

ISSUES IN ASSESSING WRITING IN A STATE-WIDE PROGRAM

The National Institute of Education contracted with Louisiana Tech University in 1980 to determine the "state of the art" of writing assessment in the nation. The research project consisted of two phases. The first phase involved sending a questionnaire to all fifty state departments of education and to selected large-scale school systems around the nation.

A major purpose of the questionnaire was to determine the extent to which large-scale writing assessment is being implemented in the country. Of the forty-two states that responded to the questionnaire, twenty-four reported that state-wide writing assessment programs were being administered. Of these twenty-four states, twenty-two reported assessing writing using a writing sample. Most of the states that used a writing sample indicated that they used holistic scoring. Only two states indicated the use of an objective measure alone, and eleven states combined an objective test with a writing sample.

In order to validate and amplify responses to the questionnaire and to extend and enrich research into the nature of large-scale assessment, the second phase of the project culminated in a national conference in the summer of 1981 to which individuals involved in large-scale writing assessment programs were invited.

Responses to the questionnaire and discussion at the conference enabled us to formulate the following observations and generalizations about the administration of large-scale writing assessment programs across the country.

USE OF A WRITING SAMPLE

All participants agreed that a direct assessment of writing carries a special message to teachers that students must write in order to learn to write. Those participants representing institutions or agencies that did not include a writing sample in their assessment programs indicated that it was the cost of administering,

scoring, and analyzing a writing sample that prohibited its inclusion in their respective programs. In short, participants agreed that there are qualities in writing that cannot be measured by objective tests. This viewpoint seems to be supported by leaders in government and industry. On the issue of legality of assessment measures, participants pointed out that all assessment results, whether from direct or indirect measures, must have established indices of reliability and validity to be used as legal evidence.

DEVELOPING WRITING TASKS

Since the purpose of a writing sample is to determine how well students can write and not how much knowledge they have about a particular subject, most participants agreed that a major consideration in developing topics was to try to eliminate pressure on writers to provide information. Further, participants stressed the importance of field testing topics with a representative sample under carefully controlled conditions.

The guiding criterion for the types of writing tasks assigned appeared to be the philosophy of the agency or school with reference to using holistic or analytic scoring procedures. Where testers leaned toward holistic scoring, the tasks were usually open-ended, and designed as a twenty-minute task; in such instances, only one essay was obtained from each student. Where analytic methods, such as primary trait scoring, were preferred, tasks were narrowly defined and designed to elicit specific responses. As the student was expected to write only a short passage, two or three different tasks were assigned, often including at least two modes of writing. Virtually all participants indicated that their students were assigned the same topics at a given grade level.

(Continued on page 22)

SURVEY OF TESTING WRITING PROFICIENCY: A PROGRESS REPORT

In 1979, the CCCC Task Force on Testing conducted a survey of ten percent of the active membership of CCCC, randomly selected, in order to gather information on the state of the art of testing proficiency in college writing; a follow-up study was done in 1981. A report of the study by the Committee's Chief Investigator, Rosentene Purnell, appears in the December 1982 issue of CCC. The following is a brief summary of its findings:

- *Testing of writing for placement, diagnosis, and certification is widely accepted and on the rise among colleges nation-wide.*
- *There is strong resistance against the use of a test as a single criterion for advancement or graduation.*
- *The positive side effects of growth of the testing movement include: an impetus for the establishment of writing centers; proliferation of developmental and remedial courses; an increased awareness of and commitment to a shared responsibility for improved writing proficiency; a positive impact on writing curricula.*
- *The majority of respondents reported the use of direct measurement of writing skills (the essay) as all or part of tests used for diagnosis, placement, and certification.*
- *Holistic scoring of essays is the preferred assessment procedure as reported by a majority of respondents; one-third reported the use of both an essay and an objective test.*
- *Most respondents reported that writing tests are broadly based*

assessments of writing competency rather than tests of skills practised in any single course.

- *The length of time students are given to write varies greatly from one to two and a half hours.*
- *Remuneration of readers is not a common practise; in those few institutions which reported its use, students are assessed a fee to cover the costs.*
- *The end of the freshman year is most commonly used as the cut-off point for certifying minimum competency.*
- *Internal motivation was reported by most institutions as the source of the upsurge in testing, although pressure by state and employment agencies was also cited.*

The findings suggest the need for further research on the uses and abuses of both the products and processes of testing writing proficiency including the possible misuse of testing to abrogate equal access to educational opportunities and the potential for misrepresentation by writing tests of the abilities of some students—in short, to what extent testing of writing may adversely affect the equity and integrity of the educational process.

Rosentene B. Purnell is a Professor of English at California State University, Northridge; she is also 1983 Program Chair of the CCCC.

McCREADY, MELTON (continued)

SCORING THE ESSAYS

For the most part, participants are adapting standard procedures for scoring essays to local situations. For instance, combining holistic scoring with clearly defined evaluation criteria is a common method.

Much discussion focused on the problems posed by attempts to score mechanics in large-scale assessments. Where holistic scoring is used, a general impression of a student's ability to use mechanics correctly is considered when assigning a score. Where analytic scoring is used, a separate score for mechanics is assigned. The National Assessment of Educational Progress scores mechanics by determining a percentage of errors. Scoring mechanics presents problems because a writing task does not necessarily stimulate all students to attempt the same mechanical conventions. In fact, research indicates that better writers will attempt more complex constructions and therefore might make more errors than less competent writers. Therefore, applying a uniform system of scoring mechanics may elicit misleading results.

The logistics involved in scoring large numbers of essays was also discussed. One state agency brought in teachers from all over the state to evaluate essays while in others outside contractors were responsible for scoring. Participants seemed to agree, however, that directly involving the classroom teachers in the scoring process had a favorable impact on instruction.

Time and cost of scoring varied to such a degree that no generalizations can be drawn. Criteria seemed to be the number of students being tested and the method of scoring employed.

SOME GENERAL CONCLUSIONS

A number of questions, both technical and meta-physical, remain unanswered. The nature of writing tests which are designed to improve writing, presents a major area of needed research. Very little is known about how carefully norms are determined, how stable the results are, how the results might advantage one group over another, or how accurate their predictive value is. Because competency testing programs are being used to make important decisions about people's lives, it is imperative that administrators of testing programs, as well as the users of the information yielded, scrutinize their programs with care.

The NIE Assessment project with Louisiana Tech was one effort in this direction. A great deal of data about specific programs was gathered in the course of the two-year project which, in the interest of conserving space for this publication, we have omitted here. Readers are welcome to write to us at Louisiana Tech for more information.

Michael A. McCreedy is Professor of Education at Louisiana Tech University in Ruston; he has been involved in large-scale writing assessment at both the state and national levels and has recently completed research at the national level for NIE.

Virginia S. Melton is Associate Professor of Education at Louisiana Tech where she teaches graduate and undergraduate courses in elementary language arts; she has served as a consultant, trainer of readers, and analyzer of data for the Louisiana Writing Assessment since its inception.