
FROM HOLISTIC TO PROFILED SCORING OF SPECIFIC ACADEMIC WRITING

Speaker: Liz Hamp-Lyons, University of Michigan

Introducer/Recorder: John E. Nace, Community
College of Philadelphia

Liz Hamp-Lyons presented conclusions from her research on the scoring procedures for the British Council Proficiency Test of the English Language Testing Service. She explained that the ELTS is charged with responsibility for testing overseas, non-native English speaking students who wish to enter Britain for post-graduate study. As such, its tests are roughly parallel to TOEFL instruments in the United States. The British Council Proficiency Test is composed of five components: reading, listening, study skills, writing, and oral interview (one-to-one). For the last three components, students choose to be tested for skills within one of the several disciplines, including life sciences, medicine, social studies, technology, physical science, and general academic knowledge. The test is administered on demand to about 12,000 candidates per year in about ninety countries around the world. One person in each country is designated ELQ (English Language Qualified) and is responsible for scoring all tests administered in that country. In many countries, as few as five students per month take the test.

For the writing component in every module, there are two readings to which candidates compose written responses. The first reading requires a "divergent" approach, and candidates must apply outside experience, opinion, judgment, or analysis to their reading when they write in response to this text. The second reading requires a text-based approach; candidates must restrict their written analysis to features in the reading itself. Scoring of the writing test component is criterion-referenced and scaled for nine levels (bands) of performance (from "non-writer" to "expert writer").

The test, originally administered in 1980, posed many

problems at the outset. First, it was designed by English language professionals who lacked specific qualifications in test development and measurement. Secondly, persons designated ELQ in countries which administered the test would sometimes see only one candidate per month, making it impossible to collect a sufficient number of test performances for effective rating comparison. Finally, only one set of "benchmark papers" (a sample representing bands 8-3) was available for raters' reference. In addition, at this early stage of administration (1980-83) there was no test for reliability. It was not until 1984 that Hamp-Lyons was charged specifically with tightening up reliability and improving rating procedures.

In order to accomplish this, she started by looking at the descriptors across which tests were rated in the several bands; her purpose was to tease out the criteria that were implicit in the original descriptors. These descriptors were as follows: communicative quality (essentially holistic), organization (showing the extent to which logical structure carries the message), argumentation (effectiveness of argument), and linguistic accuracy and appropriateness (mechanics, syntax, etc.). She also discovered that some of the greatest problems candidates had in writing for the test were with length of response and plagiarism. Accompanying problems in these areas was the disturbing tendency of raters to assign a range of scores (for example, 4/5/6) on the writing test instead of confidently fixing a single score.

In answer to these immediate difficulties with the test, Hamp-Lyons wrote a training manual to accompany the test stating what the criteria meant, what the problems were, and how to deal with them. In order to effect these changes, she had to persuade the British Council to publish new short paragraph band descriptors for each of the nine levels. At the bottom of the revised, "profiled" scoring sheet is a box labeled "Final Band," for the average score on the writing. Admissions officers at British universities are interested in this score while the British Council looks at subscores to inform their decision whether to accept the foreign graduate student. The English Language Center at the student's university receives a copy of this sheet so that the university knows what the student's abilities are.

Based on the student's performance across the criteria, the British Council can make a tentative decision that, for example, the student is a competent writer but probably a slow reader. In this case, the Council will fund the student to take an English "topper" course before starting at the university in October. At the University of Edinburgh, students were placed into sections according to the band at which they scored. Some students were at various bands upon admission to the university of their choice. Hamp-Lyons concluded by saying that she is attempting to introduce this "profiled" scoring procedure in placement testing at the University of Michigan, where she is currently Testing Coordinator.◊