



$$6 \times \frac{2}{3} = \frac{6 \times 2}{3} = \frac{12}{3} = 4$$

Using Effectiveness Research in Teacher Evaluation

Shirley B. Stow

Student achievement rose sharply when the West Des Moines schools incorporated the findings of effectiveness research in their teacher evaluation system.

Don Medley (1979) claims that "... if teachers know the criteria on which decisions affecting their careers are based, they will meet the criteria if it is humanly possible to do so." Could teachers be taught to use direct instruction in connection with the appraisal process, and would that lead to increased student achievement? We in the West Des Moines (Iowa) Community School District decided to find out.

Our project, which was funded by a National Science Foundation Grant,¹ formed a partnership among the district's administrators, board of education, the teacher association, and university-based researchers. The idea was to create a teacher performance evaluation system that used criteria from existing research and a cycle of operating procedures that included the setting of job improvement targets. The system was to be validated by use of student gain scores on custom-tailored, criterion-referenced tests.

Validation was especially important because researchers have found it difficult to go beyond rater perceptions to assessment of teacher performance in terms of student behaviors (Shavelson and Dempsey-Atwood, 1976).

Iowa State University researchers have repeatedly identified clusters of categories that can be used to rate teacher performance including productive teaching techniques, positive interpersonal relations, organized-structured classroom management, intellectual stimulation, and desirable out-of-class behavior (Manatt, and others, 1976).

In his report of meaningful findings of research in teacher effectiveness, Medley (1977) indicates that effective teachers spent more time on task-related or

¹ This research was supported by the National Science Foundation, Grant No. GU3373. Richard P. Manatt, College of Education, Iowa State University, Ames, Iowa 50011.

"academic" activities and differed from less effective teachers in group work procedures, seat assignments, praise and positive motivation, and the use of a variety of management techniques. Medley defined teacher effectiveness as "... the effects a teacher has on pupils. The more pupils learn as a result of what a teacher does, the more effective that teacher is." To use that definition, we needed to assess student performance, measure how much was learned, and determine what segment of the learning could be associated with the efforts of a given teacher.

Mathematics was selected as the subject since it is considered to be school specific; that is, less influenced by out-of-classroom learning experiences than reading. All 494 fourth graders were chosen as students because gains in mathematics are usually substantial at that level. Intelligence and reading ability were controlled, but socioeconomic status was not. West Des Moines Community School District serves a relatively homogeneous, upper middle-class, white population.

The Process

Our first step was to design the teacher performance appraisal system. A committee of teachers, administrators, board members, and citizens of the community developed an instrument based on what was then considered "state of the art" in performance criteria (Rosenshine and Furst, 1971). Those who were to be observer/evaluators were trained in a series of one-day workshops.

Next we developed criterion-referenced tests. Teachers from grades three, four, and five worked with consultants from Westinghouse Learning Corporation to select and/or write performance objectives for beginning and ending fourth-grade mathematics. Thirty-one objectives were chosen for which Westinghouse provided 69 test items to be used in the pre- and post-tests. Pre-tests were given in September with extensive diagnostic data provided to each teacher.

Duplicating

Educators are authorized to reproduce a single article from this publication without making a written request provided that 1) duplication is for an educational purpose in a not-for-profit institution; 2) copies are made available without charge beyond the cost of reproduction; and 3) each copy includes full citation of the source. Permission to reproduce more than one article will be granted under the same conditions to those who make a reasonable request in writing.

This authorization does not apply to materials, if any, for which ASCD doesn't hold the copyright. Any such materials are so identified.

Each of the 17 fourth-grade mathematics teachers was evaluated separately by his or her building principal, the director of elementary education, and the assistant superintendent. For purposes of validation, research, the 17 categories on the evaluation instrument were assigned points. A performance index ranging from 0-180 points was used. In pre-observation and post-observation conferences, teachers were encouraged to adhere to the performance criteria.

When all the necessary data had been collected, we analyzed it. Because the district had promised teachers that their formal appraisal would be based not on student achievement but on observation, test results were treated as confidential information.

The Results

There was no reason for concern, however. We found significant gains in student scores on the criterion-referenced tests in all 17 classes, both in objectives attained and in the number of items answered correctly.

To guard against the possibility that the teachers under evaluation were placing undue emphasis on mathematics to the detriment of other subjects, we examined standardized test results for the past five years. Each year during that period, students in our district had gained an average of nine to ten months on the Iowa Tests of Basic Skills. That pattern continued during the year of our study in all subjects and grades except mathematics in fourth grade. There the gains were approximately 22 months.²

Despite the remarkable increase in achievement, however, the student gain scores did *not* correlate with the composite scores of teacher performance. Most item scores on the teacher evaluation instrument had positive but low, nonsignificant correlations. Some even correlated in a negative direction. In other words, we had not been able to validate the instrument. Nevertheless, it was evident that teachers whose pupils attained substantial growth could be characterized as businesslike, task-oriented, and concerned about academic learning time.

In Retrospect

The indecisive results regarding performance criteria were not unexpected. From the outset, we considered this a "post hole" attempt to see if criterion-referenced tests could be used in teacher performance research. The statistical analysis gave us an opportu-

² Tables that show the test results and/or the correlation of teacher performance rating categories with student gain scores are available upon request from the author: Shirley B. Stow, 1101-5th Street, West Des Moines, Iowa 50265.

To help you help them...



...that's our goal at the Social Science Education Consortium. And we've developed three new publications designed to help administrators and curriculum planners achieve their goals.

Data Book: Social Studies Materials and Resources, vol. 4. Analyses of 100 K-12 classroom materials, teacher resources, and curriculum guides published during 1977-78. Pub. no. 234. Paperbound; 250 pp.; \$10.00.



Planning a Social Studies Program: Activities, Guidelines, and Resources. Step-by-step procedures and all necessary materials for designing and evaluating a curriculum. Pub. no. 224. Ringbound; 194 pp.; \$14.50.

Social Studies Evaluation Sourcebook. Detailed analyses and descriptions of 291 instruments for evaluating all aspects of a K-12 social studies program. Pub. no. 227. Paperbound; 465 pp.; \$17.50.



Write to us for a free catalog.



Social Science Education Consortium, Inc.
855 Broadway
Boulder, Colorado 80302

ity to improve the prototypic instrument we had created.

Perhaps the most salient finding of the study was the discovery that performance appraisal centering on teacher effectiveness criteria coincided with extraordinary results in standardized test scores.

It seems that schools can use basic research findings to improve their effectiveness. The more teachers and administrators studied the research on teaching, the more willing they became to use direct instruction activities as at least part of the district's performance criteria. The steps that the mathematics teachers followed encouraged the academically-focused, teacher-directed classroom, which is presently called "direct instruction." The results for West Des Moines' fourth graders were so impressive, we believe similar efforts should be made elsewhere.

References

Berliner, David C. "Instructional Time in Research on Teaching." Paper presented to the American Educational Research Association, April 1977. Available through the Far West Laboratory for Educational Research and Development, San Francisco.

Hidlebaugh, Everett. "Naperville (Illinois) Teacher Performance Evaluation Items." *The District*, May 1973.

Manatt, Richard, Everett Hidlebaugh, and Kenneth Palmer. "Improved Teacher Performance Evaluation With Improved Rating Scales." *National Association of Secondary School Principals Bulletin* 68(401): 21-23; Sept.-Oct. 1979.

Medley, Don M. "A Summary of Teacher Effectiveness Research." Charlottesville, Virginia: Department of Research Methods, University of Virginia, 1977.

Medley, Don M. *Teacher Competence and Teacher Effectiveness: A Review of Process-Product Research*. Washington, D.C.: American Association of Colleges for Teacher Education, 1977.

Medley, Don M. "The Effectiveness of Teachers." In: Penelope Peterson and Herbert Walberg, editors. *Research on Teaching: Concepts, Findings, and Implications*. Berkeley, California: McCutchan, 1979. p. 25.

Millard, Joseph E. *Teacher Evaluation*. Ankeny, Iowa: Educational Services Division-Heartland Educational Agency (AEA 11), September 1976.

Rosenshine, Barak and Norma F. Furst. "Research on Teacher Performance Criteria." In: B. Othanel Smith, editor. *Research in Teacher Education: A Symposium*. Englewood Cliffs, New Jersey: Prentice-Hall, 1971. pp. 37-72.

Shavelson, Richard and Nancy Dempsey-Atwood. "Generalizability of Measures of Teaching Behaviors." *Reviews of Educational Research* 46(4): 553-611; Fall 1976.



Shirley B. Stow is Curriculum Coordinator, West Des Moines Community School District, West Des Moines, Iowa.

Copyright © 1979 by the Association for Supervision and Curriculum Development. All rights reserved.