

## K-12 WRITING ASSESSMENT: ISSUES AND MODELS

Speakers: Charles R. Chew, *Bureau of English and Reading Education, New York State Department of Education*  
Beth Breneman, *California Assessment Program, California Department of Education*

Introducer/Recorder: Richard D. McCallum, *University of California, Berkeley*

The presentation reported here outlines the writing assessment programs of two of the nation's largest states: California and New York. Charles Chew began the session with a philosophical and historical overview of New York's statewide writing assessment program. Testing for credentials and other certifications was mandated by New York's Board of Regents in the 1860's. Currently students' writing is tested in the fifth, eighth, and eleventh grades in New York.

Familiarity with testing, both its strengths and weaknesses, has led to a "pragmatic" approach to assessment in New York. Chew stressed that "all assessment situations are artificial" and that all educators need to recognize this limitation. To account for this, at the fifth grade level two writing samples are taken. The writing of these papers is prefaced by a pre-writing session in which students are lead through the task. At the eleventh grade level three samples are taken: a business letter of complaint, a report (data provided), and a persuasive essay. To account for the artificiality of the situation the tasks are placed in a rhetorical context, with purpose and audience provided.

In regard to the strengths of testing, Chew stated that the concern for writing assessment has had "more impact on instruction at the elementary level than any occurrence in twenty-five years." Several aspects of the assessment process have contributed to this change. To aid in assessment, 10,000 teachers have been trained in holistic scoring. This has led to an increase in awareness of the nature of the process of writing. In addition, test results are immediately reported to classroom teachers. Increased awareness coupled with valuable diagnostic information has given teachers an understanding of student needs and an idea of what remains to be done. Since the inception of the assessment program, Chew has noted a change in how teachers view students' final products: there has been a move from a focus on mechanics to other issues, as well as an awareness that drill and practice is not the way to teach writing.

In New York, all writing samples are scored holistically. Papers are first scored by the district. Papers that are scored 60% or above are then re-scored by the state (approximately 75,000 per year). If the score is reversed, the school must abide by the state decision, but there is an appeal process. For those students at the high school level who fail, provisions are made for re-testing. At all levels the state mandates that remediation be provided for students who fall below the cut-off score. Teachers must document these remedial plans, and parents must be notified.

The last part of Chew's presentation focused upon other practical considerations related to developing assessment programs: overall cost and test development. The cost of the existing program has been significant. Raters are hired per diem and must be

(Continued on page 21)

trained and tested. The cost of test development is hard to judge given the elaborate process of piloting that must precede implementation. For example, the current exam consists of six prompts, which were derived from an initial pool of 130. In regard to test development, samples of student writing throughout the school year are used to obtain an idea of student abilities. These insights are coupled with research on the development of composition ability over time. Even when this task is completed and a pool of prompts is available, questions such as "Do the prompts work?" and "Will testing samples be interpreted differently in different situations?" must be addressed.

In closing, Chew noted that more can be done to utilize the results of assessment. The goal is to develop writing strategies within a "whole language approach." To achieve such a goal, one of the biggest challenges remaining is the continued in-service training of teachers in the relationship between writing assessment and instruction.

Beth Breneman presented an assessment program quite different from the New York model. Breneman discussed two assessment models: one which is in use in California, and one on which the state is working. According to Breneman, both of these models are geared toward instructional applications.

Currently, assessment in reading, math, and written expression is conducted through the California Assessment Program (CAP). Testing is conducted at the third, fifth, eighth, and twelfth grades. All CAP exams employ matrix sampling. The exam that individual students take is developed from a large pool of comparable assessment questions developed by the state. Each child takes only a "sample" of all the available questions. At grade eight, for example, there are thirty-eight comparable exams. Each child takes only 1/38th of the "whole" exam. No individual scores are provided from the results. Information is provided at the school and district levels. A major strength of this system is the testing time: usually only 30-40 minutes.

The tests developed for the CAP program are designed to provide diagnostic information as to the strengths and weaknesses in skills areas at the school and district levels. At the high school level, essay topics cover various content areas as well as literature. Students are asked to answer questions concerning the language choices and mechanics employed in their essays. Students are also required to edit and manipulate the essay. These activities provide information about pre-writing/organization, critical judgment, overall organization, knowledge of paragraphs, sentence manipulation, and sentence recognition. This information can be used to plan instruction at the school and district levels.

Such a model, according to Breneman, has both advantages and disadvantages. On the positive side, it is currently impossible to teach to the test without having students write. Further, the decisions students make on the exam require critical judgment and capture some aspects of the process of pre-writing and revision. On the negative side, the current system does not directly assess writing ability. The data generated does cover skills, but not facility with the types and uses of writing (for example, informational vs. persuasive). Because of this limitation, California is attempting to move away from multiple-choice testing. This has led to an effort to develop an alternative model, Model #2, which is slated to be implemented statewide at the eighth grade level in 1987.

The goal of Model #2 is to couple the valuable aspects of Model #1 (information at the school and



district levels) with individual assessment information. The guidelines for Model #2 are currently being developed based on the results of a pilot study conducted with eighth grade students. The purpose of the pilot was to begin to explore methodologies that could be employed in assessment and, more generally, to find out how well eighth graders can write. The pilot exam consisted of National Writing Assessment prompts, six core prompts, and experimental prompts. Results show that the following general areas should be included in an exam:

- 1) A variety of types of writing (narratives, expositions, discursive writing, personal reflection, and dialogues).
- 2) Matrix sampling at the school level.
- 3) Primary trait scoring guidelines and criteria.
- 4) Secondary trait scoring.
- 5) Provisions for state and national comparisons.

The pilot also provided an insight into current instructional needs in secondary schools: more writing assignments and more types of writing, a broader understanding of the types of writing and their demands, and a need to reflect on writing standards, with a concern for literature.

Breneman reiterated the tentative nature of the pilot results, and in response to a question from the audience, she noted that no final decisions have been made as to the type of scoring or the nature of the exam. Although the results of the pilot look promising, Breneman pointed out that questions still abound:

- 1) Should the multiple-choice format be abandoned?
- 2) Can prompt security be assured?
- 3) What are the reading demands of the prompts and passages and how can they be adequately controlled?
- 4) How can we come to understand the nature of the processes involved in writing at all grade levels?

Answering these questions, Breneman and Chew agreed, requires a concerted group effort.