DOCUMENT RESUME

ED 087 757 SP 007 742

AUTHOR Popham, W. James
TITLE Alternative Teacher Assessment Strategies.
PUB DATE [73]
NOTE 7p.

EDRS PRICE MF-$0.65 HC-$3.29
DESCRIPTORS *Criterion Referenced Tests; *Evaluation Methods; *Performance Criteria; Teacher Dismissal; *Teacher Evaluation
IDENTIFIERS Competency; *Teaching Performance Test

ABSTRACT

This document, noting that teacher evaluation has now become a terror for teachers due to legislation such as the Stull Act, reviews the major assessment alternatives for teacher competence appraisal. The author discusses the use and merits of ratings, observations, and pupil test performance and finds them all to have fatal defects. He then describes as a final alternative the use of teacher performance tests, which he first advocated for use in the mid-sixties. The rationale behind this type of test is described as follows: since one of the major difficulties of comparing teachers for evaluation is that different teachers have different instructional emphasis, a teacher's ability to accomplish prespecified instructional objectives should be measured. The teaching performance test accomplishes this by providing an identical task for different instructors. Projects exploring this method are noted, but further experimentation is advocated. (JA)

ALTERNATIVE TEACHER ASSESSMENT STRATEGIES*

W. James Popham
University of California, Los Angeles
and
Instructional Objectives Exchange

At a time when the nation's entire educational enterprise is being subjected to increasing public scrutiny, it is not surprising that the classroom teacher -- generally conceded to be the pivotal figure in most instructional settings -- is being evaluated more frequently and more rigorously than ever before. Perhaps this increased stress on teacher evaluation stems from the widespread concern about accountability in education. Possibly it is the predictable concomitant of a glutted teacher market where employers can finally be choosy, when in former years their chief concern was to get an instructor, any instructor, to cover every classroom.

Whatever the cause, concerns about teacher evaluation have become far more pronounced in the past few years than at any time during this century, even though educational researchers have been continually carrying out teacher effectiveness investigations for well over 70 years. The difference in the focus of these activities provides the key to our understanding of why today's typical teacher starts to perspire a bit when someone mentions teacher competence assessment. In the old days, most teacher effectiveness researchers were searching for a suitable criterion variable which, if located, would permit them to isolate independent variables (such as teachers' personality traits, educational experience, or instructional styles) that would contribute to more effective instruction. Such investigations were accurately perceived by teachers as <u>research</u> inquiries and, as such, were not viewed as particularly threatening. Even in those instances where the attention of the investigator was clearly directed toward teacher evaluation, few teachers were very concerned. After all, even if defensible assessment techniques were discovered, it was generally held that teacher evaluation efforts would be directed toward helping teachers, never firing them. The American public had great confidence in the nation's public schools and, although everyone knew there were differences in the abilities of teachers, a tenured teacher's position was next to inviolate. For example, a recent search[1] of California's teacher employment records revealed that during the last 40 years <u>not one</u> California teacher has been dismissed on the grounds of incompetence. It is small wonder that in the past our teachers have not been too threatened by teacher evaluation activities.

*An invited working paper for a meeting of the Multi-State Consortium on Performance-Based Teacher Evaluation, New Orleans, February 25-28, 1973.

[1]Personal communication, Research Department, California Teachers Association, Burlingame, California.

But they are threatened now -- and with good reason. Dissatisfied legislators in a good many states are beginning to enact penalty-laden laws which call for more stringent teacher accountability. The most celebrated of these recent teacher accountability laws is the so called Stull Act (named after its author, Assemblyman John Stull) passed by the California legislature during the 1971 legislative session (Assembly Bill 293). The Stull Act has generated an immense amount of educational activity among California school people[2], for its implications are serious indeed. The new law calls for the annual evaluation of all probationary teachers and the biennial evaluation of all non-probationary teachers. The evaluations must be made on the basis, as stipulated by law, of pupil progress according to district-explicated standards of achievement in all areas of study. What has happened in California as a result of the Stull Act is that an attempt has been made to operationalize incompetence so that even tenured teachers can be dismissed if they are evaluated adversely. We can expect to see other state legislatures enacting comparable teacher evaluation laws in the next few years, particularly if the California experiment seems to be working.

But even if no more states established teacher appraisal systems, there is still a strong likelihood that local districts, perhaps buffeted by school board pressures, will set up some sort of teacher evaluation system. In view of these developments at the state and local levels it does not require much prescience to be able to forecast an increasing need for the technical devices and procedures required for effective teacher appraisal systems.

Although it is generally assumed by most laymen (and many legislators) that educators currently possess adequate devices for use in evaluating a teacher's instructional effectiveness, nothing could be farther from the truth. The history of teacher effectiveness research is replete with failure after failure in efforts to devise defensible indicators of how well a teacher teaches. Space limitations preclude an exhaustive analysis[3] of the limitations of previously tried assessment schemes, but each of the chief contenders, that is, ratings, observations, and pupil test performance, have fatal defects.

Ratings. Briefly, the difficulty with ratings of teacher effectiveness (characteristically supplied by administrators, but also obtainable from students, peers, etc.) is that different raters have different notions regarding what it is that constitutes good teaching. The same teacher who is rated high by one

---

[2] See Popham, W.J., "California's New Precedent-Setting Teacher Evaluation Law," Educational Researcher, Vol. 2, No. 7, July, 1972, pp. 13-15.

[3] For a more detailed examination of the strengths and weaknesses of various teacher effectiveness assessment approaches see McNeil, J.D. and Popham, W.J., "The Assessment of Teacher Competence," Chapter 7, second edition, The Handbook of Research on Teaching, R.M.W. Travers Ed., MacMillan, 1973.

individual because of such factors as "flexible interaction with learners" and "personable, informal rapport with class" may be rated low by another individual because of "poor discipline" and "classroom anarchy." We all use our private value matrix in judging whether good teaching has taken place, and when we try to pool these disparate sets of rater expectations, chaos is the characteristic result. How many times, for example, has a classroom teacher been rated negatively by a principal because the teacher was conducting class in a way other than the manner in which the principal recalled his/her lusterous days in the classroom. Yet that same teacher may receive a positive rating by the district office supervisor who has a different idea of how teaching should be carried on. It has been observed that one person's humorist is another person's smart aleck. Similarly, one rater's Mr. Chips is another rater's Mr. Peepers.

Observations. With respect to systematic observations of the teacher's classroom behavior we encounter an interesting assumption. It runs as follows: if certain process variables can be found to correlate positively with desired outcome variables, then by ascertaining whether those process variables are present, on that basis alone we can make judgments regarding the desired outcomes. For example, if it is discovered that a teacher's provision of practice opportunities for learners generally results in desirable learner attainment, then proponents of observational teacher evaluation schemes would contend that we can, at least in part, evaluate teachers on the basis of the degree to which they provide practice opportunities.

The trouble with the logic of this approach is its tendency to force one to the position that the process variables scrutinized by classroom observers are not only necessary for securing worthwhile results with learners, but that they are essentially sufficient. For if the phenomena observed, e.g., amount of teacher talk, are viewed as means to an end, why not assess the end results directly without encountering the measurement noise associated with the extra assessment step. For although no upstanding classroom observation devotee will ever assert that those behaviors observed are without exception associated with desired outcomes in learners (such as important cognitive or affective changes), the logic of the observation strategy pushes us to place greater reliance on means-end predictive relationships than the current sophistication of our observational techniques permits. If we are really interested in the ends, why not focus our assessment energy on them?

A second difficulty with observation-based approaches to teacher appraisal is that although a teacher may display optimal use of the classroom behaviors called for in the observation system, there may be deleterious factors present, factors not built into the observation structure, whose presence will essentially cancel out the positive features of the teacher's classroom behavior. The only way to head off this assessment difficulty would be to build an observation system so exhaustive that it could pick up all (or most) negative process variables, but by that time the system would be too vast to be practical.

Another difficulty with observational approaches to the assessment of teacher effectiveness is that whereas they might prove useful in identifying some classroom practices which in general will yield beneficial results with learners, the teacher evaluation game demands personal and particular decisions, not general guidelines. A particular teacher working toward particular goals with particular students in a particular setting may break all the process guidelines and yet achieve superb results. The particularized interaction effects are too subtle for our currently unsophisticated observation systems. There have been several outstanding pro football quarterbacks whose passing form looked abysmal, yet when the receiver arrived at the appointed spot the ball was always there waiting.

Finally, there is considerable danger that when the stakes are high enough (and job security represents a big bet), many teachers will "fake good." Observation evaluation systems are particularly susceptible to such faking, for in these days of openly described criteria we can expect teachers to know what factors will be involved in the observation system. Indeed, any diligent and legally informed teachers organization should be easily able to unearth the observation dimensions involved. Having been apprised of what practices yield positive evaluations, is it so unrealistic to expect that teachers will tend toward the use of those practices when under observation? Of course, if one wished to employ constant monitoring of classroom behavior through such devices as closed circuit television, then such fakeability fears would be vitiated, but by then most schools would have been closed permanently because of the anti-1984 teachers' strikes.

Pupil Test Performance. The chief deficiency with the use of student test performance as an index of teacher proficiency has generally been that the wrong kinds of tests were employed. Since 1900 most teacher effectiveness research in which pupil test performance was employed as a criterion variable involved the use of standardized achievement tests. And since most standardized tests were designed to serve a different purpose, namely, to permit comparisons among individual learners (not among teachers) they invariably resulted in a "no significant difference" outcome.

The difficulties with standardized or norm-referenced tests, particularly for teacher evaluation, have been treated elsewhere[4], but their two most visible defects can be briefly identified. First, since commercially developed standardized tests must serve students throughout an entire nation, the generalized nature of their content coverage is often inconsistent with local curricular emphases. Incongruent measurement and curriculum results in misleading data. Second, certain psychometric properties of norm-referenced tests (such as their heavy reliance on producing among-learner variance) leads to tests which are sometimes insensitive to detecting the results of high quality instruction.

---

[4]See, for example, Popham, W.J., "Domain-Referenced Measurement and Teacher Evaluation," Educational Technology, in press: Glaser, Robert, "A Criterion-Referenced Test," Criterion-Referenced Measurement: An Introduction, W.J. Popham (Ed.), Educational Technology Publications, Englewood Cliffs, N.J., 1971, pp. 41-51.

In the past few years the development of criterion-referenced (or mastery) tests offers teacher evaluators an alternative to standardized tests for assessing an instructor's impact on learners. The judicious employment of criterion-referenced tests for teacher evaluation purposes is only beginning to be seriously investigated.

Teaching Performance Tests. In the mid-sixties the writer had reached a point of frustration regarding teacher effectiveness assessment devices and, after a reappraisal of alternative assessment strategies, had proposed the development of an alternative approach to solving this problem, namely, through the use of a teaching performance test. Two separate projects[5] were supported by the U.S. Office of Education, each designed to develop and attempt to validate teaching performance tests in different subject matter fields. While the rationale underlying the teaching performance test strategy, as well as the detailed results of these two projects are supplied elsewhere[6], a brief description of the performance test approach can be supplied here.

One of the major difficulties in comparing teachers for purposes of instructor evaluation is that different teachers have different instructional emphases, thereby making across-the-board comparisons misleading. The teaching performance test counteracts this problem by providing an identical task for different instructors, namely, the ability to accomplish prespecified instructional objectives. The teaching performance test is built on the general premise that one chief reason for a teacher's existence in the classroom is to bring about worthwhile changes in students, that is, changes in their knowledge, attitudes, skills, etc. To the extent that this is true, then one criterion by which a teacher should be judged is his or her ability to bring about such changes. By providing identical instructional objectives for teachers, then giving the teachers an opportunity to accomplish those objectives using whatever instructional techniques they wish, a measure of the teacher's ability to accomplish given objectives can be provided. One might wish to argue that the better achiever of given objectives will also be the better achiever of his/her own objectives, but this is a question which can be answered empirically. If one simply decides that an important criterion of teaching is the ability to accomplish instructional objectives, then teaching performance tests would appear to have some utility in a data-based evaluation matrix.

The steps involved in a teaching performance test are these: (1) the teacher is provided with an explicit instructional objective (and sample test item) along with any background information necessary to become familiar with the subject matter related to

---

[5]Performance Tests of Instructor Competence for Trade or Technical Education, USOE Cooperative Research Contract No. OE-5-85-051; Development of a Performance Test of Teaching Proficiency, USOE Cooperative Research Contract No. 3200.

[6]Popham, W.J., "Performance Tests of Teaching Proficiency: Rationale, Development, and Validation." American Educational Research Journal, January, 1971, 8 (1), pp. 105-117.

that objective; (2) the teacher plans a lesson designed to accomplish the objective; (3) the teacher instructs a group of learners, typically a small group of learners for a short period of time; (4) the learners are posttested with an examination based on the objective. The examination has not previously been seen by the teacher but its nature is readily inferable from the objective (and sample test item) previously given to the teacher.

In the previous USOE-supported research studies described above, the purpose of developing the performance tests was primarily research-oriented, that is, it was anticipated that these devices would be employed principally for research purposes such as the identification of relevant independent variables. Consistent with that intent, the performance tests involved in those investigations consumed a fairly large amount of learner instructional time, ranging from four to ten hours. At the conclusion of those investigations, it became clear that if teaching performance tests were to prove practical for teacher evaluation or instructional improvement purposes, they would have to be developed for much shorter periods of instructional time. As a consequence, the writer's recent development work with performance tests has featured instruments which take only 15 minutes of instructional time and are designed to be used with small groups of adults or younger learners. These teaching performance tests, frequently referred to as instructional minilessons, superficially appear comparable to the microteaching exercises developed at Stanford University some years back. In rationale, however, they are quite different. The Stanford microteaching lessons emphasize the teacher's acquisition of process skills, e.g., good questioning techniques. The instructional minilessons referred to here, on the other hand, focus more heavily on the results of the teaching than upon the instructional procedures themselves.

During the past few years teaching performance tests have been employed both in preservice and inservice teacher education settings.[7] Generally speaking, these performance tests have been of the short duration alluded to above, i.e., 15 to 20 minutes in length. But while these devices appear useful in instructional settings, for example, in helping prospective teachers become more facile at accomplishing prespecified instructional objectives, their utility for purposes of teacher evaluation has been largely unstudied.

In a recent paper[8] Glass has proferred the notion that teaching performance tests may have insufficient reliability to permit their effective use in teacher evaluation enterprises. Glass cited

---

[7]Popham, W.J., Applications of Teaching Performance Tests to Inservice and Preservice Teacher Education. A paper presented at the annual meeting of the American Educational Research Association, New Orleans, February 26-March 1, 1973.

[8]Glass, Gene V, Statistical and Measurement Problems in Implementing the Stull Act, Stanford University Invitational Conference on the Stull Act, October 1972, Palo Alto, California.

several investigations in which the reliability of teaching performance tests was clearly inadequate. Several of the investigations cited, however, had been conducted as doctoral dissertations or by novice researchers. The reliability of teaching performance tests is as yet a seriously unstudied matter. For one thing, the teaching performance tests used in those investigations have been constructed on an almost opportunistic basis, that is, whatever topics, objectives, etc., have come to the investigator's mind. No attempt has been made to carefully delineate the truly critical dimensions in teaching performance tests. Beyond that, only one investigator[9] has carefully attempted to study the reliability of even these ill-defined performance tests. Results of this investigation will be reported by Millman at the 1973 meeting of the American Educational Research Association. Examination of the Millman findings suggest that the reliability evidence, once again, is not encouraging. But, as indicated above, the nature of the performance test employed in that investigation was not rigorously explicated.

When this paper was solicited as one of several dealing with the "state of the art" in the assessment strategies suitable for performance-based teacher education I had just completed the final draft of an AERA paper describing a set of minimal competencies for a performance-based teacher education program.[10] I had even sketched alternative assessment procedures for each of the competencies. Now I just couldn't bring myself to re-write the paper or even to subtly paraphrase my original paper. I try to restrict my paraphrasing talents to the writing of others, not my own.

Accordingly, in the present effort I have attempted to focus exclusively on the major assessment alternatives for teacher competence appraisal. Since if performance-based teacher education programs cannot demonstrate that their competency-armed products are indeed better teachers, then the performance-based teacher education folk had best fold up their competencies and slip away into the night.

---

[9] Millman, Jason, Psychometric Characteristics of Performance Tests of Teaching Effectiveness. A paper presented at the annual meeting of the American Educational Research Association, New Orleans, February, 1973.

[10] Popham, W.J., Identification and Assessment of Minimal Competencies for Objectives-Oriented Teacher Education Programs. A paper presented at the annual meeting of the American Educational Research Association, New Orleans, February, 1973.