THE DEVELOPMENT OF WRITING PROMPTS

Speaker: Gordon Brossel, The Florida State University

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

Gordon Brossell’s session addressed issues of importance in developing prompts for large-scale writing assessments and for classroom assignments. Brossell noted that most essay exams offer only a truncated opportunity to engage in the processes of writing that our profession has generally come to see as essential for teaching writing. In addition, most essay tests provide only a single instant of writing performance—a “snapshot”—which may not be representative of the writer’s ability. Brossell warned that one of the questions teachers must consider is whether they want to teach students how to write or how to pass essay examinations. The two activities are not the same, though there is the potential for bringing them closer together.

Brossell stated that all writing tests are, to a certain extent, also tests of reading and interpretation. Writers read and interpret topics or prompts differently, and this affects their writing performance. Understanding a topic in one way causes a writer to approach the composing of a written response in a certain way. How much information writers have at their disposal is tied to how they interpret the prompt. Moreover, Brossell noted that different prompts ask for different cognitive and rhetorical activities. The demands on writers are different if the task is to describe an event or tell a story or to present an argument. Students who understand what a prompt is asking them to do have a better chance of writing well than those who do not.

Next, Brossell reviewed recent research on factors affecting prompt variability. First, he said that changes in the wording of prompts tend to produce no discernible differences in essays written in response to them. Second, he examined the effects of the subject matter of different prompts and explained the “content-fair” prompts that he and James Hockett developed for Florida’s statewide College Level Academic Skills Test in Writing. These prompts enable writers to select their own subject matter (for example, they chose the practice in the prompt, “Describe a common practice in American schools that should be changed”). This solves the problem of unequal familiarity with the content of most prompts in large-scale writing assessment. Then, Brossell discussed the research on the rhetorical specification of prompts. Several of the studies that he reviewed showed that providing explicit information about the purpose and audience for writing does not necessarily elicit the best writing performance from the students, a finding that calls into question conventional wisdom about rhetorical specification.

Next, Brossell pointed out some implications for the teaching of writing: (1) a good prompt is not writer-proof (so there is no substitute for extensive prewriting activities and teacher feedback), (2) teaching to a test prompt—with many in-class practice essays—may help students pass but it dramatically undercuts the learning process and students’ writing processes, (3) classroom writing assessment cannot—and should not—mimic large-scale assessment. He concluded by stating that large-scale assessment ought to be modeled after the ways in which writing is taught and learned. This means that the composing process has to be incorporated into the testing situation.

SELECTING THE BEST ESSAY TOPIC

Speakers: Gertrude Conlan, Educational Testing Service Pamela Chambers, Educational Testing Service

Introducer/Recorder: Albert Porter, Mercer County Community College, N.J.

Gertrude Conlan and Pamela Chambers described the procedures used by Educational Testing Service (ETS) to select and field test essay topics for two standardized testing programs: The New Jersey College Basic Skills Placement Test (NJCBSPPT) and the College Board English Composition Test (ECT). The NJCBSPPT is administered to 50,000 New Jersey College freshmen each year to help identify those needing remediation. In December of each year, the ECT is completed by 90,000 college-bound high school students. Students’ essays in both programs are scored holistically at ETS by trained faculty readers under controlled conditions.

Although the audiences for the two programs are quite different, the procedures used to select topics and the characteristics of the most successful questions are remarkably similar. Committees of English faculty generate an initial pool of 50-100 topics from which ten or more are selected for field testing. Using guidelines from ETS, each topic is pretested in the classroom to obtain a sufficient number of writing samples from students representative of the target population. Readers consider the success of each topic in prompting students to develop their ideas, form arguments, and defend their opinions. The best topics are accessible to students, allowing them to show how well they can write.

Typically, only a few topics survive the pretest reading. If a topic produces inflammatory responses or fails to differentiate between weak and strong writers, it is discarded. Some topics are judged inadequate because they generate interesting writing or reveal unintended interpretations of the question. Even the best topics from the pretest stage may fail to work after adoption. Unexpected events immediately preceding the test may disproportionately influence students responses to a question. For example, a major world event or a widely viewed television show can homogenize students’ viewpoints and arguments on a topic. Obviously, the field testing effort cannot control for such occurrences.

Using illustrations from actual topics and student responses, Conlan and Chambers demonstrated the features of topics that have succeeded and failed. Some topics judged marginal during pretesting succeeded after reasserting to clarify their meaning or amplify their intent. For example, the replacement of vocabulary found to be obscure to certain cultures (or reflecting regional differences) has made some
marginal topics more accessible to a broader range of students. In their closing remarks, the presenters emphasized the most important considerations in topic selection. To the extent possible, topics should be:

- clearly stated, using language familiar to the audience
- appropriate to the population being tested
- reasonable for the allotted writing time
- suited to the purpose of the test
- accessible to the broadest range of students
- interesting to write about
- capable of producing writing across the range of student abilities

Some participants were concerned about the limitation of twenty minutes for writing the essay, particularly in light of instructional emphasis on the importance of the prewriting, drafting, composing, and revising process. Although the presenters acknowledged the potential difficulty of the time constraint, they noted that it would be impractical in a testing situation to provide students with the necessary time and distance to simulate actual writing exercises. In a separate project now nearing completion, ETS is analyzing the influence of additional time on the quality of writing produced by examinees taking the NJCBSPT essay. Under controlled conditions, students in that study were given twenty minutes, thirty minutes, and twenty minutes with ten minutes of prewriting; results of the investigation will be available in late spring.

A separate question involved the comparability of scores obtained using different topics. The presenters warned against such direct comparisons, even when scores are derived using trained readers and carefully selected rangefinders. However, ETS is studying both the inter-reader reliability across different scoring sessions involving the same topic and the comparability of scores when the same topic is repeated in separate test administrations. These studies primarily focus on the effectiveness of reader training.

**TRAINING SESSION IN THE HOLISTIC SCORING METHOD**

**Speaker:** Rose Ann Morgan, Middlesex County College, N.J.

**Introducer/Recorder:** Dennis Donahue, New Jersey Institute of Technology

Rose Ann Morgan gave conferences attending this session a hands-on introduction to holistic scoring. Participants had an opportunity to read, score, and discuss examples of essays written by students as part of the New Jersey College Basic Skills Placement Test (NJCBSPT). The session was a scaled-down simulation of the training given to the readers, usually high school and college teachers, who score these essays.

Morgan explained that the essence of holistic reading and scoring is the non-subjective appraisal of the work in question. Readers are not asked to do anything but to assign numbers to discourse. The numbers (1 to 6 on the NJCBSPT) represent different levels of achievement on the test, but the significance of the numbers, the judgment regarding which number represents passing or failing or placement in one course or another, is to be made by others. Such a judgment might be made, for example, by the person at a college responsible for placing students in a course at an appropriate level. Such people see the whole range and distribution of actual scores. The function of a holistic scoring training session is to make sure that all the readers know and are able to employ the full scale and range of scores for evaluating essays.

In an actual holistic scoring session, given just before a reading, readers sit in groups of six or seven under the guidance of a table reader, an experienced reader who has participated in a preliminary holistic scoring session a day or so prior to this general reading. The readers are given copies of prescored sample essays, called rangefinders, and are asked to learn—or re-learn if they are experienced readers—the scale to be used throughout the reading. They are asked to rank order these essays and then to assign each a score ranging from 1 (lowest) to 6 (highest). There is at least one essay for each of the six scores.

When all have finished scoring the rangefinders, the readers are asked to announce their scores by a show of hands. Readers can see instantly whether they are scoring essays higher or lower than the table leaders and other readers in the room. They then make any mental adjustments necessary to bring their scoring into line with the scale being used by the group. Additional rangefinders are scored and discussed and then, when the chief reader is satisfied that readers are using the agreed upon scale, "live" essays are brought out and the actual scoring begins. At irregular intervals during the session, additional rangefinders are scored by the group to keep their idea of the scale sharp.

At the abbreviated training session given at the conference, participants were given the same information readers at an actual session get. They were told that the writers had been given just twenty minutes to write an expository response to a question relating to a task they had not previously been able to perform but now could. Participants were reminded that the writers' task was complex, that a full answer might deal with the writer's past and present and might give an explanation as to why the writer now could perform that task. But Morgan asked readers to read the essays supportivey, and not to expect all writers to respond to all three areas. Rather, the readers were to score the essay on the basis of what had actually been written. When Morgan was asked whether this was unfair, she responded that it wasn't, that a given essay might cover all three areas, not that it should. She added that a writer's covering all three areas was no guarantee that the writing was good, whereas an essay that had dealt with only one or two areas before time elapsed might in fact be well written, it incomplete.

A special method of training readers was then employed. Participants were given just three rangefinder essays to begin and were asked to rank them high, middle, and low. Then they were given two additional papers and asked to interleave them among the first three. Finally, they were given the