DOCUMENT RESUME

ED 340 109 EA 023 529

AUTHOR Ramos-Cancel, Maria L.; Duttweiler, Patricia Cloud, Ed.

TITLE Issues in Perspective, Assessment Issues Papers 18-23.
INSTITUTION Southwest Educational Development Lab., Austin, Tex.
SPONS AGENCY Office of Educational Research and Improvement (ED), Washington, DC.
PUB DATE Nov 86
CONTRACT 400-86-0008
NOTE 15p.; For other papers in the series, see ED 273 617; for the document on which these papers are based, see ED 272 511.
PUB TYPE Reports - Descriptive (141)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Academic Achievement; *Classroom Observation Techniques; Elementary Secondary Education; *Faculty Evaluation; *Incentives; Merit Rating; Minority Groups; Observation; *Performance; *Teacher Evaluation

ABSTRACT

The papers in this collection are based on the document "Perspectives on Performance-Based Incentive Plans" and offer brief overviews of the following issues in education: (1) performance-based evaluation in relation to incentive pay plans; (2) recommendations for success of performance-based evaluation systems; (3) the use of classroom observation ratings in performance-based teacher evaluations; (4) problems associated with the observation/rating procedure; (5) the use of student achievement measures as indicators of teacher performance; and (6) some concerns about the effects of performance-based evaluation on minority teachers. References accompany each paper. (LMI)

Issues in Perspective

Assessment Issues Paper #18

November 1986

THE ISSUE: PERFORMANCE-BASED EVALUATION IN RELATION TO INCENTIVE PAY PLANS

Maria L. Ramos-Cancel

Education reform initiatives have drawn renewed attention to the practice of performance-based evaluation. Assessment of teacher performance is viewed as crucial to quality education. Furthermore, as incentive plans for teachers are examined, performance-based evaluation becomes even more important.

Interest in performance-based evaluation existed long before the current furor over incentive plans. States and school districts have developed and used performance-based evaluation systems for many reasons unrelated to incentives. Many methods for assessing performance exist, but not all are suitable or appropriate for evaluating teachers for reward purposes.

In order to determine which methods will help to meet the goals of an incentive plan, the theoretical basis of evaluation needs to be examined. Also, the experiences of researchers and practitioners in developing and using performance-based evaluation systems, in general, need to be considered.

Rationale for Performance-Based Evaluation

Performance-based evaluation, as used here, refers to evaluation that focuses on assessing on-the-job performance rather than on assessing attributes or characteristics believed to predict this performance.

Using performance-based evaluation as a basis for apportioning rewards presupposes a rationale for why awards are being made and why they are linked to performance-based evaluation. The main goal of most incentive plans is to recognize teachers by rewarding them for outstanding performance. Thus, performance-based evaluation is needed for determining who is eligible for the rewards. Effective incentive plans depend on performance-based evaluation in the following ways (Barro, 1985):

- Valid, reliable, and fair measures of teacher performance are needed to guarantee that the "right" teachers are rewarded.
- Accurate measurement of teacher performance is essential to ensure that the incentive system will encourage effective teaching behaviors and discourage undesirable behaviors.
- The quality of the measurement method determines, in many respects, how the rest of the incentive system can be structured.

Summative vs. Formative Evaluation

In the context of incentives, educators are more concerned with summative evaluation than with formative evaluation (Barro,

1985). Formative or improvement-related evaluation is primarily concerned with descriptive information that identifies sources of difficulty and viable courses for change. The main purpose of summative or incentive-related evaluation is to determine how well teachers are performing rather than to diagnose teaching problems. Therefore, incentive-related evaluation must be capable of yielding objective, standardized, and externally defensible information about teacher performance (Wise, Darling-Hammond, McLaughlin, & Bernstein, 1984).

Process- vs. Product-Oriented Evaluation

Performance-based evaluation may focus on assessing actual performance (e.g., what teachers actually do), or it may focus on assessing the effects of this performance. The former evaluation approach is generally referred to as process-oriented and the latter as product- or outcome-oriented. In process-oriented evaluation, teachers are judged by *what they do*, whereas in product- or outcome-oriented evaluation teachers are judged by *the effects of what they do*. Proces oriented evaluation generally relies on teacher performance measures while product-oriented evaluation generally relies on student achievement measures. Whether to evaluate teachers by what they do, by the effects of what they do, or by a mixture of the two, is considered the main design issue concerning teacher evaluation (Barro, 1985). The specifics of the assessment instruments, procedures, standards, etc. will depend on which approach is selected.

Technical Requirements of Evaluation

Performance-based evaluations are expected to meet certain technical requirements. They are expected to provide evaluation results that are valid and reliable. However, the level of technical quality can vary depending upon the purposes of evaluation. For example, reliability should be higher in summative evaluations than in formative evaluations (Braskamp, 1980).

The validity and reliability of the evaluation results depend on the instruments and procedures used to assess performance. The results are considered valid if they were obtained with instruments and procedures that accurately and adequately measure the particular performance domains (behaviors or outcomes) that they were supposed to measure. The results are considered reliable if they were obtained with instruments and procedures that are consistent in their measurements -- that is, they always produce the same results. Both properties, that of being valid and reliable, are important in the defensibility of performance-based evaluations.

Discriminating Power of The Measurement System

In the context of incentives, methods for differentiating superior or excellent performance from average performance is required. It is not sufficient merely to determine whether teachers are minimally competent. The measurement system should be able to distinguish gradations of performance and, therefore, be able to discriminate between teachers relatively close together on the performance spectrum. An example of a system with too low discriminating power to be useful for apportioning rewards is one that can distinguish reliably only between "unsatisfactory," and "satisfactory" performance (Barro, 1985). A system with higher power might classify teachers into three, four, or more performance strata, each of which could then be associated with a different level of reward.

Unbiased and Fair Evaluation

Performance-based evaluations should be unbiased and fair. To be unbiased, evaluations should be unaffected by relationships between assessors and assessee and should neither penalize nor give undue advantage to persons by virtue of sex, age, race, ethnicity or other personal characteristics. Also, to be unbiased, the evaluations should be minimally dependent on the subjective judgment of any individual. Where subjective judgment is unavoidable, as in the case of evaluations that rely on

classroom observations, performance should be judged by multiple evaluators (Barro, 1985).

Fairness involves the sensitivity of the evaluation toward differences among people and situations. For example, identical procedures and standards cannot be applied mechanically to both beginning and experienced teachers, to elementary and secondary teachers, to special education and science teachers. Fairness also involves recognizing the complexity of teaching and the possibility that different assessment results may be the result of using different methods to evaluate (Braskamp, 1980).

Credibility and Acceptability of Evaluation

Credibility involves mutual trust between those being evaluated and those using the evaluative information. Since credibility is assessed from the perspective of those being evaluated rather than from the perspective of the evaluator, the external constraints under which the evaluation is to be implemented become important (Braskamp, 1980). If the information derived from the evaluation procedure is not viewed as credible, it will not be highly regarded or accepted by those being evaluated even if the information is technically sound. Therefore, the evaluation system must ensure credibility as it addresses the concerns of teachers.

Effects of Evaluation on Instruction

Performance-based evaluation may have both positive and negative effects on the instructional process and the condition of the schools (Barro, 1985). Possible beneficial effects include reinforcement of state or district educational priorities and direct stimulation of improved teacher performance. For example, if teacher rewards depend on student achievement in specific areas, this may help to enforce compliance with the curriculum and with the priorities officially assigned to different subjects of instruction. Furthermore, measuring and comparing teacher performance might induce teachers to do better even in the absence of performance-contingent rewards.

Possible adverse effects include making teaching methods more rigid and inhibiting innovative practices. Also, heavy reliance on achievement testing could distort the content of teaching. Teachers might be motivated to emphasize unduly those areas of the curriculum that count toward evaluations. There is likely also to be extensive "teaching to the test" -- a phenomenon that could be either desirable or undesirable depending on how well the test reflects the full range of instructional goals (Barro, 1985).

Overall Evaluation Climate

The assessment of job performance, particularly of teachers, often generates a threatening climate. Performance assessment touches on an emotionally charged activity -- the assessment of a person's competence. The signals that the person receives about this assessment may have a strong impact on his or her self esteem (Thompson & Dalton, 1970). It needs to be recognized that although some anxiety might be needed for progress, too much anxiety may be destructive to both the school system and the evaluatee as a person.

This problem can be further complicated in situations where rewards are linked to evaluation. The fact that rewards might be given to a very limited number of teachers may cause unwanted levels of competition that could lead to anger and resentment. Therefore, ways of creating a more positive measurement climate and of reducing the possibility of unwanted levels of competition and anxiety need to be considered in incentive-related evaluations.

Constraints on Evaluation

A comprehensive evaluation process that meets all the technical, institutional and personal requirements may be difficult or impossible to obtain because of time and financial constraints. The demand for eval-

uation is a demand for the expenditure of time (Gephart, 1980). This is particularly so in the case of process-oriented evaluations where performance assessment tends to rely on classroom observation. Evaluations that require extensive and repeated classroom observations are likely to require a great deal of evaluators' time (Barro, 1985).

On the other hand, product-oriented evaluations that rely on student achievement measures may require elaborate and specialized testing programs. These programs can be very costly and also time-consuming for students and teachers. How to assess performance adequately but at a reasonable cost and with a reasonable amount of time investment is one of the most difficult problems in designing a performance-based evaluation system.

Summary

In spite of the fact that great strides have been made in performance assessment, both those who are being evaluated and those who are doing the evaluations are still uncomfortable about it (Gephart, 1980). There is still some uncertainty about how teacher performance can best be assessed and whether the existing instruments and procedures are suitable or appropriate for use in evaluations that are linked to rewards. In order to determine which methods, instruments, and procedures will help to meet the goals of the incentive plans, the theoretical basis of evaluation needs to be examined. Also, the experiences of researchers and practitioners in developing and using performance-based evaluation systems, in general, need to be considered.

References

Barro, S.M. (1985). The logic of teacher incentives. Washington, DC: The National Association of State Boards of Education.

Braskamp, L.A. (1980). Guidelines for using rating scales and checklists. In W.R. Duckett (Ed.), Observation and the evaluation of teaching (pp. 97-103). Bloomington, IN: Phi Delta Kappa.

Gephart, W.U. (1980). Components of a human system for evaluating teaching. In W.R. Duckett (Ed.), Observation and the evaluation of teaching (pp. 117-129). Bloomington, IN: Phi Delta Kappa.

Wise, A., Darling-Hammond, L., McLaughlin, M.W., Bernstein, H.T. (1984). Teacher evaluation: A study of effective practices. Santa Monica, CA: Rand.

Thompson, P., & Dalton, G. (1970, January/February). Performance appraisal: Managers beware. Harvard Business Review, 48, 149-159.

These assessment issues papers, edited by Patricia Cloud Duttweiler, are based on the SEDL document, Perspectives on Performance-Based Incentive Plans. For information on purchasing that document, please contact SEDL's Publication Office.

A Product of
Southwest Educational Development Laboratory
211 East Seventh Street
Austin, Texas 78701 (512) 476-6861

This publication is based on work sponsored wholly, or in part, by the Office of Educational Research and Improvement, U S Department of Education, under Contract Number 400-86-0008. The content of this publication does not necessarily reflect the views of OERI, the Department, or any other agency of the U.S. Government.

Issues in Perspective

Assessment Issues Paper # 19

November 1986

THE ISSUE: RECOMMENDATIONS FOR THE SUCCESS OF PERFORMANCE-BASED EVALUATION SYSTEMS

Maria L. Ramos-Cancel

The literature on teacher evaluation offers various recommendations that can serve as general guidelines for ensuring the quality and legality of teacher assessments. The recommendations listed here are based on the experiences of researchers and practitioners in developing and using performance-based evaluation systems inside and outside the educational milieu. They were extracted from different sources (e.g., Braskamp, 1980; Holley and Field, 1977; Holley, 1983-84; Rumery, 1985).

- The purpose of the evaluation system should be clear, and the evaluation procedures should be in line with this purpose.
- Instruments and procedures used to measure teacher performance must meet technical standards.
- Where possible, evaluations should be based on observable job behaviors.
- The evaluation system must focus on behaviors that truly matter in teaching and learning.
- Evaluations should be supported by objective evidence of performance results.
- Inputs should be solicited from various groups (e.g., faculty members, parents, and administrators) when selecting factors to be used in conducting evaluations.
- Top management must give the evaluation system its full support.
- The evaluation system should encourage self-evaluation and self-improvement in job performance.
- The evaluation system should include a variety of techniques for assessing performance.
- Evaluations should not be discriminatory in intent, application, or results.
- The evaluative process should be carried out on a regular, continuing basis and include opportunities for both formal and informal evaluations.
- School systems should develop policies pertaining to the use of performance-based evaluation, procedures for administration, rules governing decisions, etc., and these should be communicated to the persons who will be evaluated.
- The evaluation process should be carried out on a regular, continuing basis, and should include opportunities for both formal and informal observations.
- The evaluation process should complement, not usurp, the existing organization and flow of decision making within the organization.

Sponsored by the Office of Educational Research and Improvement US Department of Education

- The school district should regularly assess the quality of evaluation, including individual and collective evaluator competence, and provide feedback to individual evaluators and input into the continuing evaluator-training process.

- The persons evaluated should be informed of the results of their evaluations as soon as possible.

- Criteria and procedures to be used in the evaluations should be communicated to the persons being evaluated.

- Performance must be assessed before any personnel decisions are made that are said to be contingent on performance-based evaluation.

- Policies pertaining to the use of performance-based evaluations should be communicated to those who will be evaluated.

- Staff members should be made aware of their right to appeal unfavorable evaluations through channels to the superintendent and, ultimately, through the School Board.

References

Braskamp, L.A. (1980). Guidelines for using rating scales and checklists. In W.R. Duckett (Ed.), Observation and the evaluation of teaching (pp. 97-103). Bloomington, IN: Phi Delta Kappa.

Holley, F.M. (1983-84). Personnel evaluation: Essentials for success. In Center on Evaluation, Development and Research Hot Topics Series, Merit pay and evaluation (pp. 187-189). Bloomington, IN: Phi Delta Kappa.

Holley, W.H., & Field, H.S. (1977). The law and performance evaluation in education: A review of court cases and implications for use. Journal of Law and Education, 6, 429.

Rumery, R.E. (1985). Measuring teacher merit: What we can learn from business and industry. In H.C. Johnson, Jr. (Ed.), Merit, money and teachers' careers. Lanham, MD: University Press of America.

These assessment issues papers, edited by Patricia Cloud Duttweiler, are based on the SEDL document, Perspectives on Performance-Based Incentive Plans. For information on purchasing that document, please contact SEDL's Publication Office.

A Product of
Southwest Educational Development Laboratory
211 East Seventh Street
Austin, Texas 78701 (512) 476-6861

This publication is based on work sponsored wholly, or in part, by the Office of Educational Research and Improvement, U.S. Department of Education, under Contract Number 400-86-0008. The content of this publication does not necessarily reflect the views of OERI, the Department, or any other agency of the U.S. Government.

Issues in Perspective

Assessment Issues Paper # 20

November 1986

THE ISSUE: USE OF CLASSROOM OBSERVATION RATINGS IN PERFORMANCE-BASED TEACHER EVALUATIONS

Maria L. Ramos-Cancel

Efforts to assess teacher performance generally rely on performance ratings based on classroom observations. In the long history of evaluating teachers through classroom observation, different observation approaches have been used. Emphasis is now being given, however, to the more formal, systematic observation techniques.

Rating Procedures

Performance rating procedures can be structured or unstructured. The unstructured procedure generally requires a high level of inference by the rater. The rater observes the teacher and notes behaviors that seem relevant. The rater combines his or her impressions into a composite picture that is then compared to the rater's personal standards of effective performance. This comparison leads to a specific rating for the teacher. There are no records of the behaviors or the standards; therefore, sources of rating problems cannot be identified.

The structured procedure generally requires a lower level of inference. It is preferred over the unstructured procedure because it tends to be more objective (Soar, 1983-84). The rater uses a checklist or a scale to guide the observation. Both the checklist and the scale contain a set of items that define specific behaviors (or categories of behaviors). These specific behaviors constitute the performance criteria on which the observation and consequent rating will be based.

Use of Checklists and Scales

A checklist allows the rater to indicate whether the target behaviors have occurred. However, qualitative judgments cannot be made about the behaviors, at least not without additional data. Thus, ratings are based exclusively on the presence or absence of these behaviors (Ingle, 1980).

A scale allows the rater not only to indicate whether or not the behaviors occurred but also to make qualitative judgments about these behaviors (Ingle, 1980). The criteria for the judgments that lead to a specific rating have been established beforehand and incorporated into a public, agreed-upon rating key that is then applied to the behaviors. These judgments can appear in the form of category responses that are provided for each item. Category responses may be evaluative adjectives, frequency adverbs, or verb phrases indicating agreement with the statement presented in the item. They may also be phrases that describe or modify the statement presented in the item. While other response modes have been used, the category response mode is the most common (Rumery, 1985).

Sponsored by the Office of Educational Research and Improvement US Department of Education

Systematic Observation Technique

Systematic observation differs from the less formal administrator and peer visits that have long been the principal means for evaluating teachers. Systematic observation requires the following (Peterson & Kauchak, 1983):

1) trained and monitored observers;

2) a reliable and representative number of observation visits;

3) demonstrably fair sampling of behavior;

4) limited observational categories;

5) systematic data recording and analysis proceedures; and

6) a conceptually coherent framework for the interpretation of the data.

Use of Ratings for Apportioning Rewards

Whether the ratings can be used as a basis for apportioning rewards depends on the degree to which the rating system is able to distinguish gradations of performance. The power of the rating system to discriminate among degrees of above-average performance is unclear. This discriminating power is believed to derive, ultimately, from the reliability of the rating procedure; thus, if its reliability is low, its ability to discriminate is also low (Barro, 1985). However, the discriminating power of some rating systems is limited by design in cases where the system allows only "satisfactory" or "unsatisfactory" ratings to be assigned to teachers.

References

Barro, S.M. (1985). The logic of teacher incentives. Washington, DC: The National Association of State Boards of Education.

Ingle, R.B. (1980). Techniques for systematic observation. In W.R. Duckett (Ed.), Observation and the evaluation of teaching (pp. 106-111). Bloomington, IN: Phi Delta Kappa.

Peterson, K., & Kauchak, D. (1983). Progress on development of lines of evidence for the evaluation of public school teaching. Paper presented at the annual meeting of the American Educational Research Association, Montreal.

Rumery, R.E. (1985). Measuring teacher merit: What we can learn from business and industry. In H. C. Johnson, Jr. (Ed.), Merit, money and teachers' careers. Lanham, MD: University Press of America.

Soar, R. S. (1983-84). Measures of quality in the classroom. In Center on Evaluation, Development and Research Hot Topics Series, Merit pay and evaluation (pp. 109-116). Bloomington, IN: Phi Delta Kappa.

These assessment issues papers, edited by Patricia Cloud Duttweiler, are based on the SEDL document, Perspectives on Performance-Based Incentive Plans. For information on purchasing that document, please contact SEDL's Publication Office.

A Product of
Southwest Educational Development Laboratory
211 East Seventh Street
Austin, Texas 78701 (512) 476-6861

This publication is based on work sponsored wholly, or in part, by the Office of Educational Research and Improvement, U.S. Department of Education, under Contract Number 400-86-0008. The content of this publication does not necessarily reflect the views of OERI, the Department, or any other agency of the U.S. Government.

SCDI

Issues in Perspective

Assessment Issues Paper # 21

November 1986

THE ISSUE: PROBLEMS ASSOCIATED WITH THE OBSERVATION/RATING PROCEDURE

Maria L. Ramos-Cancel

There are several problems associated with the use of the observation/rating procedure that may threaten the validity of performance-based teacher evaluations. These problems have important implications for systems that reward teachers for what they do in the classroom. The following are some examples:

- The two, three, or four observation visits typically performed during a year for each teacher may not be sufficient to sample teaching in all major subject areas taught by one teacher or to sample teaching in different situations and with different classes (Barro, 1985). Also, this small number of visits does not allow for sampling variations.

- The presence of any observer, but especially an evaluator, changes classroom activity (Barro, 1985). Intentional, even rehearsed, artificial behaviors may be elicited if the observation is expected. However, unexpected observation also disrupts the classroom environment, leading to unnatural behaviors from both teachers and students.

- The content or format of the rating instruments may not allow raters to translate readily the job-related behaviors they observe to a specific level on a specified dimension (Borman, 1978). Rating scales with insufficiently concrete items or response categories may force raters to lump vaguely related observations together, thus producing performance measures that do not reflect "true" job-related behaviors.

- Raters' opportunities to observe relevant job-related behaviors are an obvious prerequisite for obtaining satisfactory performance ratings. Yet, when supervisors are asked to rate subordinates, they often lack the opportunity to rate all behaviors relevant to job effectiveness. Under these conditions, raters might be required to provide ratings on dimensions of behavior they have not had an opportunity to observe. They are likely to comply with this requirement by generalizing from previous observations of a teacher's behavior or from their general impressions of a teacher to the unobserved behavior (Borman, 1978).

- Raters' knowledge of common rating errors and of methods for reducing them are important ingredients for obtaining high-quality ratings. Yet, often raters who are inexperienced in performance appraisal and ignorant of common sources of error or of ways to reduce them (Borman, 1978) are required to rate others.

- The rater may have a general tendency to make an overall judgment about the person being rated and to record

OERI

Sponsored by the Office of Educational Research and Improvement US Department of Education

consistently favorable or unfavorable ratings on a number of items intended to describe different aspects of performance. The rater may fail to discriminate among conceptually distinct and potentially independent aspects of a person's behavior (Saal, Downey, & Lahey, 1980).

- Raters may differ in the way they consistently rate those they observe. Some raters may have a general tendency to assign a higher or lower rating than is justified by the observed performance. Other raters may have a general tendency to cluster the ratings around the midpoint of a rating scale, avoiding extreme categories; or to cluster the ratings around any point on a rating scale, high, low, or midpoint (Saal, Downey, & Lahey, 1980).

- The likelihood of relationship biases is especially great when the evaluator is the building principal or another teacher from the same school. Both principal and peers may have interests unrelated to teaching performance in whether the person succeeds or fails, advances or falls behind, or remains in or leaves the school (Barro, 1985).

- The rater may have biases related to teaching styles. These biases can be particularly insidious because the rater may believe that his or her personal preferences reflect valid distinctions among more and less effective modes of teaching (Scriven, 1981). In some respects, these generic biases are more troubling than the relationship ones. While the latter can be avoided by selecting the raters appropriately, the former are much more difficult to weed out (Barro, 1985).

- Organizational constraints may cause ratings to reflect organizational demands rather than true levels of performance exhibited by ratees. A supervisor might hesitate to provide a deserved low rating to an employee in order to avoid confrontation with a disgruntled employee, or avoid the burden of replacement (Borman, 1978).

References

Barro, S.M. (1985). The logic of teacher incentives. Washington, DC: The National Association of State Boards of Education.

Borman, W.C. (1978). Exploring the upper limits of reliability and validity in job performance ratings. Journal of Applied Psychology, 63, 135-144.

Saal, F.E., Downey, R.G., & Lahey, M. A. (1980). Rating the ratings: Assessing the psychometric quality of ratings. Psychological Bulletin, 88, 413-428.

Scriven, M. (1981). Summative teacher evaluation. In J. Millman (Ed.), Handbook of teacher evaluation. Beverly Hills: Sage.

These assessment issues papers, edited by Patricia Cloud Duttweiler, are based on the SEDL document, Perspectives on Performance-Based Incentive Plans. For information on purchasing that document, please contact SEDL's Publication Office.

A Product of
Southwest Educational Development Laboratory
211 East Seventh Street
Austin, Texas 78701 (512) 476-6861

This publication is based on work sponsored wholly, or in part, by the Office of Educational Research and Improvement, U.S. Department of Education, under Contract Number 400-86-0008. The content of this publication does not necessarily reflect the views of OERI, the Department, or any other agency of the U.S. Government.

SCDE

Issues in Perspective

Assessment Issues Paper # 22

November 1986

THE ISSUE: USE OF STUDENT ACHIEVEMENT MEASURES AS INDICATORS OF TEACHER PERFORMANCE

Maria L. Ramos-Cancel

The use of student achievement measures as indicators of teacher performance is gaining increased attention. However, there are several problems associated with using measures of student achievement as part of the assessment procedure for incentive-pay or career-ladder promotion plans.

Definition of Student Achievement

The term "student achievement" is used to refer to whatever a student knows and can do in a specified subject area as a consequence of instruction. It is distinguished from the term "student competency," which is used to refer to whatever a student knows and can do in a subject area independently of how the knowledge and skill are acquired (Messick, 1984).

Problems With Using Objective Tests As Measures of Student Achievement

The most practical tools for objectively determining student achievement are written examinations. Their use in assessing teacher performance is limited, however, by the fact that well-established and accepted objective tests of student achievement are concentrated mainly in such basic-skill areas as reading, language, and mathematics.

Even at the elementary level, one cannot judge teachers fairly by students' progress in basic skills alone. At the secondary level, teaching basic skills is peripheral to most teachers' assignments. Consequently, teacher evaluation based on the outcomes of teaching would require much broader-ranging achievement testing than is now the practice in most states and school systems (Barro, 1985).

Standard achievement tests are also unlikely to reflect the full range of instructional goals in their subject areas. In particular, they rarely test for the learning of those higher-order skills that presumably follow from superior teaching. Therefore, even where the relevant subject areas appear to be "covered" by existing tests, it cannot be taken for granted that the products of teaching are being adequately or completely measured.

In addition, in designing teacher evaluation systems that rely on written examinations to determine student achievement, student test performance must be taken into account. It is not achievement alone that is being measured but also student test-taking ability (Haertel, 1986).

Non-Teacher Influences on Achievement

It is not easy to separate teacher contributions from other influences on students' learning. The following are examples of non-teacher factors that may influence student achievement (Barro, 1985):

OERI

Sponsored by the Office of Educational Research and Improvement US Department of Education

1) the characteristics of the students themselves (e.g., their abilities, prior educational experiences, economic circumstances, home environments, interests, attitudes, learning styles, and prior knowledge);

2) the resources available to the teacher (e.g., materials and supporting staff); and

3) school circumstances that are external to the classroom and are not under the control of the teacher.

These factors need to be taken into account when comparing teachers on the basis of their contributions to student progress. If this is not done, the validity and fairness of the comparison may be questioned.

Control for Non-Teacher Factors

If student achievement measures are to be used as indicators of teacher performance, methods of controlling for non-teacher factors need to be examined. Specific statistical methods that yield adjusted achievement-gain scores have been suggested. These adjusted achievement-gain scores are, in essence, statistically based predictions of the gains each teacher would have produced with a "typical" class in a "typical" teaching situation (Barro, 1985). The adjusted scores, rather than the original raw scores, would then be used to determine which teachers deserve performance-based rewards.

Other methods have also been suggested. One method, for example, is to assess teachers according to the actual achievement gains made by students during the period in question relative to the initial achievement levels or prior rates of gain of these same students (Barro, 1985). The initial achievement level or prior rates of gain would serve as models for expected gains by the same students. Therefore, comparing actual gain against expected gain measures the amount by which a teacher exceeds or falls short of expected performance. However, the problems of adjusting for prior standing are extremely serious and rarely recognized (Iwanicki, 1986).

References

Barro, S.M. (1985). The logic of teacher incentives. Washington, DC: The National Association of State Boards of Education.

Haertel, E. (1986). The valid use of student performance measures for teacher evaluation. Educational Evaluation and Policy Analysis, 8 (1), 45-60.

Iwanicki, E.F. (1986). Continuities and discontinuities among measures of teacher performance. Paper presented at the annual meeting of the American Educational Research Association, San Francisco.

Meesick, W.M. (1984). The psychology of educational measurement. Journal of Educational Measurement, 21, 215-237.

These assessment issues papers, edited by Patricia Cloud Duttweiler, are based on the SEDL document, Perspectives on Performance-Based Incentive Plans. For information on purchasing that document, please contact SEDL's Publication Office.

A Product of
Southwest Educational Development Laboratory
211 East Seventh Street
Austin, Texas 78701 (512) 476-6861

This publication is based on work sponsored wholly, or in part, by the Office of Educational Research and Improvement, U.S. Department of Education, under Contract Number 400-86-0008. The content of this publication does not necessarily reflect the views of OERI, the Department, or any other agency of the U.S. Government.

SCDL

Issues in Perspective

Assessment Issues Paper # 23

November 1986

THE ISSUE: SOME CONCERNS ABOUT THE EFFECTS OF PERFORMANCE-BASED EVALUATIONS ON MINORITY TEACHERS

Maria L. Ramos-Cancel

Minority representation in the teaching profession, already low, may become even lower as states increase their reliance on standardized tests to screen candidates for teacher-training programs. Minority candidates fail these tests at a disproportionately high rate (Gifford, 1985). This problem can become further complicated if performance-based evaluations of currently employed teachers produce similar results -- that is, if minority teachers also fail these evaluations at a disproportionately high rate.

Impact on Minority Students

The large and increasing number of minority students in the public schools (Gifford, 1985) and the small and declining number of prospective minority teachers, makes even low minority failing rates on performance-based teacher evaluations a very serious problem. As the number of minority students continues to increase, so does the need for specialized programs and competent minority teachers to help these students reach their potential and to provide access to exemplars of success, especially for those students who need to see successful role models.

Minority Evaluation Results

Educators have expressed concern about minority groups getting lower performance ratings than the majority population in performance-based evaluations that rely on observational rating procedures (Wicker & Doss, 1984). Where either observational rating procedures or paper-and-pencil tests are used, the fairness and biased of the evaluation system has been questioned.

Factors to Consider

Several factors need to be considered when trying to determine whether an evaluation system is unbiased and fair with respect to minority teachers. Some of the factors that may account for evaluation results that are linked to ethnicity, race, or gender are related to the evaluation system; others are related to the evaluatees themselves. The following are examples of some of these factors:

Observational rating procedure. It has been noted that in classroom observation situations there is a general tendency for raters to make an overall judgment about the person being rated and to record favorable or unfavorable ratings that are consistent with this overall judgment (Saal, Downey, & Lahey, 1980). The biases that observers bring to the classroom can heavily influence this overall judgment. Although, in general, rater biases can be avoided by selecting and training evaluators appropriately, rater biases related to teacher characteristics (e.g., race, ethnicity, and gender) present problems because they are difficult to weed out (Barro, 1985).

OERI

Sponsored by the Office of Educational Research and Improvement US Department of Education

Evaluatee's background and experience. Another reason why minority teachers may get low ratings in performance-based evaluations is that they may, in fact, be unable to meet new teaching standards being established by school systems. Some minority teachers may have been hired on the basis of minimum teaching standards through affirmative action plans that gave emphasis to minority hiring quotas. Now, these teachers are expected to compete for rewards based on above-average teaching standards. The use of such evaluation systems to determine eligibility for incentives may seem unfair to those minority teachers who were hired by school systems in an effort to increase the number of minority teachers.

Recommended Strategies

Designing an unbiased and fair evaluation system for both minority and nonminority teachers is an important element in defending the legality of the evaluations and in creating a positive evaluation climate. The following strategies should be considered in trying to develop a fair and unbiased evaluation system:

- Ensure that highly qualified minority teachers participate in designing and implementing the evaluation system and in establishing the performance criteria or standards.
- Select both minority and nonminority evaluators carefully and provide them with intensive training.
- Develop and vigorously monitor teaching improvement programs designed to help minority teachers acquire those effective-teaching behaviors that are highly regarded in U.S. school systems, while allowing them at the same time, to maintain those behaviors that have proven to be helpful in teaching minority students.
- Recognize the value of those teaching behaviors that may be helpful in teaching minority students.

References

Barro, S.M. (1985). The logic of teacher incentives. Washington, D.C.: The National Association of State Boards of Education.

Gifford, B. (1985). Minority representation in the teaching profession: An affirmative program. American Education, 9 (2), 17-19, 43.

Saal, F.E., Downey, R.G., & Lahey, M.A. (1980). Rating the ratings: Assessing the psychometric quality of ratings. Psychological Bulletin, 88, 413-428.

Wicker, M.R., & Doss, D. (1984, August). Personnel evaluation systems in AISD: 1983-1984 final summary report (Pub. No. 83.72). Austin, Texas: Office of Research and Evaluation, Austin Independent School District.

These assessment issues papers, edited by Patricia Cloud Duttweiler, are based on the SEDL document, Perspectives on Performance-Based Incentive Plans. For information on purchasing that document, please contact SEDL's Publication Office.

A Product of
Southwest Educational Development Laboratory
211 East Seventh Street
Austin, Texas 78701 (512) 476-6861

This publication is based on work sponsored wholly, or in part, by the Office of Educational Research and Improvement, U.S. Department of Education, under Contract Number 400-86-0008. The content of this publication does not necessarily reflect the views of OERI, the Department, or any other agency of the U.S. Government.