The Journal of Educational Thought, Vol. 12, No. 3 (1978)

The contemporary behavioristic analysis of teaching and teacher effectiveness has assumed that there is a necessary empirical relationship between *differences* in teachers' classroom behaviors and *levels* of student achievement. It is argued that this assumption lacks both empirical support and conceptual plausibility. The consequences of "accepting the null hypothesis" of differential teacher effectiveness are discussed with reference to their implications for teacher evaluation, teacher accountability and teacher education.

James T. Sanders*

Teacher Effectiveness: Accepting The Null Hypothesis

The contemporary behavioristic analysis of teaching as well as its derivative empirical research on teacher effectiveness have been constrained within the so-called "process-product paradigm." They have been constrained in the sense that this paradigm has directed and delimited the study of teaching to the search for stable, empirical relationships between antecedent teacher behaviors (process variables) and consequent student outcomes (product variables). This conception of teaching and teacher effectiveness is acknowledged to be "the heart of the performance- and competency-based approaches to teacher education, teacher evaluation and teacher accountability (which) has to be the empirically established relationship between teacher behavior as an independent variable and student cognitive and affective outcomes as dependent variables."[1]

The expectation that there should be an empirical relationship between teacher behaviors and student outcomes would seem to represent a rather innocuous working assumption. One could, of course, argue intelligibly that there can be learning without teaching and by some definitions, teaching without learning. Nevertheless, the assumption that teaching behaviors and student learning outcomes are, at least, likely to be empirically related seems to be a modest, non-controversial premise. It could be considered as simply a special case (although a complicated one) of the fundamental S-R connectionism presupposed by behaviorism — with teaching as the "complex stimulus" and student outcomes as the "complex response." Indeed, much of the credibility and *prima facie* appeal of the behavioristic view of teaching no doubt derives from the persuasive simplicity of the conception.

Beneath the stimulus-response simplicity of the process-product paradigm, however, is a less obvious suppressed premise. In particular, what this simple characterization of the behavioristic conception of teaching obscures is the implicit expectation that there should be a consistent empirical relationship between quantitative and/or qualitative *differences* in teacher behavior and *levels* of student achievement outcomes. Or to put the matter squarely, the real "heart" of the performance-based approaches to teacher education, teacher evaluation and teacher accountability is the assumption that different values of the independent variable are necessarily related to different values of the dependent variable.

---

*James T. Sanders is an Associate Professor of Educational Psychology at the University of Western Ontario, London, Ontario, Canada.

This suppressed premise is most clearly revealed when one considers the accountability imperative that derives from the behavioristic definition of teacher effectiveness. The behavioristic view requires that teacher effectiveness be defined exclusively in terms of increases in student achievement. The proof of teacher effectiveness is in the pudding of student achievement. Consequently, teachers *should* be held accountable for their effectiveness (or ineffectiveness) in producing demonstrable or measurable gains in student achievement.

It is important to note that this product or outcome criterion of teacher effectiveness, and thus accountability, *requires* that all other plausible and potent sources of variation in student achievement be controlled or held constant. That is, to determine how much of the variance in student achievement is 'explained by' or 'accounted for' by the variance in teacher behavior necessitates removing all other significant sources of student-achievement variance, *e.g.,* curricular content, modes of instruction, student abilities, etc. For example, it would obviously make little sense to judge a teacher of the mentally retarded as less effective than a teacher of gifted children because the former failed to produce comparable levels of reading achievement in his students. Clearly, ability differences 'account for' much, if not most, of the variance in reading achievement between the two groups of students. Strictly speaking then, the empirical meaning of teacher effectiveness, and thus accountability, depends upon a *ceteris paribus* clause with respect to all other plausible determinants of student achievement.

Given the widespread commitment to the behavioristic model of teacher effectiveness and to performance-based teacher education it is obviously crucial to ask: "Does the empirical research support the general hypothesis that teacher behavior is *differentially* related to student outcomes?" And the answer, with far less equivocation than is usually the case, seems to be "No!" For example, one of the most recent and comprehensive reviews of the empirical research concludes bluntly that "there is no established relationship between teacher behavior and student achievement."[2] As Hertzberg observes, it is ironic that this behavioristic analysis of teaching "which so stresses accountability is not itself accountable to a body of research which would firmly support its central premises."[3] Furthermore, the failure of empirical research to support its central premises seems neither to embarrass nor to discourage proponents of this conception of teacher effectiveness. On the contrary, they "admit freely — even cavalierly — that the evidence in terms of pupil outcomes is lacking."[4] For example, a recent book assertively entitled, *Teachers Make a Difference* begins:

> We would like to suggest how teachers make a difference. . .Unfortunately, research has not yet linked teacher behavior with student achievement in a direct associative way. Thus it is impossible to say that teaching behaviors x, y and z are associated with distinct areas of student achievement.[5]

As one might expect, the proponents' typical response to these discouraging empirical returns has been to cite an assortment of methodological problems, broadly construed, that may have thus far prevented confirmation of the general hypothesis. (It seems that hypothesis, like MacArthur's old soldiers, "never die; they just fade away.") Thus, Berliner maintains that "before researchers can adequately establish those relationships (between teacher behavior and student achievement) they need to deal with the problems of instrumentation, methodology and statistics."[6] This sanguine interpretation of null findings, of course, inevitably leads to that most familiar of empiricist rejoinders: "further research is needed. . ."

The more reluctant conclusion, however, that one might draw from the cumulative empirical evidence is that the general hypothesis of differential teacher effectiveness is simply no longer tenable and should be discarded. To put the conclusion in perhaps the strongest terms: If the empirical research on teacher effectiveness is meant to be taken seriously, then the evidence is now sufficient to "accept the null hypothesis." Unfortunately, the logical requirements of statistical inference dictate that "failing to reject" does not entail "accepting" the null hypothesis. And this logical imperative, no doubt, has made it that much more difficult to discard this less-than-promising hypothesis about teacher effectiveness.

If, however, one temporarily suspends the logic of statistical inference, then it may prove more instructive to reverse the usual inductive procedure and consider the evidence and arguments that tend to "confirm" the null hypothesis. Lest this heuristic suggestion be misunderstood, the null hypothesis being recommended for "test" is not that: Teaching has *no* effect upon student achievement. It is assumed that an experimental group that experiences "teaching" will demonstrate superior achievement when compared with a control group that experiences "no teaching." Rather, to repeat, the version of the null hypothesis here being proposed for "test" is that *differences* in teacher behavior (other things being equal) do *not* account for significant *differences* in student achievement.

Perhaps the most compelling support of the null hypothesis comes from the so-called Coleman Report[7] which concluded that differences in school quality did *not* account for much of the variance in student achievement over and above what could be predicted from measures of student ability and socio-economic status. Christopher Jencks' further judgement is that "variations in schools" fiscal and human resources have very little effect on student achievement — probably even less that the Coleman Report originally implied."[8] And finally, some researchers have gone so far as to speculate that if students' background variables were controlled, then *any* effects of schools *per se* would ultimately disappear.[9]

What seems to follow syllogistically from these conclusions and speculations is that *if* schools have a minimal effect upon student achievement, and *if* teachers are "nested" within schools, *then* teachers must have a minimal effect upon student achievement. In rebuttal, Brophy and Evertson have insisted that studies, like the Coleman Report, (1) typically have not obtained measures of teachers' classroom behavior and (2) may have masked potential teacher effects by aggregating data within and/or across schools. After summarily dismissing the implications of such studies, Brophy and Evertson, undaunted, urge that "what is needed now are studies that systematically record what teachers do in the classroom and relate these behavioral data to measures of student outcomes."[10] As Kurt Vonnegut says, "so it goes. . ."

This rebuttal, however, seems to miss the point — the point being that most, if not all, of the variation in student achievement seems to be accounted for by sources *other than* school and, by implication, teacher variables. That is, there is very little residual variance in student achievement left over to "explain" by a more thorough or detailed analysis of teacher behavior or by disaggregation of the pooled data on teachers. As Heath and Nielson argue, "given the well-documented, strong association between student and achievement and variables such as socioeconomic status and ethnic status, the effects of techniques of teaching on achievement are likely to be inherently trivial."[11] Given the minimal amount of variance in student achievement that remains to be explained, it is difficult to

understand what further "process-product" research on teacher effectiveness would accomplish.

Against the background of the persistent belief in differential teacher effectiveness, the null hypothesis that differences among teachers do *not* make a difference (at least, in student achievement) is likely to be perceived as counter-intuitive. Thus, it may serve understanding more to look behind the empirical evidence and to suggest an interpretation of the evidence that renders the null hypothesis in this instance intuitively more plausible.

In re-examining the predicted relationship between antecedent teacher behavior and consequent student achievement, it is important to keep in mind the well-documented, statistical-empirical behavior of the dependent variable — student achievement. In particular, there is considerable evidence that measures of student achievement *do* correlate substantially with other variables, for example, IQ or tested intelligence. Thus, aside from any of their intrinsic shortcomings, measures of student achievement have, at least, yielded sufficient individual variability to sustain substantial statistical-empirical relationships with other variables.

Statistical logic, however, requires sufficient variability on *both* variables as a necessary (but insufficient) condition for a statistically significant relationship. Therefore, it is relevant to consider whether the independent, process variable, namely, teacher behavior, yields or is likely to yield sufficient variability to hold up its end, as it were, of the predicted, empirical relationship. In short, does an analysis of what is meant by "teacher behavior" in the context of the behavioristic paradigm suggest that it is reasonable to expect substantial individual variation on measures of this independent variable?

The behavioristic analysis of teaching, as classroom process, has typically focussed upon discrete, observable, verbal-communicative behaviors displayed by teachers in the classroom, such as asking questions, giving directions, criticizing or praising students' responses, etc. The units of analysis may vary in their specificity and complexity, but in general the behavioral domain of interest is essentially the teacher's verbal behavior in the classroom. More importantly, however, the behavioristic conception of teaching has insisted upon designating these verbal-communication behaviors as "performance skills" or "competencies." This designation implies that these behaviors (1) constitute an ability domain and (2) admit of a range of individual variation from skillful to inept, or from competent to incompetent.

Calling these behaviors "skills" or "competencies," of course, does not make them so. Indeed, what is at issue is whether it is at all plausible to assume *a priori* that such behaviors as asking questions, giving directions, etc. constitute the kinds of skills, abilities or competencies that are likely to manifest important individual differences. Does it make conceptual sense to construe these classroom teaching behaviors as analogous to such skills as musical talent or mathematical ability or mechanical aptitude or any of the other abilities for which wide individual variation in performance levels are expected? It would seem unlikely. As Johnston argues:

> Most of the acts performed by teachers in the classroom could probably be performed by an intelligent adult and by some children, if they knew what should be done. Teachers probably vary little in their abilities to execute performances called for in a detailed instructional plan. The improvement that is possible in the actual performance is limited primarily to the achievement of somewhat greater poise and efficiency of action.[12]

Similarly, in his "theory of spontaneous schooling," J. M. Stephens contends that the communicative 'skills' required for classroom teaching are

> distributed throughout the general population. The basic mechanisms responsible for teaching reside in some very earthy, primitive tendencies. Although more pronounced in some people than in others, these tendencies are quite prevalent.[13]

Furthermore, when one considers that selection criteria make teachers a somewhat homogeneous sub-group of the general population with regard to communicative 'competence' and that classrooms represent highly stylized (if not ritualized) communicative environments, the range of individual variation in these 'skills' is likely to be even further attenuated.

If this analysis is correct, then there is nothing really very surprising or counter-intuitive about the null hypothesis. It simply implies that the domain of communicative behaviors required for classroom teaching are relatively commonplace and more or less equally distributed throughout the teaching population. As such, these behaviors are unlikely to admit of wide individual variation or to be susceptible to substantial improvement through training.

Lest acceptance of the null hypothesis be construed as tantamount to pedagogical nihilism, it remains to point up some of the broad (and perhaps even salutary) implications of its acceptance for teacher education, teacher evaluation, and teacher accountability. Clearly, acceptance would imply that the notion of "teacher effectiveness" must be defined in terms of other than the garden-variety, classroom behaviors that have been presumed to be related instrumentally to student achievement. The failure of classroom interactive variables to provide an adequate empirical definition of teacher effectiveness suggests that perhaps a more defensible conception of effectiveness is to be found in the pre-interactive phase of teaching. This is *not* to recommend a retrogressive return to the ill-fated, empirical research which has attempted to link measures of teacher personality to measures of student achievement. Rather, it is to suggest that a potentially more plausible conception of teacher effectiveness and teacher accountability might be grounded in individual variation in the preemptive knowledge and skills required for curriculum selection, design and justification. While such an alternative conception of teacher effectiveness may be necessarily more dependent upon non-instrumental, rational or intrinsic criteria for its articulation and justification, it would seem to offer a potentially more differentiated basis for assessing competence.

Acceptance of the null hypothesis also implies that teacher education, as a whole, might become less obsessed with the problems and pragmatics of classroom performance. It could begin to relax its emphasis upon the rehearsal and evaluation of student teachers' classroom performance secure in the knowledge that the improvement of classroom performance is, as Johnston suggests, "limited primarily to the achievement of somewhat greater poise and efficiency of action." The absence of an adequate empirical justification for classroom practices in terms of student achievement would seem to weaken the basis for the rigid distinction between "theory" and "practice" in teacher education — a not altogether undesirable outcome. The "practical" component in teacher preparation, like its "theoretical" counterpart, would be required to articulate an alternative justification for its curriculum beyond its assumed relevance to student achievement.

Notes

[1] D.C. Berliner, "Impediments to the Study of Teacher Effectiveness," *Journal of Teacher Education* 27 (Spring 1976): 5.

[2] R.W. Heath and M.A. Nielson, "The Research Basis for Performance-Based Teacher Education," *Review of Educational Research* 44 (1974): 463-484.

[3] H.W. Hertzberg, "Competency Based Teacher Education: Does It Have a Past or a Future?," *Teachers College Record* 78 (1976): 3.

[4] Hertzberg, p. 18.

[5] T.L. Good, B.J. Biddle, and J.E. Brophy, *Teachers Make A Difference* (New York: Holt, Rinehart & Winston, 1975), p. 8.

[6] Berliner, p. 12.

[7] J.S. Coleman *et al., Equality of Educational Opportunity* (Washington, D.C.: U.S. Office of Education, 1966).

[8] C.S. Jencks, "The Coleman Report: A Reappraisal of the Most Controversial Educational Document of Our Time," *The New York Times Magazine,* August 10, 1969, pp. 12-13, pp. 34-35.

[9] Phillip C.Schlechty, *Teaching and Social Behavior: Toward an Organizational Theory of Instruction* (Boston: Allyn & Bacon, 1976), pp. 1-2.

[10] J.E. Brophy and C.M. Evertson, *Learning from Teaching: A Developmental Perspective* (Boston: Allyn & Bacon, 1976), pp. 8-9.

[11] Heath and Nielson, p. 481.

[12] M. Johnston, "Conceptual Confusion and Premature Policies," in *Regaining Educational Leadership: Critical Essays on PBTE / CBTE, Behavioral Objectives and Accountability,* ed. R.A. Smith (New York: Wiley, 1975), p. 50.

[13] J.M. Stephens, *The Process of Schooling: A Psychological Examination* (New York: Holt, Rinehart & Winston, 1967), p. 137.