

VALIDITY ISSUES IN DIRECT WRITING ASSESSMENT

- Speakers:** *Karen Greenberg*, NTNW and Hunter College, CUNY
Stephen Witte, Stanford University
- Introducer/**
- Recorder:** *Joanne Van Oorsouw*, College of St. Catherine, Minnesota

Karen Greenberg began with what she deemed a radical statement: "I have examined more than 600 writing tests and have yet to see one that I would consider to be a valid one." She went on to state that it seems impossible for writing tests, with their narrow subjects, implausible audiences and severely restricted time frames, to reflect the natural processes of writing in either academic or personal contexts.

Greenberg explained her position by pointing out that writing consists of the ability to discover what one wishes to say and to convey one's message through language, content, syntax and usage that are appropriate for one's audience and purpose. In light of this, she said, it is particularly distressing to note that teachers at many institutions find themselves administering tests that bear little resemblance to this definition or to their curricula and pedagogy. For example, many schools still use multiple-choice tests of writing even though this type of testing does not elicit the cognitive and linguistic skills involved in writing.

She stated that writing sample tests, on the other hand, can assess writing capacities that cannot be measured by existing multiple-choice tests. They, however, also, have flaws, and many problems result from our reliance on single-sample writing tests for placement and proficiency decisions. She warned that a single writing sample can never reflect a student's ability to write on another occasion or in a different mode. Yet, according to surveys conducted by NTNW and CCCC, thousands of schools across the country continue to assume that "writing ability" is stable across different writing tasks and contexts and continue to use a single piece of writing as their sole assessment instrument.

Greenberg then went on to suggest what those involved in large-scale direct assessment of writing should do about validity. The first step in establishing a test's validity is to determine its purpose: what

information is needed by which people and for what purposes? The next step is developing a clear definition of the writing competence that is being assessed, one that will vary according to the purpose and context of the assessment. Developing this definition is a critical step in creating a valid assessment, but it is easier said than done for there is as yet no adequate model of the various factors that contribute to effective writing in different contexts. Finally, after coming to agreement on their definition of writing competence, faculty need to establish consensus about the writing tasks that are significant in particular functional contexts.

Greenberg noted that she deliberately chose to talk about faculty rather than test developers, for she believes that the people who teach writing should be the ones who develop the assessment instruments. Faculty need to work together to develop tests, to shape an exam they believe in so that they can be sure its principles infuse curriculum and classroom practice. Even when faculty work together, however, Greenberg said that definitions of competent writing may vary dramatically. Locally-developed essay tests show incredible variability in the skills measured, due to difference in the range of skills assessed and the criteria used to judge those skills. For example, faculty often differ about the range of discourse structures that they should teach and that a test should assess. One way to sample students' ability to write different types of discourse is to use the portfolio method, in which writers select three or four different types of drafts and revisions for evaluation. This kind of assessment reflects a pedagogy that emphasizes process over short, unrevised products. Thus, this kind of test stimulates writing teachers and programs to pay more attention to the craft of composing.

Greenberg's final point was phrased as a question: What is the relationship between what we teach and what we test? We cannot, and should not, separate testing from teaching, and we as a profession must be more concerned with the validity of both of these efforts.

Steve Witte summarized a study begun in 1982 which sought to answer two research questions: (1) Do writing prompts that elicit different types of writing and that elicit written texts of the same quality cause writers to orchestrate composing in different ways? and (2) Do comparable prompts that elicit the same type of writing and elicit written texts of the same quality cause writers to orchestrate composing in different ways? Witte stated that although this study did not investigate naturally occurring discourse, this type of experimental study can inform the kinds of conceptualizations we can make beyond the experimental study.

The first step in conducting this research was to create two comparable writing tasks of two types:

expository and persuasive. Prompts were created after consultation with students, writing teachers, high school teachers, and pre-service high school teachers. Those prompts found to be comparable by these groups were then pretested and were found to elicit comparable ranges of writing quality. The subjects were 40 volunteer college freshmen at the University of Texas who were randomly assigned to one of the four tasks. Think-aloud protocols and rough drafts were collected and analyzed according to a coding scheme developed by the experimenters. The results of a multivariate ANOVA showed that 16 variables distinguished between the persuasive and expository tasks; these variables included generating ideas, setting content goals, reviewing text. Writers tended to set more content goals and generate more ideas for the expository tasks and set more rhetorical goals for the persuasive tasks. A discriminant analysis was done to determine which variables distinguished among all four tasks. Eleven variables were found to do this.

Witte stated that findings indicate that writers engage in different kinds of processes for different kinds of tasks. In terms of writing assessment, each prompt we use to assess ability will be measuring different dimensions of that ability. The obvious conclusion, then, is that there is no way to assess writing ability with only one task or prompt. We do not yet know how many prompts or tasks might be needed. Witte also noted that this study should make us question models of the writing process that are based on protocols from just one task. More research of the type presented here--studies that examine the effects of context on process--are needed. In Witte's study, context was limited to the writing prompt, a part of the context important to writing assessment. He said that we need more research that will help us identify how writing processes are circumscribed by other aspects of context.