

RESEARCH DESIGNS FOR STUDYING WRITING ASSESSMENT

Speakers: Ron Basich, *Office of the Chancellor, California State University*
Linda Polin, *Pepperdine University, Los Angeles*

Introducer/Recorder: Judy I. Jones, *University of California, Berkeley*

Ron Basich opened the session by reviewing various research designs for the study of writing assessment. He drew upon his professional experience as research analyst in the California Chancellor's Office where he has been instrumental in developing a multi-test data base for analyzing the entry level placement and admissions tests for the California State University System.

He delineated three major factors that aid research design: 1) the identification, understanding, and control of variables; 2) the use of accurate measuring devices and the understanding of variance within these instruments; and 3) project management. In order to identify and account for variance factors in research design, one must identify data sources and the impact of those sources on the design. Questions to ask while setting up a design are: 1) Will the data and its analysis prove generalizable? and 2) is consideration of interference effects built into the design? With the identification and development of appropriate measuring devices, it is essential to know the subject area being assessed. This in turn aids construct validity or the extent to which the measuring device captures the heart of the questions being studied. Further, context validity is strengthened because the many facets of the subject matter studied are identified and their effects accounted for.

While redundancy in test items should be avoided, reliance on one item to measure a particular content question may be insufficient. Related items, therefore, should elicit information that serves to elaborate content area from many different angles. In this way, unambiguous responses are facilitated while duplication across items is avoided. Test instruments should be piloted early on in a project in order to ensure adequate time to facilitate improvement of responses to test items. Early checks on items help lessen the probability of external and extraneous variance in responses. The means for testing should be set up and clearly established ahead of time. Some functional issues to consider are the compactness of the test, time allowances scheduled for test administration, means of distributing tests to subjects, and proper coding for essay assignment during analysis.

The key to effective project management is to encourage cooperation among project members. This becomes essential in large-scale research designs in which efficient data collection becomes contingent upon the cooperation of each project member. A further concern of the project manager is to generate the comprehensive cooperation of project participants. Allowing flexibility to enter the design can result in useful feedback from project members and participants. Paying strict attention to the realities of funding and the distribution of funds over time and coordinating funds between grant cycles and federal fiscal years' end all contribute to managing a project smoothly.

Basich outlined several steps for developing an effective writing assessment project: 1) develop a plan, 2) develop a schedule, 3) supervise personnel, 4) write checks, 5) facilitate cooperation, 6) allow for flexibility in the design, 7) provide room for feedback, and 8) cover details.

Linda Polin specifically addressed issues of validity and reliability in research design. She is the co-director of the NIE funded Research in the Effective Teaching of Writing Project and her fields of expertise include research in writing, computers in education, and research methodology. In considering the validity of assessing high school and post-secondary writing, she indicated the importance of recognizing the process orientation of writing.

A researcher should be concerned with item reliability or the extent to which an item always elicits the same performance from subjects. Could there be cultural bias difficulties contained in an item? Might there be a problem with the topic selected? Common checkpoints should be set for inter-rater reliability. Scales should be used to retrain raters throughout the holistic scoring procedure to avoid rater drift.

The process orientation of writing should also be taken into account. For example, students may know a process of writing before they are able to perform it. How then can a measure capture this? One suggestion is to collect worksheets used during the administration of the measure. Another suggestion is to measure compositions over time. This can be as simple as measuring different compositions over one through four consecutive days or may be as complex as a longitudinal study. Informal pilot testing, talking to teachers and students concerning their thoughts on the writing process, and collecting writing samples other than text samples all enrich data sources and help provide insight into processes.

Both speakers elaborated upon the Research in the Effective Teaching of Writing project. The project's major premise is to explore the attitudes and practices of college composition programs at California universities. Two questions asked are: what are the properties of a college composition program, and what is the impact of these properties on college compositions? A two-level approach was taken in constructing the research design: a taxonomy of composition skills was developed, and the measuring instruments chosen were strictly descriptive. A campus fact sheet was developed that took into account enrollment of students in college courses, and the faculty involved in these programs. A seven-page faculty survey was administered aimed at eliciting attitudes and practices of college writing faculty. Sampling bias was accounted for by addressing the entire population of faculty involved in college courses for the participating campuses. Final descriptive instruments included the interview protocols of campus and composition administrators. Thirteen factors yielding information regarding instructional styles were analyzed from a factor analysis of the questionnaire.

Analysis of the performance of students on campuses proved complex. Factors taken into account were student population variables, variation in composition instruction across campuses, and differentiation of students into remedial and non-remedial categories. Content valid test items were administered

(Continued on following page)

RESEARCH DESIGNS (continued)

and 1500 essays were collected for analysis. Each essay was holistically scored and inter-rater reliability checks were conducted throughout the scoring procedure. Two discourse features were analyzed. One category, development and focus, consisted of the ability to state generalities and support those generalities. Another, high level mechanics, included evidence of proper sentence embedding and varieties of sentence constructions. This study is still in progress, so no final report was made. However, it proved extremely useful as a model of how to set up an effective research design and helped illustrate issues to consider in managing a complex research project.