

Pupil Achievement as an Index of Teacher Performance

WILLIAM H. LUCIO

RESearch on predictive measures of teacher effectiveness has been pursued for many decades, yet the numerous efforts to identify the characteristics of an effective teacher or to describe and measure effective teaching performance generally have led to inconclusive results.

One major difficulty in predicting teacher effectiveness has been the lack of a solution to the criterion problem (Ryans, 1960). Traditionally, two general types of criteria have been employed: (a) *process*—teacher effectiveness assessed against some standard of performance or particular overt teaching acts generally assumed or inferred to relate to effective teaching: the assumption being that if a teacher *performs* certain specified acts, pupil behavior—and teacher effectiveness—can be predicted; and (b) *teacher characteristics*—various personal characteristics such as intelligence, years and quality of schooling, personality traits, aptitudes, and other personal attributes of the teacher assumed to relate to or predict teaching performance: the assumption being that if a teacher possesses particular personal traits, learner growth—and teacher effectiveness—can be predicted (Lucio and McNeil, 1969).

In general during the period between 1900 and 1960, the focus of investigators was on various aspects of teacher behavior,

and the study of the outcomes of instruction (pupil achievement), as a correlate of teacher effectiveness, was neglected. Investigators found that administrators and supervisors most often judged teacher competence on the basis of: (a) teaching ability, (b) disciplinary ability, (c) scholarship, and (d) personality, and demonstrated in their studies that commonly employed teacher rating instruments and raters' assessments were unreliable (Eysenck, 1953; Morsh and Wilder, 1954; Howsam, 1960; Ryans, 1960; Barr *et al.*, 1961; Kleinman, 1966; Popham, 1971b).

Since about 1960, following the lead of industrial organizations which have made use of performance objectives as the basis for judging personnel effectiveness (Flanagan and Burns, 1955; Adams, 1959; Meyer *et al.*, 1965), various educational researchers, in attempting to seek solutions to the criterion problem, have shifted from studying primarily what the teacher does (means of instruction) to examining changes in learner behavior as a result of instruction (outcomes of instruction). Change in pupil behavior, thereby, becomes the criterion for evaluating teacher performance (Morsh, Burgess, and Smith, 1958; McNeil, 1966; Hastings, 1969; Popham, 1971a).

In answer to arguments against the merit of using performance criterion mea-

tures in teacher evaluation (referring to studies before the 1960's), Justiz (1968) designed a study to test the validity of nine identified criticisms. He interpreted his results as having invalidated most of the criticisms, and as presenting evidence (statistically not the result of chance) that, in general, teaching ability of teachers could be measured reliably in terms of pupil achievement. He also found a relationship between student teacher attitude (as measured by the Minnesota Teacher Attitude Inventory) and pupil achievement.

An experimental study, based on the concept of supervision by objectives and conducted by McNeil (1967) with 72 secondary school student teachers, revealed that a larger number of student teachers supervised by objectives were perceived by their supervisors as achieving greater success as measured by pupil performance (significant at the .01 level) than those supervised by a rating scheme. The student teachers who contracted with their supervisors to achieve explicit changes in pupil performance were perceived also as being more successful in applying the principles of learning.

A second investigation conducted by McNeil (1967), using 44 elementary student teachers as subjects, revealed that pupils of student teachers whose success was measured by their pupils' attainment of agreed-upon objectives achieved significantly better in punctuation skills than pupils whose teachers were assessed not by pupil achievement criteria but by the usual rating scale. These same student teachers completed a questionnaire to measure their perceptions of the supervisor and the methods of evaluation, and were almost unanimous (98 percent) in their preference for the pupil progress criterion as the measure for evaluation of their teaching. Both groups tended to report the same amount of preparation time given to the punctuation lessons, felt free to select their own teaching procedures, and found their supervisors' suggestions helpful.

Moffett (1966) conducted a study in which he compared the performance of student teachers evaluated on the basis of attaining agreed-upon instructional objec-

tives with that of teachers evaluated by means of a rating instrument. Data were obtained on the extent of pupils' attainment of instructional objectives in geography skills and their attitudes toward subject matter, and on teachers' attitudes toward supervisory help, satisfaction with midterm grades, and preferences for types of performance rating. Secondary school pupils in grades 7 through 12, selected by a randomized sampling technique, served as subjects. After pretesting pupils on geography skills, 36 student teachers were randomly assigned to either an experimental or control group. Teachers in the experimental group executed pre-instructional contracts with their supervisors based on the instructional objectives to be achieved, as revealed by the pretests, and their teaching performance was evaluated in terms of pupil achievement. Control group teachers, while informed about pupils' pretest deficiencies and the need to correct them, did not enter into an instructional contract; their performance was evaluated by means of rating scale measures.

Among the findings reported by Moffett were the following: (a) the pupils in the experimental group performed significantly better on a post-test of geography skills than did pupils of teachers in the control group; (b) teachers in the experimental group expressed more confidence in supervisory help and reported satisfaction with their midterm grades significantly more often than did teachers in the control group; and (c) 94 percent of all teachers, regardless of whether they were in the experimental or control group, expressed a preference for having their teaching performance evaluated on pupil achievement as a result of instruction rather than on rating scale measures.

Smithman (1970), in a study employing the strategy of supervision by objectives, attempted: (a) to determine the extent to which agreement on instructional objectives stated in behavioral terms by supervisor and teacher prior to teaching would increase pupil performance, and (b) to discover whether the process of supervision by objectives resulted in evaluations of teaching more germane to instructional performance.

GETTING IN TOUCH

-the drug education program that works from the inside out.

GETTING IN TOUCH, a new pre-K-6 attitudinal drug program, turns kids on—to life and to themselves. Authors Barbara Milbauer and Karren Kartell Jacobson have developed a student-involving program that is simple, non-moralizing, and fun! It is comprised of four 32-page non-graded sequential activity books illustrated with large photographs and humorous art and a teacher's manual. **GETTING IN TOUCH** is so flexible that it can fit into any elementary school classroom, for use on a one-to-one basis, with groups, or with the entire class. No additional materials and no extra teacher preparation are needed before **GETTING IN TOUCH** goes to work for you.



EDUCATIONAL SERVICES

A Division of Franklin Watts, Inc.
Dept. EL, 730 Fifth Avenue
New York, N.Y. 10019

Please send me _____ copies of "Feeling Alive"
_____ "Being Me" _____ "Life—Read All About
It!" _____ "Come To Your Senses" @ 95¢ each.
(Minimum order of 30 assorted books.) And _____
copies of Teacher's Manual @ 75¢ each. I enclose
_____ check _____ money order for \$ _____ total.

Name _____

School _____

Address _____

City _____ State _____ Zip _____

The study was predicated on two assumptions suggested by Lucio and McNeil (1969): (a) learning is evidenced by a change in behavior; and (b) teaching is successful only when the instructor's predetermined and intentional changes sought in the learner actually occur.

Twenty classroom units in nine Canadian schools, consisting of 20 certified teachers and 558 pupils in their sixth year of school, were selected as the sample using a stratified randomization technique. The 10 teachers assigned to the experimental group together with their principals selected the instructional objectives and agreed upon the criteria which were to be accepted as evidence that learners had reached the desired behavior; and they were evaluated accordingly. A control group of 10 teachers, employing the same instructional objectives as the teachers in the experimental group, but without executing an instructional agreement with their principal, were evaluated by means of the school district rating scale. Pupils in both the experimental and control groups were administered pretests and post-tests of mathematical skills.

Based on results on the post-test in which a mean difference of 10.3 favored the experimental group (significant at the .05 level), it was concluded that pupils whose teachers were evaluated by objectives outperformed those pupils whose teachers were evaluated on a rating scale. On five measures of teacher attitude toward evaluation, as measured by a post-test attitude questionnaire, no significant differences were found between the experimental and the control group. As to partiality for the subject matter and the amount learned, pupil responses to the questionnaire revealed that no differences existed between the experimental and the control group. The amount of individual instruction provided the two groups was approximately equal, and it was concluded that pupils of the teachers teaching by objectives experienced no undesirable side effect, such as, for example, a dislike of subject matter.

In an attempt to isolate a readily usable indicator to assess a teacher's instructional

skills by measuring pupils' attainment of instructional objectives, Popham (1971b) developed a *teaching performance test*. His approach to assessing instructional performance was "... predicated on the assumption that the main reason for a teacher's existence is to promote beneficial changes in learners." He constructed his tests in the fields of social science, electronics, and auto mechanics, and used them to compare the performance of credentialed and experienced teachers with that of persons neither credentialed nor experienced. Each of the three teaching performance tests consisted of a set of specific instructional objectives measured by a post-test. The numbers of objectives and post-test items varied with the particular subject fields. All three tests contained a set of resource materials which could be used by the teacher in planning the instructional sequence to accomplish the objectives.

In validating this method of assessing teacher effectiveness, a construct validation strategy was selected, the researcher noting that, "Considering the nature of the requirements of teaching performance tests, it seemed that these tests ought to be able *at least* to distinguish between grossly disparate groups such as credentialed, experienced teachers and those who were neither credentialed nor experienced." Thirteen high school social science teachers, 16 high school and junior college electronics teachers, and 28 high school and junior college auto mechanics teachers constituted the experienced group. Identical numbers of non-teachers included: (a) state college students who were social science majors or minors, but who had never taught or completed any course work in professional education, for the social science group; (b) garage mechanics, for the auto mechanics group; and (c) television repairmen and electronics industries workers, for the electronics group.

Performance tests were subjected to validation contrasts in public school situations, with each participating teacher having at least two sections of a class, one randomly assigned to the nonteacher, the other to the regular teacher. Having received the ob-

jectives for the performance test and the resource materials two weeks prior to beginning instruction, teachers then planned a short unit of instruction to accomplish as many objectives as possible in the allotted time, nine instructional hours for electronics and auto mechanics tests, and four hours for the social science test.

Test results revealed that the experienced teachers did not markedly outperform the non-experienced teachers on any of the three teaching performance tests. Nor did students' responses to the affective questionnaire reveal any significant differences between the experienced and non-experienced teachers. These unexpected results led Popham to state that there was no way of glossing over the possibility that, "Experienced teachers are not particularly skilled at bringing about specific behavior changes in learners." He concluded that perhaps insufficient attention had been given to training teachers in the special expertise necessary to achieve desired changes in a learner's behavior.

Overall, the findings of early investigators provided relatively little evidence to demonstrate that particular teaching acts or teacher characteristics, assumed to relate to teacher effectiveness, were associated consistently with learner achievement. More

Manuscripts Needed for "Research in Review"

Manuscripts for the "Research in Review" column are being solicited by the coordinator. The column is devoted to providing a *summary* of research completed in an area of interest. The column is *not* designed to report on one particular research study (such manuscripts are best sent to the "Research Supplement").

Manuscripts should be 1,400 words in length (approximately five typewritten double-spaced pages), including references. Submit manuscripts in duplicate and with return envelope and postage. Decisions on materials will be made as promptly as possible.

Materials for "Research in Review" should be addressed to: Dr. Marcella L. Kysilka, Associate Professor, College of Education, Florida Technological University, P.O. Box 25000, Orlando, Florida 32816.

recently, investigators have shown that by specifying desired changes in learners, arranging instructional events to produce the changes, and assessing the learners' attainment of instructional objectives, selective indices of teacher performance (based on pupil achievement) can be derived (Cohen and Brawer, 1969; McNeil, 1967, 1971; Popham, 1971a).

Systematic efforts in the direction of analyzing teacher performance as a correlate of predicted changes in learners might be expected: (a) to establish appropriate cri-

teria for assessing teacher performance, (b) to improve teachers' skills in defining and achieving instructional objectives, (c) to provide more explicit evidence of pupil learning, (d) to define better the degrees of accountability for school personnel in accomplishing the goals of schooling, and (e) to provide evidence for the public that schools are achieving stated objectives.

—WILLIAM H. LUCIO, Professor, Graduate School of Education, University of California, Los Angeles.

References

- H. F. Adams. "Merit Rating for Personnel Development." *Journal of American Society of Training Directors* 13 (9): 4-10; September 1959.
- A. S. Barr et al. *Wisconsin Studies of the Measurement and Prediction of Teacher Effectiveness*. Madison, Wisconsin: Dembar Publications, Inc., 1961.
- Arthur M. Cohen and Florence B. Brawer. *Measuring Faculty Performance*. Washington, D.C.: American Association of Junior Colleges, 1969.
- H. J. Eysenck. *Uses and Abuses of Psychology*. Baltimore: Penguin Books, Inc., 1953.
- John C. Flanagan and Robert K. Burns. "The Employee Performance Record: A New Appraisal and Development Tool." *Harvard Business Review* 33 (5): 95-102; September-October 1955.
- J. Thomas Hastings. "Assessment of Learning Outcomes." In: William H. Lucio, editor. *The Supervisor: New Demands, New Dimensions*. Washington, D.C.: Association for Supervision and Curriculum Development, 1969. pp. 47-57.
- Robert B. Howsam. "Who's a Good Teacher?" *Joint Committee on Personnel Procedures*. Burlingame: California School Boards Association and the California Teachers Association, 1960.
- Thomas B. Justiz. "A Method for Identifying the Effective Teacher." Unpublished doctoral dissertation, University of California, Los Angeles, 1968.
- Gladys S. Kleinman. "Assessing Teacher Effectiveness: The State of the Art." *Science Education* 50 (3): 234-38; April 1966.
- William H. Lucio and John D. McNeil. *Supervision: A Synthesis of Thought and Action*. Second edition. New York: McGraw-Hill Book Company, 1969.
- John D. McNeil. "Antidote to a School Scandal." *Educational Forum* 31 (1): 69-77; November 1966.
- John D. McNeil. "Concomitants of Using Behavioral Objectives in the Assessment of Teacher Effectiveness." *Journal of Experimental Education* 36 (1): 69-74; Fall 1967.
- John D. McNeil. *Toward Accountable Teachers*. New York: Holt, Rinehart and Winston, Inc., 1971.
- Herbert H. Meyer et al. "Split Roles in Performance Appraisal." *Harvard Business Review* 43 (1): 123-29; January-February 1965.
- George McHatton Moffett. "Use of Instructional Objectives in the Supervision of Student Teachers." Unpublished doctoral dissertation, University of California, Los Angeles, 1966.
- Joseph Morsh and Eleanor Wilder. *Identifying the Effective Instructor: A Review of the Quantitative Studies, 1900-1952*. Research Bulletin AFPTRC-T-54-44. San Antonio, Texas: Lackland Air Force Base, 1954.
- J. E. Morsh, G. G. Burgess, and P. N. Smith. *Student Achievement as a Measure of Instructor Effectiveness*. Research Bulletin AFPTC-TN-55-12. San Antonio, Texas: Lackland Air Force Base, 1958.
- W. James Popham. "Performance Tests of Teaching Proficiency: Rationale, Development, and Validation." *American Educational Research Journal* 8 (1): 105-17; January 1971a.
- W. James Popham. "Teaching Skill Under Scrutiny." *Phi Delta Kappan* 52 (10): 599-602; June 1971b.
- D. G. Ryans. *Characteristics of Teachers: Their Description, Comparison, and Appraisal*. Washington, D.C.: American Council on Education, 1960.
- Harold Henry Smithman. "Student Achievement as a Measure of Teacher Performance." Unpublished doctoral dissertation, University of California, Los Angeles, 1970. □

Copyright © 1973 by the Association for Supervision and Curriculum Development. All rights reserved.