R E P O R T R E S U M E S

ED 014 466 SP 001 364

CRITERION PROBLEMS IN VALIDATING TEACHER SELECTION POLICIES AND PROCEDURES.

BY- MITZEL, HAROLD E.
NEW YORK CITY BOARD OF EDUCATION, BROOKLYN, N.Y.

PUB DATE JUN 67

GRANT OEG-1-6-061665-1624
EDRS PRICE MF-$0.25 HC-$0.28 5P.

DESCRIPTORS- CLASSROOM TECHNIQUES, EFFECTIVE TEACHING, *LESSON OBSERVATION CRITERIA, *PREDICTIVE ABILITY (TESTING), PREDICTIVE MEASUREMENT, PREDICTIVE VALIDITY, TEACHER EMPLOYMENT, *TEACHER RATING, *TEACHER SELECTION, TEACHING QUALITY, *TEST CONSTRUCTION, TEST VALIDITY,

IN DISCUSSING A RELIABLE AND VALID PRESERVICE PREDICTOR OF TEACHING EFFECTIVENESS, IT IS NOTED THAT MOST TEACHER SELECTION PROCEDURES DEPEND SOLELY ON PAPER AND PENCIL MEASURES OF VERBAL AND SYMBOL MANIPULATION ABILITY. THEREFORE, A NEW MULTIVARIABLE EVALUATION PROCEDURE IS OFFERED. IT CONSISTS OF (1) DEVELOPING EMPIRICAL DESCRIPTIONS OF CLASSROOM PERFORMANCE BY SYSTEMATIC RECORDING OF THE OBSERVABLE BEHAVIOR OF ALREADY EMPLOYED TEACHERS, (2) CONSTRUCTING A COMPREHENSIVE SAMPLE OBSERVATION INSTRUMENT FROM THE DATA OBTAINED BY STEP 1, (3) ADMINISTERING THE SAMPLE OBSERVATION DEVICE TO ALL TEACHER CANDIDATES IN A NATURALISTIC CLASSROOM SETTING (E.G., STUDENT TEACHING) SEVERAL TIMES, (4) EMPLOYING ALL CANDIDATES (NECESSARY FOR VALIDATION OF INSTRUMENT), (5) SYSTEMATICALLY OBSERVING AND SCORING THE BEHAVIOR OF ALL IN THE SAMPLE, (6) USING THE DATA OBTAINED TO RANK ORDER THE SAMPLE AND ESTABLISH A TRIAL SCORE DIFFERENTIATING BETWEEN A SATISFACTORY AND AN UNSATISFACTORY GROUP, AND (7) ANALYZING THE PRE-EMPLOYMENT AND EMPLOYMENT RATING OF INDIVIDUALS TO TEST THE PREDICTIVE VALUE OF THE INSTRUMENT. THIS DOCUMENT APPEARED IN GILBERT, H.B., AND LANG, G., "TEACHER SELECTION METHODS," 1967. (RF)

TEACHER SELECTION METHODS

Project No. 6-1665
Grant No. OEG 1-6-061665-1624

Harry B. Gilbert

Pennsylvania State University

Gerhard Lang

Montclair State College

June 1967

The research reported herein was performed pursuant to a grant with the Office of Education, U.S. Department of Health, Education, and Welfare. Contractors undertaking such projects under Government sponsorship are encouraged to express freely their professional judgment in the conduct of the project. Points of view or opinions stated do not, therefore, necessarily represent official Office of Education position or policy.

BOARD OF EXAMINERS
Board of Education of
The City of New York

New York, New York

Criterion Problems in Validating Teacher Selection Policies and Procedures

Harold E. Mitzel

The Pennsylvania State University

Instead of following my assigned topic for this brief paper with its emphasis on problems, and in order to stimulate discussion, I have chosen to propose a research and development procedure which would lead to a new set of teacher selection procedures. By adopting this stance, I hope to convey the notion that many of our so-called teacher selection criterion problems are of our own making and generally stem from an inadequate conceptualization of the task. My proposed solution is, I'm sure, not without difficulty in implementation. Further, I'm well aware that the proposal is deceptively simple, fantastically expensive, and politically difficult to bring about.

If we look at teacher selection as a data-gathering device instead of as a procedure, and if we view the result as multiple variables instead of as a single binary selected-rejected scale, then it is clear that we have been looking in the wrong behavior domain for the critically relevant information about teacher selection. Psychologically oriented disciplines have taught us that the best predictors of a person's future performance are variables based on his performance in a prior similar situation. Thus, we use a student's high school academic record as a major trustworthy indicator of his college success. In industry, the best predictor of performance on a specific job is the quality of a person's performance on previous closely related jobs. In the military, expensive simulators and training aids have been used as major sources of data for forecasting adequate performance under battle conditions. To be sure, the data derived from these sources are not perfect forecasters of direct on-the-job behavior, but as a class of variables they seem to stack up better than the currently used predictors of teaching success. The real problem of teacher selection is to choose from among all the candidates who have met the same eligibility standards (i.e., baccalaureate degree, passed a set of state-approved certification requirements) the ones who will provide the best pattern of teacher behaviors in the classroom. Of course, what teacher behaviors are wanted have to be identified and made explicit, but this task should be easily accomplished by the educational leadership in any given school system.

After screening out eligible teaching candidates with obvious physical and psychiatric defects, I consider that symbol

manipulating ability as assessed by paper-pencil devices is the major variable which saturates the selection procedures of most large school systems. Systems that use undergraduate grade point averages as indicators of teacher quality are undoubtedly tapping the same academic variable. Large school systems that employ a written test, particularly the short answer variety, are banking on the same general factor, but attempting to measure it as reliably as possible. Indeed, many of the dubiously valid interview rating techniques require the rater to form some sort of global assessment of mental ability. I am not advocating the selection of unintelligent teachers for our schools, but it does seem to me that we have exhausted bookish intelligence in its wide variety of forms as a viable predictor of teaching performance. I am suggesting that we turn to measures of classroom performance derived from systematic classroom observation as a data source for improving the selection of teachers. I am reminded of an analogy recently in the news. For years the annual slaughter on the nation's highways has been attributed to a class of variables associated with the personal adequacy of the driver, but suddenly, in spite of the powerful propaganda influence exerted by auto makers, it becomes evident that there is an important class of accident and injury related variables directly traceable to the mechanical condition of vehicles and their lack of safety engineering. For teacher selection, we need to discover a whole new set of variables.

To exploit the potential of a new and different set of variables for teacher selection first requires a decision to quit trying to find valid predictors of teaching performance in school grades, interview rating scales, oral and written examinations and U. S. Office of Education guidelines. Second, we must attempt to develop classroom performance variables in the context of the classroom situation, thus giving ourselves an optimum chance of producing valid predictors. From the standpoint of the scientific approach, the procedure is straightforward and can be listed in eight steps:

Step 1. Develop empirical descriptions of classroom performance by systematically recording samples of observable behavior for existing teaching staff members. These objective records of teacher behavior may be process-oriented or content-oriented, and preferably both. It is extremely important that the arbitrary global evaluations of "good teacher" and "poor teacher" be avoided in any attempt to record teacher behavior objectively. The research on systematic observation reported by Withall, Flanders and Amidon, Smith and Meux, Mitzel and Medley, and Bellack would make good starting points.

Step 2. Construct a comprehensive sample observation

instrument from the data obtained by Step 1. Here the judgment of educators who know the teaching process and the subject matter will have to be employed in order to select those behaviors which fit together to make up patterns deemed desirable by the school. For instance, if pupil participation in classroom activity planning is wanted in a school system, then it should be fairly easy to identify a cluster of specific teacher behaviors which clearly foster pupil participation. Conversely, a cluster of specific behaviors which inhibit pupil participation can also be identified. Experience shows rather conclusively that these behaviors can be put together with unit scoring to form a reasonably reliable scale.

Step 3. Administer the sample observation device to all teacher candidates in a naturalistic classroom setting (i.e., student teaching, internship). Of course, if some of the candidates are obviously unfit for reasons unrelated to their classroom performance (i.e., infection with a communicable disease, defect in moral character), then these may be eliminated in advance. This step is important since it enables the selection staff to obtain a wide range of behavior patterns. Research by Medley and Mitzel shows that it is necessary to obtain multiple samples of the observed behavior of a given teacher candidate in order for a reliable pattern to emerge.

Step 4. Employ all candidates. This is perhaps the hardest decision to make, even in times of a teacher shortage, but it is a necessary step in order to validate a new instrument.

Step 5. Systematically observe and score the on-job behavior of all teachers in the employed sample. The scoring should be done on a priori determined trial scales which were generated from the extensive observations of the behavior of in-house teachers. This step should be accomplished as soon as feasible after employment in order to minimize the self-selection effects of early dropouts.

Step 6. Using the observation data generated in Step 5 as a criterion and employing optimum weights for the several behavioral components, rank order the sample of new teachers and establish a trial demarcation between a satisfactory group and an unsatisfactory group. (The procedure itself does not demand that anyone be dismissed.)

Step 7. Analyze the pre-employment observation scale data against the on-job observation criterion obtained from the same individuals. This step makes it possible to assess the predictive power of the pre-employment observation information.

Step 8. Repeat the process beginning at Step 2 on a new sample of teacher candidates utilizing improved observation techniques and behavior scales.

Now the oversimplified eight-step process described above has a number of pitfalls. If the pre-employment teaching situation in which the predictive data are gathered has in it a lot of content-specific or situation-specific elements, then the predictors do not have a maximum opportunity to be closely related to the criterion. For example, if a candidate is observed during student teaching in a comfortable, middle-class, suburban school and then is measured on-the-job in a slum school, one would expect this difference in situation to have some impact on his instructional behavior pattern. Similarly, if the predictive measures for a candidate were obtained from a series of mathematics lessons and the criterion data were generated for the same person on several art lessons, some of the observed differences might well be attributed to the change in subject matters.

These problems of the specificity of schoolroom situation and content can probably be lessened by careful planning in large school systems.

That the use of classroom observation as a source of selection data would be a revolutionary development is confirmed by the recent survey reported by Gilbert and others. This 1966 report shows that, in almost sixty per cent of the large public school systems, candidates are not at all observed. In an additional twenty per cent, only one observation is made per candidate and the data derived from this procedure are undoubtedly a mass of subjective non-predicting rating scales which tell more about the rater than about the candidate. It seems to this observer that we must strike out in new directions on the task of predicting teacher performance and discontinue our perseveration with unproductive sources of data.