

Elma S. Herradura

Analyzing a Study of Measurement of Teacher Merit

The study being discussed here was done in Raleigh, North Carolina, but is presented here for the implications it may have for the evolvement of teacher-rating scales.

This study of teacher merit was authorized by the State Education Commission which was created by the General Assembly of North Carolina. The latter body wanted to know what convenient indices of teacher worth are valid enough to justify their use for salary purposes. Although the researchers failed to fully realize this objective, the study was of value in making those concerned critical of existing procedure for merit rating. It also led to several findings of educational significance about the characteristics of a good teacher.

The research did serve one immediate purpose—to determine the validity of the rating of teachers' merit by their official superiors, an index which the General Assembly had thought of using to determine teachers' salaries. The researchers concluded that this method of rating lacks

sufficient validity to justify its use for salary purposes.

The researchers felt that the proved ability of a teacher to make desirable, balanced changes in pupils is the most acceptable criterion of teacher merit; however, they did not recommend its use because of the time, trouble, and expense involved. The study did show that a battery of direct tests like the personality paired-traits method used by the principal and by the teacher's peers, and the pupils' ratings of their teachers on the social behavior scale, would be more valid than the State's existing system of measuring merit by training and experience. Here again, the researchers believed that the expense and the complexity of such a battery makes its use prohibitive.

(Please turn to page 10)

In failing to recommend any system of measuring teacher merit as basis for paying teachers' salaries, McCall reflects the growing consensus among educational writers that the major objective of teacher appraisal is to improve performance and that it is likely to be handled effectively if it is not tied to salary problems.

The research director points out at the outset that those who cannot agree that the proved ability to produce desirable growth in pupils is the most acceptable criterion of teacher worth, will not be able to accept any of the conclusions. It is easy to agree with McCall and Campbell¹ that the effectiveness of the teaching performance can be determined by the growth in achievement that pupils make while under the direction of the teacher. However, one cannot help questioning the use of the change-in-pupil achievement as the only criterion for determining teacher worth. For one thing, this method of measuring teacher worth does not adequately recognize the influences on pupil learning other than the teacher, influences like home and out-of-school experiences, and emotional and psychological factors.

Moreover, pupil growth cannot really be fully measured, notwithstanding the rather wide variety of tests used. The researcher, in anticipation of this criticism, expressed the belief that it was not necessary to measure every growth produced in the pupils, and that the research should provide only for a reasonably adequate sampling of all types of good growth.

¹Roald Campbell, et al., *Introduction to Educational Administration* (Boston: Allyn and Bacon, 1958), p. 106.

Conceding this, one can still argue that growth in terms of attitudes and values developed in children may elude measurement and may not even be evident until years later.

All these "misgivings" about the use of change-in-pupil achievement as index of teacher merit are expressed not so much to question the wisdom of the choice as to show the complexity of pupil growth and the difficulty of measuring it. This fact should give an idea of the difficulties involved in merit pay.

Another objection to the use of the growth criterion is that differences of class size and of the capacity of classes to "grow" may invalidate the criterion. To answer this objection, the researchers had taken measures to make corrections for variations in class size and capacity to grow. How effective these measures were is hard to tell. As there was consensus that considerable growth in handwriting, for example, was not worth so much as an equally large growth in social behavior, various growths were "weighted" on the basis of common sense as well as on judgment of test experts.

Although one cannot accept the growth criterion without hesitation, one is forced to concede after considering other criteria, that it is difficult to find another criterion that will call forth less criticism. For example, experience may immediately evoke such objections as "Ten years' experience may be only one-year experience repeated ten times"; college marks may call forth, "Bright students do not necessarily make good teachers; in fact many bright students make poor teachers." Ratings made by teachers' superiors or peers are often decried as subjective and often based on personality characteristics rather than on performance. Thus, in spite of misgivings about the validity of the growth criterion, one is forced to agree

with the researchers that it seems to be the best criterion for evaluating worth of teachers.

An interesting part of the research is the surprising results of the evaluation of conventional methods of measuring teacher merit. The validity of five conventional methods and of 20 new methods of measuring teacher merit was determined on the basis of the criterion-score (the teacher's score as determined by the growth produced in pupils). The study showed that the index of validity of training is 10 per cent;² amount of experience, 12 per cent; rating by principals, -6 per cent; rating by peers, -11 per cent; confidential rating of teacher by himself, 39 per cent.

Of the new methods only the following had indices of validity of 20 per cent and higher.

- Personality paired-traits method used by the principal 20%
- Pupils' ratings of their teacher on the social behaviour scale between 22 and 39%
- Test of political tolerance.. 24%

The following methods had negative indices of validity:

- 1. Personality plus-minus technique used by the principal -6%
 - 2. Personality plus-minus techniques used by peers -10%
 - 3. College marks -28%
 - 4. Morris trait index -7%
 - 5. Questionnaire on variety of games played -43%
- (Please turn to page 26)

1. The index of validity is reported in per cents in this study. Whether the researchers meant 10 instead of 10% or whether the index of validity was computed in terms of per cents rather than *r*'s, was not made clear in the research report.

Elma S. Herradura..

(Continued from page 10)

- 6. Questionnaire on amount of time spent in sports activity -29%
- 7. Comprehensive test . . . -13%
- 8. Composition test -5%
- 9. Age -11%

From these findings, it seems that the best judges of the worth of teachers are the pupils and the teachers themselves. It is obvious why McCall could not recommend these two measures for salary purposes. As McCall pointed out, the propriety of having teachers' salaries determined by pupils' opinion is highly questionable.

Toward the solution of determining teachers' salaries, McCall recommends that teacher-training institutions should accept the responsibility for making their curricula functional and their marks and measures valid. Then, McCall believes, salaries of young teachers coming into service should be based solely on training.

Until all teacher-training institutions seriously accept such a responsibility, what bases should be used for determining pay? Unless care is exercised to make sure that instruments for measuring pupil growth are valid and unless corrections for variations in class size and pupils' native capacities are made, the use of the growth criterion may prove unjust. Even if these conditions are met, it seems necessary to supplement it with one or two criteria. The question is "which criteria?" Further research similar to McCall's study is needed before much weight can be given to the criteria which McCall found to be positively related to the growth criterion.