

DONALD MUSELLA

Ontario Institute for Studies in Education

Improving Teaching: An Alternative To Supervisory Evaluation

Improved teaching, one of the ultimate objectives of much educational activity, continues to pose problems for theorists and researchers, as well as for practitioners. Although the practitioners need to look to researchers and theorists for guidance in the decision-making activity necessary to the operation of the institution, no cause-effect relationships have come forth to assist the evaluator of the teaching process. Decision-making affecting personnel in the school setting must continue to rely on experience-based intuitive prediction and assessments that constitute the best judgments at the time.

What is proposed here is (1) an examination of the problems inherent in teaching effectiveness research, and (2) a proposal for improving teaching that abrogates the reliance on empirical research, that minimizes the threat intrinsic to the superordinate-subordinate role relationship, and that enhances the chances for changed teacher behavior in the direction of teaching improvement.

Obviously, all research and experience has not gone for naught. We know that certain factors can account for the improvement of the teaching-learning situation. We know what elements in the physical environment need to be considered — ventilation, lighting, special requirements, height of desks, and so forth. We know something of the physiological, psychological, and social needs of teachers and learners. We know something of the nature and measurement of physical and mental growth.

We know something of personality factors, social perception, role relationships, and aspirational levels. We know of critical thinking, divergent thinking, convergent thinking, creative thinking, productive thinking, inductive thinking, and deductive thinking. We know of perceptual, cognitive, and affective processes and of certain relationships among these.

We can measure the learner; we can measure the teacher; we can measure relationships and interactions between these two — *but* we cannot say which teaching behaviors cause which changes in student behaviors, or what teaching leads to what learnings.

Hence, the problem — how can we evaluate and improve teaching if we cannot identify those teacher behaviors which lead to identifiable learning outcomes? Before attempting to provide the “answers,” a compendious review of the directions research has taken during this past half-century might highlight some of the problems to be overcome.

TEACHER-PUPIL: CAUSE AND EFFECT

Many studies continue to be directed toward identifying variables related to learning outcomes. There are those who believe that teacher effectiveness can only be measured in terms of learning outcomes. They contend that there is no teaching without learning. Bloom represents the majority of researchers when he states:

The writer takes the position that unless the criteria of effectiveness are related to changes in students, the researcher has avoided the primary criterion and has used only proximate criteria. . . .

Teaching and learning experiences are not good or poor in their own right. They are good or bad because of the ways in which they affect the learner.¹

In summarizing a chapter on what is learning, Clayton states:

Learning depends upon what the learner does. This involves how he perceives, how he thinks, how he feels, and how he acts. *There can be no learning unless he responds in some way.*²

Consequently, if one accepts this position, learning must be measurable before one can say that teaching exists. The dilemma: the only approach to improving teaching is through improving learning, and the only learning of use in measuring teacher effectiveness is that which is measurable. Using this frame of reference to look at teaching, the evaluation of teaching and teacher is, in fact, an evaluation of the product — the student.

However, this model of viewing teacher effectiveness has produced a host of unanswerable questions. How can specific learning outcomes be related to specific teaching acts? If pupil gain is the effect, one might be able to say that the effects indicate the presence of identifiable causes, but can the influence of myriad variables operating in human interaction ever be similarly identifiable?

What teacher behavior(s) can be the cause, and indeed, if one or more behaviors are found to be partial determinants, can other determinants and interaction among determinants be accounted for?

If so, can these be the criteria against which we measure and judge *all* aspects of teaching, *all* teachers, only teachers of x, teachers of xy, and/or just teacher b of pupil c in classroom d of subject e, during f-g clock hours, on h of i year?

¹B. S. Bloom, “Testing Cognitive Ability and Achievement,” *Handbook of Research on Teaching*, ed. N. Y. Gage (Chicago: Rand McNally and Co., 1963), p. 379.

²T. E. Clayton, *Teaching and Learning: A Psychological Perspective* (Englewood Cliffs, N.J.: Prentice-Hall, Inc., 1965), p. 45. Italics supplied.

Much of the efforts of researchers has been directed toward measurement of what Broudy *et al.* describe as the replicative uses of knowledge.³ Measurement of this aspect of schooling demands a restating of original learning upon presentation of appropriate cues. Two reasons for this approach have been the lack of adequate measures of certain behaviors and the limitation of existing measuring devices. However, perhaps an even more fundamental reason is the belief that all learning is measurable or, more appropriately for some, that which is measurable is learning.

In summary, there is considerable lack of confidence in the ability to measure teaching effectiveness by use of student-growth criteria. Although we have some measures of teaching behaviors, and some measures of student-growth (and even these are questionable), the cause-effect link poses a problem that seems insurmountable at this time.

THE TEACHER

Another approach used to considerable extent in teacher effectiveness studies has been to define the characteristics, attitudes, traits, or qualities that are associated with judgments of effective or ineffective teachers. Looking for the good teacher is something like looking for the good life. What is good becomes a product of the anticipated expectation to be derived as a result of the goodness. Thus, if we are to judge teachers with x,y,z, characteristics, we must be able to say that teachers with these characteristics cause certain desirable or undesirable behaviors in others.

The result, invariably, is an endless list of traits and their corresponding opposites in black-white fashion.

An example of the measurement and use of qualities factorally identified with teachers is the study undertaken by Ryans.⁴ Effective teachers have been described as fair, democratic, responsive, understanding, kindly, stimulating, original, alert, attractive, responsible, steady, poised, confident. Conversely, ineffective teachers may be described as partial, autocratic, aloof, restricted, harsh, dull, stereotyped, apathetic, unimpressive, evading, erratic, excitable, uncertain. We undoubtedly could add other adjectives to this list — honest, moral, intelligent, and so on.

If we are to look to research for definitions of the most effective teacher and the least effective teacher (as well as various degrees of both) certain questions need asking. How do we decide on the desired characteristics, on the hierarchical weighted arrangement of these characteristics, the extent to which each characteristics (or group of char-

³H. S. Broudy, B. O. Smith, and J. R. Burnett, *Democracy and Excellence in American Secondary Education* (Chicago: Rand and McNally and Co., 1964).

⁴D. Ryans, *Characteristics of Teachers: Their Descriptions, Comparison, and Appraisal* (Washington: American Council on Education, 1960).

acteristics) becomes manifest or becomes the manifestor of certain teaching behaviors, and, most important, which ones are determinants of specific learning outcomes? Ideally, and ultimately, can the description of the good teacher be anything less than god-like, and can these god-like characteristics be the determinants of the teaching behavior(s) that produce the desirable learning outcome(s)?

Once again an examination of one variable, in this case the teacher, is not enough. Although there is considerable evidence that particular personality characteristics of teachers have a discernible influence on pupil behavior, there is little evidence that certain personality characteristics are more desirable than others for teaching in general.⁵ We cannot look to research at this time for assistance in decisions about teacher evaluation from information of their personality, attitudes, values, and other personal characteristics.

Getzels and Jackson represent the concensus of researchers when they say:

Despite the critical importance of the problem and a half-century of prodigious research effort, very little is known about the nature and measurement of teacher personality, or about the relation between teacher personality and teaching effectiveness.⁶

TEACHING PROCESS

Recently, there has been a renewed emphasis on the teaching process and teacher-pupil interaction. Some researchers have been successful in developing systems for categorizing teacher and student verbal behaviors (verbal is used here in its broadest sense). A variety of systems designed to analyze pupil-teacher interaction in the classroom have been developed by Amidon and Flanders, Anderson, Aschner, Hughes, Medley and Mitzel, Smith, Whithall, among others.⁷ For the most part, the

⁵A. S. Barr, et al., *The Measurement and Prediction of Teacher Effectiveness* (Madison, Wisconsin: Dembar Educational Research Services, Inc., 1961); N. L. Gage, Ed., *Handbook of Research of Teaching* (Chicago: Rand McNally and Co., 1963); H. F. Siberman, Ed., "A Symposium on Current Reseach of Classroom Behavior of Teachers and Its Implications for Teacher Education," *The Journal of Teacher Education*, XIV, No. 3, (1963).

⁶J. W. Getzels and P. W. Jackson, "Teachers Personality and Characteristics," *Handbook of Research of Teaching* (Ed. N. L. Gage, Chicago: Rand McNally and Co., 1963), p. 574.

⁷E. J. Amidon and N. A. Flanders, *The Role of the Teacher in the Classroom* (Minneapolis: Paul Amidon and Associates, 1963); H. H. Anderson, "The Measurement of Domination and of Socially Integrate Behaviors in Teachers' Contacts with children," *Child Development*, 10 (1939), pp. 73-79; M. J. Aschner, "The Analysis of Classroom Discourse: A method and its uses (Unpublished Doctoral Dissertation, University of Illinois, 1959); M. Hughes and others, *Development of the Means for Assessment of the Quality of Teaching in Elementary Schools* (Salt Lake City: University of Utah, 1959); D. M. Medley and H. E. Mitzel, "A Technique for Measuring Classroom Behavior," *Journal Ed. Psychology*, 49 (1958) pp. 86-92; B. O. Smith, "A Concept of Teaching," *Teachers College Record*, 61 (1960), pp. 229-241; J. Whithall, "The Development of a Technique for the Measurement of Social-Emotional Climate in Classrooms," *Journal of Experimental Education*, 17 (1949), pp. 347-361.

criteria for the selections of categories have followed from a sound frame-of-reference or theoretical model of behavior.

One of the primary concerns of these researchers is the determination of the relationships between the behavior of the teacher and such outcomes as pupil attitude and achievement. However, has this research offered reliable criteria for the practitioner — the evaluator of teaching personnel? Can categorization of verbal activity differentiate effective teaching? Can the teacher who displays x verbal discourse be said to be more effective than the teacher who displays y verbal discourse? Can x verbal discourse be more effective than y verbal discourse? Can xy discourse be more effective than xy discourse in another situation with other people and at other times?

Most studies of this type have attempted to isolate behaviors regarded as more or less effective in stimulating pupil growth. Because the description and measurement of the teaching process is designated in terms of conditions, climates, attitudes, attitude change, and situations involving social interaction and communication between student and teacher, the approach depends almost entirely on direct observations of the behavior in the classroom situation. Evidence based on test scores or other indirect behavior cannot be adequate measure of the teaching process.

Therefore, research in this area has been hindered by the limitations of direct classroom observation. Some of these limitations are as follows: (1) insufficient observation time upon which to base judgment; (2) inadequacy of recorded observations as valid and reliable samples of the total teaching experiences of the teacher; (3) uncertainty of the validity and reliability of the observers — assessors, evaluators, judges, raters. This last point needs to be considered further.

RATER

How certain are we of the judgment of the rater? Are “outside experts” the best judges of teacher effectiveness? Studies which have used raters who are not a part of the daily milieu have admitted the inadequacy of observation time and the lack of reliability in judgment. And what of validity — was their observation an adequate sample of that teacher’s effectiveness?

Research of the relationships and differences between kinds of raters and teacher rating has not yielded results which can assist one in determining who should do the rating — superordinates (supervisors, administrators, department chairmen), subordinates (students), peers (teachers), or “experts” (visiting personnel not identified as part of the immediate school setting).

The work done in this area, although quite extensive, has not produced any significant consistent evidence, other than the lack of corres-

pondence between ratings by supervisors, colleagues, students, and teacher-training specialists.⁸

Perhaps the subjectivity of rating prevents consistency in rating results; that is, the problem is not so much the lack of consensus on criteria, but the influence of variables related to the personal characteristics of the raters.

Remmers, in reviewing the use of rating methods in research on teaching, states:

In addition to any limitations imposed by the rating form itself, ratings are limited by the characteristics of the human rater — his inevitably selective perception, memory, forgetting, lack of sensitivity to what may be psychologically and socially important, his inaccuracies of observation . . .⁹

Ryans' work (*Characteristics of Teachers*) in training his observers exemplifies the principle that the measuring instrument in the rating method is not so much the paper device itself as it is the rater, judge, or observer.¹⁰

In one study undertaken by this writer, it was found that certain personal characteristics of the rater and ratee are related to the rating of teacher effectiveness by administrators in the school setting.¹¹ Support was obtained for rating, defined as including all the physiological and psychological processes that go into the final outcome, as a function, in part, of the perceptual-cognitive style of the individual rater.

One weakness evident in the use of formal rating forms (criteria) as part of a research design is that regardless of the formal criteria accepted and used, one must rely on the rater's perceptual-cognitive view of the ratee, of the criteria, and of the relationship (similarity-difference) between the two. At this time this seems to be an uncontrollable variable — how sure are we that the judgment and decision of the rater is based indeed on the stated criteria.

What can the practitioner do? Although research has failed to provide useful criteria for the measurement and evaluation of teacher effectiveness, the rating of teachers continues to be one of the professional responsibilities of the administrator and supervisor.

Superordinates, be they in industry, government, or education, must make decisions regarding personnel. They must make decisions regarding selection, placement, transfer, re-assignment, in-service training

⁸G. C. Stern, "Measuring Noncognitive Variables in Research on Teaching," *Handbook of Research on Teaching* (ed. N. L. Gage, Chicago: Rand McNally and Co., 1963), pp. 398-447.

⁹H. H. Remmers, "Rating Methods in Research on Teaching," *Handbook of Research on Teaching* (ed. N. L. Gage, Chicago: Rand McNally and Co., 1963), p. 329.

¹⁰*Ibid.*, p. 332.

¹¹D. F. Musella, "Open-and-Closed Mindedness as Related to the Rating of Teachers By Administrators: Implications for Administrative Theory Based on Superordinate-Subordinate Role Relationships," (Unpublished Doctoral Dissertation, State University of New York at Albany, 1965); D. F. Musella, "Open-and-Closed-Mindedness as Related to the Rating of Teachers by Elementary School Principals," *Journal of Experimental Education*, 35 (1967).

promotion, dismissal, and retirement, as well as those that are more directly related to evaluation for the purposes of improvement of process.¹²

ONE POSSIBLE ANSWER

Since the ultimate objective should be the improvement of process, that is, the improvement of teaching, and until we have conclusive evidence on certain cause-effect variables, it might be more productive and desirable to direct our efforts toward enhancing and extending opportunity for self-improvement. Certainly we cannot be far off when we state what seems like a reiteration of the obvious — the attempts at improvement that have the greatest chances for success are those which come from “self” — the person whose behavior is to be changed.

From this point of view, the superordinate-rater must accept certain basic concepts and must assume certain responsibilities: (1) although the superordinate — be he principal, supervisor, department chairman or master teacher — must assume the leadership role, the development of teacher effectiveness criteria must be a cooperative effort of both rater and ratee; (2) the superordinate-rater must provide the means for describing and categorizing the teaching act in terms which the teacher and rater can understand and accept — in other words, terms which do not connote values of effective or ineffective teaching; and (3) the superordinate-rater must provide the teacher with the opportunity for self-assessment based, in part, on the criteria previously decided upon.

DEVELOPMENT OF CRITERIA

Obviously, the above statements need considerable explanation. First of all, the development of ground rules — criteria — is the necessary prerequisite to any assessment. If an effective teacher is one who has control of the classroom, and if control is defined as the determinant of behavior, then this must be understood at the outset. If an effective teacher is one who displays flexibility in the classroom, and if flexibility is defined as shifts in presentation and climate as demanded by shifts in student response, then this must be defined, stated, and understood previous to the assessment. If an effective teacher is one who is permissive and if permissiveness is defined as accepting student initiative and leadership, then this must be understood as such by both parties. The question at this time is not the “universal” acceptability of effective teaching definition and description, and, to considerable extent, the ques-

¹²P. M. Blau, *The Dynamics of Bureaucracy* (Chicago: The University of Chicago Press, Revised Edition, 1955); N. J. Powell, *Personnel Administration in Government* (Englewood Cliffs, N.J.: Prentice-Hall, Inc., 1956); J. W. Getzels, “A Psycho-Sociological Framework for the Study of Educational Administration,” *Harvard Educational Review*, 22 (1952), pp. 235-246; T. L. Whistler and S. F. Harper, (Eds.) *Performance Appraisal* (New York: Holt, Rinehart and Winston, Inc., 1962).

tion at this time is not the acceptability by both parties in the rating situation; rather, the objective is the awareness and understanding of the stated criteria. Consensus, acceptance, and compliance are individual matters that can be considered at the time of decision-making by the superordinate. The responsibility lies in the development and communication of the criteria for teacher effectiveness for the specific situation in which both rater and ratee are active participants.

PERCEPTION OF SELF

The second point mentioned above involves the techniques for viewing "self." The usual procedure consists of rater observance of the classroom followed by a discussion of selected aspects of the teaching-learning situation. The primary purpose of the follow-up discussion is to change future teacher behavior in the direction of "improvement" as stated in the assessment. The rater is telling and the ratee is listening and verbally agreeing or disagreeing, depending on the latter's perception of the expected, acceptable role behavior.

Although one cannot predict the changes in teacher behavior caused by the rater-ratee discussion, the rater has done his job — has assumed his responsibility. The acceptance depends to considerable extent, on the relationship between the rater and the ratee, as well as on the contents of the comments.

A note should be inserted at this point. There are those who say that rating often is not going on in the situation described. They say that generally, the discussion focuses on specific aspects of teaching and consists of positive comments directed toward the teacher's presentation at the time. However, can there *ever* be interaction without assessment — judging? Certainly we do not always indicate our judgments according to formal rating scales, and we are not always conscious of rating others. Nevertheless, we *are* judging others (and things) at all times — this is part of thinking. What augments the threat and minimizes the freedom for cooperative effort in the superordinate-subordinate setting is that the rating factor (conscious or unconscious) is always present — it is an inherent part of this role relationship. No amount of rationalization can eliminate or subrogate the element of judgment that is part of social perception; certainly the ratee is aware of this. This accounts for the extensive role playing in "live" human interaction.

Consider now another procedure, one which may be in lieu of or in addition to the one presented above. Researchers have provided us with certain tools or techniques for organizing feedback data: The categorizing and coding of verbal and non-verbal discourse as developed by Flanders, Aschner, and others.¹³ These are but several examples of

¹³Mary J. Aschner and J. J. Gallagher, et al., "A System for Classifying Thought Processes in the Context of Classroom Verbal Interaction" (Institute for Research on Exceptional Children, University of Illinois, 1962); Arno A. Bellack,

methods of providing a "mirror" of teacher-pupil behaviors in the teaching-learning situation.¹⁴ Although the question of acceptance of the basic assumptions and resulting interpretations is argumentative, the use of coding techniques for abstracting and displaying teacher-pupil interaction can provide the teacher and rater with certain common dimensions for viewing behavior.¹⁵

Using Flanders' categorization of verbal teacher-pupil interaction as an example, the following plan is presented as one method of approach.

What if we instruct the teacher (and rater) in the use of this technique of analysis, provide the equipment and materials for recording the classroom discourse, provide the time for an analysis of classroom interaction, provide the criteria (through cooperative effort) against which the teacher (and rater) can compare the teaching acts, and *leave the assessing to the teacher*.

Obviously the last part of the statement would cause certain superordinates to look askance at such a proposal — and rightly so, perhaps. Unfortunately, the improvement of teaching cannot be left to the teacher alone. Institutional requirements place the responsibility for this improvement directly on the principal, supervisor, department chairman, and/or other designated superordinates. In order to guarantee the success of this process, the institution usually requires formalized assessments made periodically. Nevertheless, this assessment need not detract from the efforts to stimulate self-improvement. The plan focusing on self-improvement should not become a part of (nor can we expect it to replace) the requirements of the institution for the assessment of teaching. To the extent that superordinate rating is used for purposes of decisions regarding status, salary, retentions, and so on — , to that extent does it constitute a threat to the improvement of teaching.

Thus we have two entirely different purposes, hence, kinds of assessment: (1) one which satisfies the institutional needs for perpetuation and purpose achievement; and (2) one which satisfies the needs of improvement of process for both the individual and the institution.

et al., *The Language of the Classroom: Meanings Communicated in High School Teaching, Part II* (New York Institute of Psychological Research, Teachers College, Columbia University, 1965); F. G. Cornell, C. M. Lindvall and J. L. Saupe, *An Exploratory Measurement of Individualities of School and Classrooms* (Urbana, Ill.: College of Education, Bureau of Educational Research, 1952); N. A. Flanders, *Teacher Influence, Pupil Attitudes, and Achievement* (Washington: Government Printing Office, 1965); Donald M. Medley, "Experiences with the Oscar Technique," *The Journal of Teacher Education*, 14 (1963), pp. 267-298; B. O. Smith, "A Conceptual Analysis of Instructional Behavior," *The Journal of Teacher Education*, XIV, No. 3 (1963), pp. 294-298.

¹⁴In addition to methods which abstract actual behaviors, researchers have proven the value of video systems and instantaneous response systems in providing feedback of the classroom situation.

¹⁵This paper will not discuss the characteristics, uses, or pros and cons of specific methods.

Through the use of this procedure, we might predict the following results to take place: (1) the teacher and rater become the developers of the criteria (before the fact) against which certain behaviors are to be assessed; (2) the teacher and rater focus on actual classroom events and specific behaviors, not generalities; (3) there is a reduction in the possibilities for disagreement due to the lack of communication; (4) the teacher and rater have a common frame-of-reference for viewing and judging teaching in terms that are relatively free of connotative value dimensions;¹⁶ and (5) the situation, as well as the relationship between rater and ratee, remains relatively free of threat since reference to effective-ineffective teacher behaviors are obtained through one's own perception and not simply from direction and/or implications presented by the rater.

SUMMARY

Research has provided the practitioner, whose responsibility it is to improve teaching, with little evidence to assist him in identifying the cause-effect relationships in the teaching-learning situation. Consequently, since he cannot refer to research to tell him what "effective" teaching should be, the decisions he makes are based on intuitive experience-based judgments.

This author has indicated that a re-focus of research and practice in this area be directed toward providing the opportunity and means for the teachers to examine their own teaching.

Techniques for providing feedback of the teaching situation have offered us heretofore unavailable opportunity to redirect our efforts in the improvement of teaching. Because of the availability of first-hand "objective" information of one's teaching, and because of the concomitant reduction of threat inherent in the superordinate-subordinate role relationship, the re-focus of efforts toward the self-improvement of teaching seems to be the most profitable direction to go at this time.

¹⁶If the teacher and rater are presented with information indicating the percentage of teacher-asked questions which are answered by the teacher during a specified amount of class time, this fact alone is a non argumentative one. However, if the rater places value on the fact in terms of effectiveness or ineffectiveness, then the advantage of feedback, as well as the primary purpose of this procedure, will be lost.