

## THE CBEST ESSAY EXAM: PERILS, PROBLEMS, AND PROMISES

**Speakers:** Paul Ramsey, Educational Testing Service, New Jersey

Claire Pelton, Los Altos High School, California

**Introducer/Recorder:** William Hughes, Cuyahoga Community College, Ohio

Claire Pelton began by describing the events which prompted the creation of the California Basic Educational Skills Test (CBEST) writing exam. The proficiency exam was suggested by California's Gary Hart, now a state senator, in response to the charge that some teachers were illiterate. As evidence, he circulated a teacher's letter which contained numerous spelling errors and sentence fragments. The bill resulting from his efforts states that teachers and teaching candidates must demonstrate proficiency in mathematics, reading, and writing. An advisory committee of thirty-five members, selected from the disciplines of reading, writing, and mathematics, some of whom were political appointments, wrote the initial test specifications. Along with objective sections on mathematics and reading, the proposed test included a one-hour essay to be scored holistically. California's Superintendent of Education recommended that there also be an objective part to the writing section.

Paul Ramsey said that he shared the Superintendent's view, expressing his concern about a single writing test determining people's futures in their chosen careers. Although a multiple choice test on writing defies the intuition of most writing teachers, there is a high correlation between scores on certain multiple choice tests and writing skills. New types of objective tests can, for example, measure skills in revision and in selecting stylistic options, which are indicators of the writer's linguistic ability. Therefore, Ramsey advocated a larger test with at least 47 objective questions in addition to the essay. This would enhance the reliability of the essay and make it more valid for making a judgment about prospective teachers' competencies.

The advisory committee, of which Pelton was a member, rejected Ramsey's proposal and recommended a writing test composed of two essays in two different modes, each to be evaluated by two readers. One of the essay topics would have an expressive aim and pertain to a remembered experience (such as an event in college which made the writer a different person). The other would have a referential aim and require analytical thinking. Pre-testing established that neither essay topic should designate an audience, since writers became angry with audience specification, knowing full well that their audience consisted of the scorers. There was much debate on this issue, but Pelton maintained that the audience specification, used effectively in classroom assignments, is not necessarily good for mass testing. Au-

dience specification becomes a problem for holistic scoring when well-written papers fail to consider the audience. Even letters to the editor, furnishing false audiences, can create problems, with different editors prompting different registers. A test task without an audience designation, according to Ramsey, is closer to a real writing situation.

For pre-tests, candidates were encouraged to bring a watch, but there were no time announcements during the hour. There was some concern that they would spend too much time on the first topic. The first time the test was administered, the expressive topic preceded the referential, and candidates slighted the latter in favor of the former. In subsequent administrations, however, the topic order was reversed, and the problem disappeared. Also on the subject of pre-testing, Pelton said that the state planned to offer the test free to 2,000 upper-level college students who were candidates for teaching, and to 275 teachers taking administrative positions. She stressed the importance of testing borderline candidates, students with SAT scores of 360 verbal and 380 math, and B's and C's in academic work. The advisory committee anticipated that these students would be its main challenge. It also searched for a significant number of minority students for pre-testing, since it wanted to pre-test a sample similar to the final testing population, with at least 150 responses for every topic.

In preparing to score the test, much time is spent on training readers. The day before the reading, table leaders choose single essays which reflect the scoring guide and exemplify the four possible scores: Pass, Marginal Pass, Marginal Fail, and Fail. Passing essays demonstrate the proficiency of upper level college students. Pelton said that they have little difficulty finding readers with previous experience in holistic scoring, but there are always new readers for CBEST, some of whom have never scored so many essays at one time (possibly 10-12,000 in one weekend). Even readers with some experience feel somewhat insecure on the first day, which makes it more important that table and room leaders come to a consensus on scores for the sample essays. Leaders spend much of the first day discussing with the readers how the samples reflect their scores. Ramsey added that reading teams on topics one and two also exchange their samples and, if they disagree on scoring, decide whether or not to change samples.

Ramsey concluded the presentation by asking the conferees about the issue of comparability, emphasizing that one of the biggest challenges in working on CBEST is to be certain that tests given at different times are comparable. Standards must remain constant across administrations and topics so that the ability of candidates to pursue their careers is not determined by the luck of taking the test at one time rather than another. At the same time he wants to improve the test without invalidating past results. This can be accomplished, Ramsey suggests, by comparing scoring samples to those of previous years and by retaining a small cadre of chief readers. ■