the proficiency exams. In interpreting the results, Rigsby noted that the pattern of increased error rates is consistent with other researchers' observations that old errors reappear in students' writing as the writing tasks become more difficult. As for the differing error rates in men's and women's writing, he speculated that these patterns of change reflect persisting differences between men and women in socialization patterns and socially constructed goals for which education may be more or less relevant. The pattern is also consistent with conclusions from previous research that women exhibit less change in response to many kinds of educational tasks than men do, because of women's greater tendency to do as well as possible on educational tasks from the beginning.

Frank Sullivan, of the English Language Enrichment Center at Temple, then discussed his study of the ways in which readers respond to writers' attempts to establish coherence in their placement-test essays. Sullivan found three statistically significant patterns of interactions between writers and readers: first, readers evaluated positively the essays of writers who "over-identified" information by including details they had reason to think were redundant for readers. Sometimes this happened because the writers had to identify information for two sets of readers with different amounts of knowledge about the topic—the readers named in the topic question and those who actually read the essays. For instance, reference to "the School Board of Emmaus High School" would be clear to the fictitious readers in the assignment but not for the people actually reading the students' papers. In other cases, over-identification seemed to function as a way for readers to establish writers' identities as students who know things. For example, although a reference to "Oedipus Rex" or "the Russian Revolution" would suffice to identify what the writer meant, readers responded more positively to "Sophocles' 'Oedipus Rex'" or "the Russian Revolution of 1917." Sullivan also found that readers reacted positively to writing which contained references to things that could have been more substantively and concisely identified, such as "anything to do with books that contain foul language," "some sort of program," or "a book report following questions written up by the school." This kind of representation seemed to function as a way for readers to identify with writers who could act as if they had something to write about, even when they didn't.


Speakers: Richard Swartz, GED Testing Service
Wayne Patience, GED Testing Service

Introducer/Recorder: Karen Greenberg, National Testing Network in Writing

Richard Swartz began the session by asking the following question: How do you design an essay test that can be administered virtually any day of the year and scored—according to the same standards—in one of several hundred locations throughout the United States and Canada? This is the question that the researchers who developed the new essay section of the GED Writing Skills Test had to answer. Swartz noted that the current Writing Skills section of the GED includes multiple-choice questions intended to measure a person's ability to use Standard Written English clearly and effectively. The essay section of the GED test will be
administered beginning in 1988. The GED created a Writing Committee whose responsibilities included reviewing and evaluating potential essay topics and selecting rangeland developers and training papers. The Committee gave considerable attention to selecting the best possible method for scoring the writing samples obtained from high school students, and holistic scoring was selected as the optimal method. Swartz pointed out that holistic scoring is a quick and relatively simple procedure, so it is an efficient method for scoring large numbers of essays. In addition, holistic scoring offers the most accessible method of attaining high degrees of inter-rater reliability, and these were necessary in order to answer the primary research question involving the correlations between essay and multiple-choice scores. In order to control the drifting of standards from one reading to another, a scoring guide was developed for each topic. Because the guides were created inductively—by describing the common characteristics of agreed-upon model papers—they could be regarded as descriptive rather than prescriptive.

Wayne Patience then discussed the evaluation of potential essay topics. He noted that in order to minimize the problem of controlling topic variability, it was decided to restrict GED essay topics to one type—expository. A sample topic is as follows (the time allotted is forty-five minutes):

The automobile has certainly been responsible for many changes in the United States. Some of these changes have improved our lives and some have made life more difficult or unpleasant. Write a composition of about 200 words describing the effect of the automobile on modern life. You may describe the positive effects, the negative effects, or both. Be specific, and use examples to support your view.

After potential topics are reviewed and edited by the GED Writing Committee, they are field-tested in administrations to both high school seniors and GED examinees. If topics survive these evaluations, statistical evaluations are performed (i.e., a Kruskall-Wallis test of the homogeneity of score distributions and a test of the equality of correlations between multiple-choice and essay scores).

Both Swartz and Patience noted that one of the most important conclusions gained from their experience is that constant vigilance is required to maximize essay score scale stability and reading reliability. They concluded their session by stating that the development of the GED national essay testing program posed unique problems in large-scale writing assessment, and some of the solutions may apply throughout the field of writing assessment.