DOCUMENT RESUME

ED 128 335 SP 010 407

AUTHOR Kennedy, John J.; Bush, Andrew J.
TITLE Overcoming Some Impediments to the Study of Teacher Effectiveness.
NOTE 14p.; Paper presented at the National Invitational Conference on Research on Teacher Effects: An Examination by Policy Makers and Researchers (Austin, Texas, November 3-5, 1975)

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS *Academic Achievement; Classroom Research; *Educational Research; *Effective Teaching; Learning Processes; Measurement Techniques; Performance Specifications; Predictive Validity; Research Methodology; *Research Problems; Statistical Analysis; Student Behavior; Student Characteristics; Student Improvement; Student Teacher Relationship; *Teacher Influence

ABSTRACT

The study of teacher effectiveness is confronted by a number of problems that are generally associated with the conduct of behavioral research. It is possible in some instances to resolve or circumvent some of the current methodological stumbling blocks that tend to reduce the credibility of research findings. This paper discusses three methodological problems: (1) the importance of the teacher relative to his ability to affect student growth; (2) the attempts to operationalize constructs that appear to be related to student outcomes; and (3) the statistical problems associated with measuring student growth. Several alternative solutions to these problem areas are presented. (JMF)

# OVERCOMING SOME IMPEDIMENTS TO THE STUDY OF TEACHER EFFECTIVENESS

John J. Kennedy and Andrew J. Bush

The Ohio State University

Researchers who study how teachers affect student behavior are confronted with the great majority of problems that are generally associated with the conduct of behavioral research. A perusal of the methodological article by Berliner, contained within this issue, provides one with an appreciation of the many impediments that are inherent in teacher effects research. Although, as the Berliner paper reveals, the problems are many and serious, it is possible in some instances to resolve or circumvent some of the current methodological stumbling blocks that tend to reduce the credibility of research findings and discourage many able educators from conducting research in this area.

The purpose of this note is to address three methodological problems that were frequently discussed, both formally and informally, at the National Invitational Conference on Research on Teacher Effects. The three problem areas briefly discussed are somewhat representative of the wide range of existing impediments. The importance of the teacher relative to his or her ability to affect student growth constitutes the first problem. The second is that of attempting to operationalize constructs that appear to be related to student outcomes. The third and final topic concerns a statistical problem associated with measuring student growth.

---

John J. Kennedy, Professor, Faculty of Educational Development, College of Education, The Ohio State University and Andrew J. Bush, Graduate Assistant, Faculty of Educational Development, College of Education, The Ohio State University.

This note is primarily addressed to two types of educators: (a) those who are contemplating doing teacher-effects research and (b) those who are presently doing research in this most important area. Relative to incipient researchers, our goal is to simply acquaint them with three of the many methodological problems that they will shortly confront. To those able researchers currently attempting to link teacher variables to student outcomes, we hope to be able to propose an idea or two that might assist them in their important work.

## The Relative Importance of Teaching Variables

Perhaps the most fundamental problem relative to the conduct, interpretation, and appreciation of research into teacher effects is the fact that the boundaries of this field of inquiry have not been clearly established. Unfortunately, at present, we cannot answer with confidence the following question: what influence can a teacher *per se* exert on a child's learning and development? It is obvious from the studies contained in this issue, and elsewhere (1), that teachers *do* influence the quality and quantity of student learning. But, how important are teacher variables in comparison to other known correlates of student achievement such as socio-economic class, ability level, etc.? Empirically based answers to this question are needed if we are to establish realistic expectations for the potential results of future studies on teacher effects and if we are to convince public audiences and policy makers of the need to support future research efforts in this area.

Parenthetically, the need to determine the relative contribution of teacher variables to v[illegible] in pupil achievement is particularly important with respect to the public audiences. As Berliner (2) has pointed out, inferences drawn from the well publicized research of Coleman (3) and Jencks (4) have promoted the misleading impression that teacher and school variables contribute little to the academic, and even economic, attainment of students. Further, the impression has been created that the greatest ultimate educational payoff will result from working with social and attitudinal factors at the expense of school and teacher variables. Fortunately, there exists today a growing awareness that these impressions have been overdrawn. Not only can the studies upon which these inferences are based be challenged on both theoretical and methodological grounds, but there exists an increasing body of research which both advances contradictory findings and demonstrates the promise of additional research into the nature of school and teacher variables.

Recognizing the importance of this issue, recent attempts have been made by educational researchers to assess the relative contributions of teacher variables to pupil achievement. McDonald (5), for example, has hypothesized that teachers may account for as much as 25 percent of the variance in reading achievement socres at the elementary level. McDonald admits, however, that the estimates used in his argument were crude. But, there is a more fundamental limitation to current attempts to identify the relative contribution of teaching to achievement sc variance. Briefly, efforts such as McDonald's concentrate on attempting to explain the 20 to 40 percent of achievement score variance which is not accounted for by the relationship between beginning-of-year and end-of-year performance. Not only are teacher "main effects" not represented in this pool of residual variance, but there are simply too many methodological problems inherent in the regression-type analysis (e.g., multicollinearity) to fee. confident that the residual variance reflects differential teacher effects.

We believe that there is a more appropriate methodology to assess the contribution of teaching to achievement variance. It is a method- o[illegible] [illegible]ly in fields such as agriculture and ar[illegible] [illegible]ilar goals. In animal genetics, for [illegible] in determining, over a generation, t[illegible] [illegible]ed trait in cows, he or she employs a hie[illegible] or [illegible] design in which the lead factor is comprised of a represen[illegible] [illegible]mple of bulls. Several cows are then <u>nested</u> within each level of the lead variable and the resultant progeny of the inevitable liasons are measured for the trait under study. Specifically, through the estimation of <u>variance components</u>, the relative genetic contributions made by both bulls and cows are established.

Similar types of educational studies can be readily designed to specifically estimate the relative contribution to variance of such factors as schools, teachers, and classrooms. The first step of such a study might be to assemble a representative sample of schools so that the influence of school and community context variables can be estimated. Since teachers are naturally <u>nested</u> within schools and classes are usually <u>nested</u> within teachers, these latter two random variables can also be built into the design in a hierarchical fashion. To avoid problems associated with change scores, pretest and posttest achievement scores could be treated as a repeated measurement variable and crossed with all levels of the school and classroom variable. Finally, it would be intersting to entertain one or more important context variables such as social class or ability le[illegible] students. Such variables could be built into the design as a _wit[illegible]n-class_ variable. Student test scores would then be subjected to ANOVA procedures. Relative to the proposed study, by computing _variance component estimates_ it would be possible to assess the proportion of achievement score variance attributable to: (a) school-community factors, (b) teacher effects, and (c) classrooms. Using techniques app[illegible]priate for fixed variables (e.g., eta squared or omega squared coefficients), the relative contributions of context variables could also be estimated. In conclusion, by conducting a few studies along the lines proposed above, it would be only a matter of time before one of our most troublesome methodological problems and public issues would be resolved.

## The Problem of Construct Definition

In one of the most comprehensive reviews of teacher effects research, Rosenshine and Furst (6) synthesized the results of approximately fifty studies which, for the most part, were studies which correlated teacher process variables with student achievement gain. The synthesis produced eleven categories of teacher behaviors that were apparently related to student achievement. The categories were further broken down into a principal set of five behaviors construed to have strong research support, and a secondary set of six behaviors judged to have weaker support. Members of the principal set in decreasing order of apparent strength are as follows: clarity, variability, enthusiasm, task-oriented and/or businesslike behaviors, and student opportunity to learn criterion material. The six weaker but promising behaviors are: use of student ideas and general indirectness, criticism, use of structuring comments, types of questions, probing level of difficulty of instructions.

Although the fifty reviewed studies are subject to criticisms on methodological grounds, in toto they do represent the most solid body of evidence for consistently demonstrating that teacher behavior is related to measures of student achievement. Unfortunately, the behavioral complexes supported are just that -- complex. Thus, a body of our most promising research is plagued by problems of definition and operationalism.

As a case in point, consider the teacher-behavior construct with most research support, clarity. Suppose teacher clarity is defined as "being clear and easy to understand." Obviously, such a definition is circular. Yet, this is an example of the most common kind of clarity definition to be found in clarity research. A construct defined in this manner cannot be readily observed or measured. In fact, an observer must infer its existence. From a measurement perspective, an observer is required to make a rating rather than a simple record of occurrence. Since behaviors that demand rating procedures -- termed high-inference behaviors -- are by nature ambiguous, their use in research sets the stage for evaluating the findings of such studies with suspicion. One potentially profitable method for escaping the inherent problems of using high-inference variables is to identify their low-inference constituents, i.e., behaviors which are amenable to direct observation and tallying. To the extent to which low-inference constituents can be determined, the potential for conducting research which will yield more definitive results is increased.

Using the clarity construct as a working example, we propose the following blueprint for so reducing this high-inference construct. First, a tentative mapping of the domain, in low-inference terms, is necessary. One way of getting such a mapping might be to ask a large number of students to think of their most "clear" teacher and list some specific behaviors that make that particular teacher "clear." Similarly, the same operation can be carried out for the most "unclear" teacher. Subsequent to obtaining these behaviors, experienced educators can analyze and categorize the results into sets containing well-defined, easily observable behaviors.

Next, the tentative mapping can be put to empirical test by first asking large samples of students to think of their most "clear" teacher and to relate how often that teacher exhibits each of the behaviors. Once similar observations are obtained from students who are instructed to think of their most "unclear" teacher, the two sets of data can be aggregated and subjected to discriminant function analysis. This multivariate technique can be used to discover if the tentative mapping distinguishes significantly between teachers perceived to be "clear" and those perceived to be "unclear." If so, then those behaviors that contribute heavily to the discrimination can be regarded as a set of low-inference behaviors that at least map one portion of the clarity construct.

Such an approach should be heavily replicated for at least two reasons. First, apparent significance may always be the result of chance alone; consequently, replications with similar results are needed to strengthen the conjecture that such findings are not chance artifacts. Secondly, unknown biases may be in operation when students suggest behaviors or when educators refine them into workable form. It is possible that such unknown factors may serve to restrict the scope of responses, thus preventing comprehensive mapping of the high-inference construct. Again replication helps by increasing the potential for broad coverage of the construct domain.

In conclusion, it is difficult to argue with the spirit of Rosenshine's recent contention that the greatest current need is to conduct more research which is designed to link teacher variables with student outcomes. However, studies which attempt to further explore the relationships between student growth and the eleven or so correlates advanced by Rosenshine will be greatly hampered, and their value possibly reduced, until serious attention is given to the problem of defining these abstract constructs in terms of low-inference behaviors. Parenthetically, studies whose purpose is to identify the specific components of several high-inference constructs are currently being conducted at The Ohio State University.

## The Change Score Dilemma

The emerging paradigm for teacher effects research consists of relating promising teacher presage or process measures to measured changes in pupil learning. The objective of these studies is to identify teacher variables which correlate meaningfully with student change or, if the study is an experiment, to identify treatment conditions which are responsible for maximal gain. Measures of change or gain are sometimes calculated by simply subtracting pretest scores from posttest scores. However, due to the growing awareness that raw change scores are susceptible to regression effects (7), more often researchers attempt to "adjust" raw scores for regression toward the mean by partialing out differential pretest performance.

Unfortunately, it is known that even adjusted or residualized gain scores, despite their intuitive appeal, are not suitable measures of change (8). A major problem, as mathematically demonstrated by Bereiter (9), is that change scores based on residuals "over-correct." Specifically, to the extent to which error of measurement is reflected in pretest scores, residualized gain scores will be spuriously large for low-pretest performers and spuriously small for students who earn high-pretest scores. Consequently, if the research is descriptive and calls for computing correlations between teacher variables and residualized gain scores, to the extent to which teacher variables covary with pretest performance, the resultant correlations will be spurious. By way of simple illustration, consider a hypothetical situation in which it is desired to estimate the correlation between teacher age and mean student gain in reading over a school year. Suppose reasonable samples of classrooms are studied where, for each, the age of the teacher and the residualized gain in reading for the year are obtained. Now assume that it turns out that the youngest teachers in the sample tend to be located in inner-city schools and oldest teachers in outer-city schools. Suppose further that outer-city pretest scores are higher on the average. In this case, the computed correlation coefficient between age and gain would be biased in a negative direction, suggesting falsely that students of younger teachers experience greater achievement gains. In sum, those who study teacher effects are confronted with a rather serious methodological problem -- a problem which is particularly serious because of the modest and fragile nature of the correlations that are usually obtained.

In discussing ways in which the change scores might be minimized, it is important to separate experimental and correlational research. Measuring change is more tractable within an experimental context; in fact an experimenter is presented with several alternative methods which can completely circumvent the use of change scores. An approach which is most justifiable when pupils have been randomly assigned to repsective treatment conditions is to perform an analysis of covariance on posttest scores using pretest scores as the covariable. Essentially, this was the strategy employed by Gage in the experiment reported in this issue. Even though the over-correction phenomenon mentioned earlier is still reflected in adjusted posttest scores, the random distribution of adjusted scores among treatment conditions selectively controls for pretreatment inequalities. Analysis of covariance should be used cautiously, however, and only by data analysts who are familiar with its many subtle limitations.

A most direct alternative with less demanding statistical assumptions is to create a blocking variable from pretest scores and build this variable into the design of the experiment. In the simplest case where there is a treatment variable crossed with the pretest variable, a standard two-factor ANOVA is performed on the posttest scores. If the subjects have been randomly assigned to treatment conditions, not only is the analysis capable of documenting significant gain, but it is also capable of detecting interactions between levels of pretest performance and treatments. Feldt (10) has discussed several advantages of this design in comparison to using analysis of covariance.

A third experimental option is to treat pretest and posttest scores as a single factor and to build this factor into the design as a repeated measurements variable. In the simplest case, a treatment variable consists of the pretest and posttest scores. If there should be greater gain under some treatment conditions, it will be detected by the presence of significant treatment by pre-post testing interaction in the ANOVA. If it has not been possible to initially equate treatment groups, this option is particularly attractive because the means to detect pre-treatment biases are readily available.

It is clear that for experimental work, there are ample alternatives to the use of change scores. Hence, as Cronbach and Furby (8) have concluded, "There appears to be no need to use measures of change as dependent variables and no virtue in using them."

Unfortunately, overcoming problems associated with change scores is not as easy when the research is of the correlational type. There exists, however, a method which has been shown by Lord (11) to be superior to computing correlations which involve residual gain scores. The method consists of: (a) completely correcting zero-order correlation for the unreliable variance in each measure (i.e., correcting for attenuation) and (b) using these corrected correlations to compute semi-partial correlation coefficients where pretest performance has been partialed out of posttest performance.

To illustrate the Lord method, consider again the relationship between teacher's age and gain in reading during the school year. One first obtains reasonable estimates of the reliability of age, pretest, and posttest variables. Using the reliability information within the context of standard formulae (12: p. 155), the semi-partial correlation between teacher's age and posttest scores is calculated. The resultant semi-partial correlation represents the correlation between age and reading achievement subsequent to removing initial reading ability from the posttest reading measure and further possesses the advantage that it is least vulnerable to the "over-correction" problem mentioned earlier. Granted, greater labor is expended in using this approach, but considering the importance of the relations being sought, this methodology should be used far more extensively in descriptive studies of teacher effects.

## Concluding Remarks

In this brief note we have only been able to mention three out of the vast array of impediments associated with scientific inquiry into the nature of teacher effects. Admittedly, we selected these three because, in our view, potential remedies lie close at hand. If nothing else, our purpose has been to show that some of the methodological obstacles confronting educational research can be overcome and to demonstrate that methodologically respectable teacher-effects research can be conducted. We hope that this article will encourage others to respond to this crucial need and challenge.

# REFERENCES


1. Dunkin, M.J., and Biddle, B.J. The Study of Teaching. New York: Holt, Rinehart and Winston, 1974.

2. Berliner, D.C. Impediments to the Study of Teacher Effectiveness, Paper presented at the National Invitational Conference on Teacher Effects, The University of Texas at Austin, November 1975.

3. Coleman, J.S., Campbell, E.Q., Hobson, C.J., McPhartland, J., Mood, A.N., Weinfeld, F.D., and York, R.L. Equality of Educational Opportunity. Washington, D.C.: U.S. Department of Health, Education and Welfare, Office of Education, U.S. Government Printing Office, 1966.

4. Jencks, C., Smith, M., Acland, H., Bane, M.J., Cohen, Gintis, H., Heyns, B., and Michelson, S. Inequality -- A Reassessment of the Effect of Family and Schooling in America. New York: Basic Books, 1972.

5. McDonald, F.J. Research on Teaching and Its Implication for Policy Making. Paper presented at the National Invitational Conference on Teacher Effects, The University of Texas at Austin, November 1975.

6. Rosenshine, B., and Furst, N.F. The Use of Direct Observation to Study Teaching. In M.W. Travers (ed.), Second Handbook on Research on Teaching. Chicago: Rand McNally, 1973.

7. Lord, F.N. Elementary Models for Measuring Change. In C.W. Harris (ed.), Problems in Measuring Change. Madison, Wis.: The University of Wisconsin Press, 1963.

8. Cronbach, L.J., and Furby, L. How We Should Measure Change -- Or Should We? Psychological Bulletin, 1970, 74, 68-80.

9. Bereiter, C. Some Persisting Dilemmas in the Measurement of Change. In C.W. Harris (ed.), Problems in Measuring Change. Madison, Wis.: The University of Wisconsin Press, 1963.

10. Feldt, L.S. A Comparison of the Precision of Three Experimental Designs Employing a Concomitant Variable. Psychometrika, 1958, 23, 335-353.

11. Lord, F.N. Further Problems in the Measurement of Growth. Educational and Psychological Measurement, 1958, 18, 437-454.

12. Nunally, J.C. Psychometric Theory. New York: McGraw-Hill, 1967.