Abstract

Sanders' recent argument that differences in teacher behavior do not have any significant effect on levels of student achievement is flawed. His criticisms of research into teacher effectiveness are cogent but apply no less to the research into social factors that he cites to support his view. It is argued that what is really being revealed is the inevitable inadequacy of empirical research for contributing to knowledge of cause and effect in the classroom. Consequently, teacher education should concentrate more on developing conceptual finesse to enable teachers to judge for themselves what they should do in the classroom and to treat empirical research more sceptically.

*Robin Barrow

Teacher Judgement and Teacher Effectiveness

James T. Sanders has recently argued that we should accept "the null hypothesis of differential teacher effectiveness"[2] By this he means that we should conclude that differences in teachers' classroom behavior do not have any appreciable effect on levels of student achievement. I shall suggest that his argument is centrally flawed and that we cannot legitimately draw that conclusion. However I shall argue that there is nonetheless reason to accept the gist of his view that teacher education should become "less obsessed with . . . the pragmatics of classroom performance" and more concerned with the "knowledge . . . required for curriculum selection . . . and justification."[3]

His argument involves the following three claims:

1. Empirical research into teacher effectiveness does not in fact support "the general hypothesis that teacher behavior is *differentially* related to the student outcome."[4]
2. There is research that provides some "compelling support" for the null hypothesis.[5]
3. Despite seeming to be counter-intuitive, the null hypothesis is in fact intuitively plausible.

While it is the second of these claims that is most obviously vulnerable, nevertheless I shall comment on each of them in turn.

First, Sanders maintains that constraints within the process-product paradigm of North American research into teacher effectiveness have led to countless studies that actually suggest that "differences in teacher behavior (other things being equal) do not account for significant differences in student achievement."[6] At any rate, empirical research has failed to produce convincing support for the hypothesis that teacher behavior *is* differentially related to student outcomes. He cites many commentators who concede that they are not getting what they are looking for in their research and points out that "one of the most recent and comprehensive reviews of the empirical research"[7] concludes that "there is no established relationship between teacher behavior and student achievement."[8]

In this general observation Sanders is clearly correct and one could cite further examples to support the contention that research into teacher effectiveness has not established the kind of relationship between teacher behaviors and student achievement that it presupposes.[9] Sanders is also right to point out that empirical researchers have a tendency to acknowledge the shortcomings in most of what has been done and the serious difficulties in the way of doing better, and then to propose carrying on as before. They tend to be unwilling to question either their fundamental hypotheses or the appropriateness of their manner of research, and to presume that it is just a question of doing the job properly next time. That might conceivably be true, though I shall argue it

*Faculty of Education, Simon Fraser University

is not. But it is certainly difficult to feel comfortable with the lack of fit between, for example, Dunkin and Biddle's sober review of a mass of research,[10] which is peppered with criticisms of instruments, unwarranted inferences, designs, and gross conceptual confusions, and their conclusion that "at long last we are beginning to know what is actually going on in the classroom, as well as what produces and results from classroom events."[11] Even if it were true that we could in the future attain to such knowledge, a close study of their own text makes it quite clear that it is false to say that "we are beginning to know."

Second, Sanders does not merely wish to say that research so far has told us little. Observing that researchers are mistaken in moving from the assumption that teacher process affects student product to the view that "*differences* in teacher behavior" affect "*levels* of student achievement outcomes", he suggests that differences in teacher behavior (other things being equal) do *not* account for significant differences in student achievement."[12]

I think he is wrong. But I am less concerned with arguing the point that differences in teacher behavior do make a difference that with arguing that he has not made a case for adopting his conclusion.[13] The main flaw in his argument is that he ignores the force of the *ceteris paribus* clause. This "other things being equal" caveat is both one of the main problems in making sense of such research and the main reason why we cannot accept the null hypothesis. For in the classroom other things never are equal. Nor could they be.

We cannot accept the null hypothesis, because it has not been shown that different teaching behaviors have no effect on level of student achievement for *precisely the same reason* that it has not been shown that they do: no research has met the demands of the *ceteris paribus* clause. No research has successfully isolated variables.

There are a number of factors that stand in the way of success in such research. Some of them are merely contingent matters such as the practical difficulty of observing large enough truly random samples,[14] of ensuring that participants in the research do not behave uncharacteristically because of the research, and of effectively controlling some of the variables. Those sorts of problems are acknowledged by researchers, although the implications for the value of their work are not so often acknowledged. Then there are constraints that are unavoidable in the world as we know it: for example, a student or a teacher in any field study or experiment will have his particular personality and all that that entails. We cannot nullify its effects. Thirdly, there are logical constraints such as the impossibility of actually observing cause and effect.

Human ingenuity might deal better with some of the contingent problems than it has so far done, but it cannot do anything about the logical ones. And in particular one doubts whether it could solve the problem of juggling with the thousands of variables that actually are in play in a classroom setting.

But the heart of the problem is conceptual. Researchers, in admitting this, sometimes seem to think that the answer is greater clarity and specificity. However, that is not an adequate response. It is true that if we are to observe or experiment in a scientific manner we need clear concepts, such as we seldom get. It is not true that they necessarily should be specific. And it is, in any case, not true that that is all that is required: in addition we need readily observable and readily isolatable concepts. The fact is that these requirements cannot be met.

The emphasis on the importance of specificity in research presumably arises out of a recognition of the shortcomings of early research into such general notions as "directness".[15] A major criticism of such research has been that by concentrating on such a broad notion it was unable to say much about what aspects of directness were having any effect; more particularly, an unthinking assump-

tion that warmth and indirectness were co-extensive has been noted.[16] Such criticism is entirely well founded, but greater specificity does not necessarily mean greater clarity: a broad concept such as "directness" could in principle be clearly articulated; a relatively specific concept such as "receptiveness to pupils' responses" might be, and in the event has been, very far from clear.[17]

Here we move on to the second point: if we are going to observe, we need to be looking for things that we can observe. How does one observe teacher receptiveness? There are two well-worn answers. The first is that one provides an operational definition in terms of concrete observable behaviors. The trouble with that is that in many cases, such as this example, it does not do justice to the concept. Being receptive is not just a matter of doing certain particular things; it is a question of doing them in a certain way, recognition of which is dependent on knowing something about the intentions and attitudes of the agent and the expectations and perceptions of those providing the responses. Appreciation of that point leads to the other answer: one observes impressionistically: one judges. The problem with that is that it makes our research a vulnerable and variable matter of human judgement.

The third point is that if we want to make sense of our research we have to be able to see what we are looking for in isolation. This is rarely possible: one might in principle conceptualise "being receptive" in a manner that distinguishes it from other things, but teachers in practice will seldom be simply "receptive"; they will at the same time be "encouraging", "paying attention", "drawing others in" and so on. Meanwhile the students will necessarily be being affected by other factors such as their background, events at home that day, the fact that they are tired or have just fallen in love, aspects of their individual personality, etc. And then there is the question of whether the individual students *recognise* the receptiveness of the teacher that the researcher claims is being exhibited. Paradoxically research into broad concepts is better equipped to deal with this problem of isolation and recognition: it is easier to say with legitimacy "that is a 'direct' teacher" than it is to say "that is a 'receptive' teacher". Just as it is easier to recognise something as a building than it is to specify what kind of building it is [18]

Any research that wishes to tell us something reliable about correlation in teacher-student behavior must be able to operate in terms of a set of clearly articulated concepts that are properly conceived in behavioral terms and can be recognised independently. I believe that some such research, probably with broad concepts, is in fact possible, but it will be very limited in scope, of relatively little help to the practising teacher, involve no information about cause and effect, and unless the contingent difficulties are overcome, not be reliably generalisable. Certainly the vast bulk of research to date has not met these requirements.

Some, notably Jan Smedslund and Kieran Egan, have tried to argue that empirical research into such matters is necessarily pointless, since any hypothesis will be a fusion of the analytic and the arbitrary.[19] To some extent the claim will be true by definition and therefore empirical verification is irrelevant, and, in so far as it is not, it will be based upon local interpretation and therefore not be generalisable. One example given by Egan presumes that we wish to enquire into whether presenting lists in an ordered manner will facilitate learning them.[20] He argues that the fact that it will does not need to be empirically verified. Because of what we mean by "learning" and "order" it is logically necessary that learning should be ordered; the parrotting of an unordered jumble of items does not count as learning. On the other hand what will count as ordered may vary from one person to another, and therefore we cannot generalise from the observation that certain children learn a list if it is organised in a certain particular way. To others no order may be apparent.

This seems to me obviously correct as a characterisation of many instances of empirical research into teaching effectiveness. We obviously cannot conclude that all empirical research in education is

pointless, but what we can do is gain a sharper focus on what kind of research is feasible. And the answer is that empirical research can only be worthwhile if it operates with broad, clear, observable, distinguishable, concepts, and refrains from inference about cause and effect, except in those rare instances where some form of controlled experiment is possible. Those conditions will make such research a very limited affair indeed.

I return now to the relevance of this to Sanders' paper. He believes that "the most compelling support of the null hypothesis" comes from the Coleman report, and he further cites the "judgement" of Jencks and the "speculation" of Schlechty as support.[21] But the Coleman report cannot possibly be said to have *established* that differences in school quality do not account for much variance in student achievement. This is partly because it, no less than the teacher effectiveness research with which it is being contrasted, can be severely criticised on *ad hoc* grounds (for instance, the conclusions are based on differences among schools rather than individual teachers). But it is also because all that it purports to reveal is a high correlation between student ability and socio-economic status. The inference of cause and effect, let alone an inference of socio-economic status being the crucial or sole determinant of achievement, is quite unwarranted.[22] Who is to say that schools in high socio-economic areas did not make more effort, that teachers in such schools were not better matched to their students, that parents of such students were not more supportive, that teachers of such students did not presume they could achieve more, and so on and so forth? Other things were not equal in this study. Other things cannot be equal in research into education, except on those rare occasions that the stringent conditions noted above are met. And at this point we should add that any such research would have the demerit of not speaking to the real world. Suppose we did discover in experimental conditions that when all other variables are controlled teacher behaviour X is followed by student behaviour Y. What are we supposed to make of that, given that in the classroom teachers have to do X when other variables are not controlled? The effect then might be quite different or might differ from individual case to individual case.

What research, whether into teacher behavior or social context, could possibly hope to establish what aspects of a given result are due to what combination of social, personal and teaching factors? One only has to look at the variations within such research, whether broadly sociological, psychological or classroom based, to see that the factors that need to be considered added to the various possible combinations they give rise to are incalculable.

Third, Sanders' point is that the conception of teacher effectiveness in such research, being confined to observable "garden variety" skills, lacks plausibility.[23] He is correct in pointing to the inadequacy of seeing teaching as a set of performance skills: a set of such skills does not fully characterise the act of teaching. Not all the putative skills are truly skills, and many of the skills focussed on in the mainstream of research would not be likely to make noteworthy differences in pupil achievement. Likewise he may well be correct to suggest, following Johnston, that "most of the acts performed by teachers in the classroom could probably be performed by an intelligent adult,"[24] at any rate if the "acts" of teachers are confined to specifiable and observable performances.

But why does he conclude that "the improvement that is possible . . . is limited primarily to the achievement of somewhat greater poise and efficiency of action"?[25] To do that or to say as Sanders does that "the domain of communicative behaviors required for classroom teaching are (sic) relatively commonplace . . . and unlikely to admit of wide variation" is to forget that this is merely the implication of the narrow conception of teaching that has just been criticised. The reasoning supports the conclusion that this particular kind of research will not get us far and not the conclusion that we should accept the null hypothesis.

The most obvious and important possibility that is being ignored is that teacher effectiveness

depends upon judgement. That is to say that, even if we accept that teaching consists in a number of fairly commonplace techniques, what distinguishes the good from the bad teacher is *knowing when* to question, listen to, correct, inform, etc., *which* pupils in *what* manner for *how long*. One must agree with Sanders that no research helps us to answer such questions as yet, and for the reasons given above, I do not see how it could. But that is an argument for abandoning such research, for taking the emphasis off so-called skills and techniques, and for placing it instead on seeking a more thorough conceptual grasp of the nature of education. It is not an argument for concluding that how one teaches makes no difference to levels of student achievement. Indeed the intuitively plausible suggestion that the exercise of judgement by the teacher in making use of his battery of commonplace skills and strategies makes a profound difference (though not in isolation, since part of the need is to judge what other social, psychological and organisational factors make what behaviour appropriate at a given time) is not challenged by anything that Sanders says.

Such a view may suggest to some readers a sympathy with research into aptitude-treatment interactions (ATI). Certainly it implies sympathy with the belated recognition by some researches that "individual differences need to be considered at all points"[26] and that "interactions are not confined to the first order; the dimensions of the situation and of the person enter into complex interactions."[27] But ATI studies, though arising out of a more realistic appraisal of the situation than traditional field surveys or experiments in relation to teacher effectiveness, are in a sense merely the last gasp of those determined to maintain, against all reason, that empirical research into the finer points of cause and effect in the classroom is feasible. Leading exponents of ATI, while sometimes remaining grimly optimistic, themselves acknowledge the problems. The variables that may play a part in the ultimate achievements of students are "so complex as to push generalisations beyond our grasp, practically speaking,"[28] and, as noted above, the magnitude of our task is in fact greater still, since every varying combination too must in principle be taken into account. (Here I am simply ignoring the problems in conceptualizing them and thinking only of the problems of disentangling and handling them).

Snow is simply wrong when he writes "Note also that if by 1984, say, the population served in the school changes so that all tenth graders turn out to be high ability - high anxiety - high conformity . . . the instructional theory (generated by ATI studies) . . . just prescribes that in that year everyone is taught by the same method."[29] He is wrong because the fact that all tenth graders are the same in those respects tells us nothing about other factors that may affect how they individually respond to different teachers, who may themselves, because of factors to do with them, be better advised to do something different well rather than to do the prescribed thing badly.

Cronbach remarks, in reference to work by McKeachee and others, "In so far as a generalisation can be glimpsed through the tangle of evidence, it is this: the constructively motivated student (who seeks challenge and takes responsibility) is at his best when an instructor challenges him and then leaves him to pursue his own thoughts and projects. In contrast the defensive student tends to profit when the instructor lays out the work in detail."[30]

That is so clearly essentially analytic as to make further comment superfluous. Snow actually appears to be aware that what he is proposing is really formative evaluation: what will be gained by ATI studies will be "a rather different sort of theory from what we usually think of as theory . . . The kind of theories that come out of this are quite specific, limited in both time and place."[31] But those are not "a different sort of theory." They are not theories at all. They are, as is made explicit, accounts of what appears to be going on in particular situations, arrived at with a welcome recognition of some of the variety of variables at issue and the importance of their interrelationship, but without due allowance made for the inherent and insuperable difficulties in such an undertaking.

Sanders turns finally to the implications for teacher education. I am in full agreement with him that the practical component of teacher preparation needs to set its house in order, and cannot legitimately continue to preach ways of proceeding on an alleged scientific basis. But this is not because ways of proceeding make no difference. It is because research has not told us much that is both useful for the practising teacher and reliably validated, nor does it seem that it could. (I would draw a distinction here between on the one hand the claims of psychology and teacher research, and on the other some of the claims of sociology. The latter, because it has often confined itself to producing evidence for correlations between broad concepts, has produced information about sex, class and language, for example; such information, though it rapidly becomes popular knowledge and hence seems commonplace and though it does not in itself enable us to determine what to do about it, does at least seem to be information that a teacher ought to be in possession of. Psychology and teacher effectiveness research have largely thrown up truths of definition masquerading as empirical and claims which though they may be true have not been satisfactorily established.)[32]

But, if the preceding paragraphs seem generally sensible, then it surely follows that the main thrust of teacher education should be towards developing in teachers a far greater conceptual grip on education and other aspects of schooling combined with a far clearer grasp of the nature of research, such that its inevitable shortcomings may be appreciated. The former is required because the only way to hope to make rational judgements about how to proceed in particular situations is in the light of a thorough grasp of what the whole enterprise is about and what one's own part is supposed to contribute to it.[33] The latter is required to enable teachers to understand, not that the null hypothesis is true and that they are merely insignificant automatons, but that their contribution has to be based on their own judgement of the situation in which they find themselves.

Notes

[1] I am grateful to the reviewers of the *Journal of Educational Thought* for their assistance.

[2] James T. Sanders, "Teacher Effectiveness: Accepting the Null Hypothesis," *The Journal of Educational Thought*, Vol. 12, No. 3 (1978), p. 184.

[3] Ibid. p. 188.

[4] Ibid. p. 185.

[5] Ibid. p. 186.

[6] Ibid. p. 186.

[7] Ibid. p. 185.

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

[9] E.g., H.W. Hertzberg, "Competency Based Teacher Education: Does it have a Past or Future?" *Teachers College Record*, 78 (1976):3.
T.L. Good, B.J. Biddle and J.E. Brophy, *Teachers Make a Difference*, (New York; Holt, Rinehart and Winston, 1975). See also references in ch. 6 of Robin Barrow, *Giving Teaching Back to Teachers*, (Brighton: Wheatsheaf, 1984).

[10] M.J. Dunkin and B.J. Biddle, *The Study of Teaching*, (New York: Holt, Rinehart and Winston, 1974).

[11] Ibid. p. 418.

[12] Sanders, op. cit. p. 186.

[13] Sanders is however surely correct to suggest that we should not be hidebound by the logic of statistical inference to the extent of insisting that negative results are only that and do not imply the null hypothesis. There tends to be a remarkable inconsistency in empirical research here: while the rule that a negative correlation does not itself establish no connection is generally observed, the equally basic rule that a positive correlation cannot be construed as cause and effect is regularly broken.

[14] "Unfortunately, to our knowledge *no* studies of teaching have yet made use of representative samples", Dunkin and Biddle, op. cit., p. 80.

[15] See, for example, N.A. Flanders, "Teacher influence in the classroom" in E.J. Amidon and J.B. Hough (eds), *Interaction Analysis: Theory, Research and Application.* (Reading, Mass.: Addison-Wesley, 1967).

[16] See Dunkin and Biddle, op. cit., chapter 5.

[17] "Acceptance of pupils' ideas" was, for instance, a category in much of Flanders' research. But what constitutes acceptance of an idea? Providing some examples for observers does not suffice for defining the concept, as Plato was at great pains to point out. One needs to know what makes an example an example in order to recognise further examples. And then, even assuming one has an adequate conception, there is still a further problem of recognising instances. I know what happiness is, but I cannot always detect it in others.

[18] I do not offer this as an invariable rule. Some specific concepts are easier to handle than some broad ones. But generally speaking broad human attributes are easier to recognise than specific ones. I may more confidently classify a person as kind/unkind, generous/mean, quiet/loud than as tender, jealous, or even anti-semitic. There are many reasons for this. (Jealousy, for example, is difficult to recognise because it involves knowledge of the person's state of mind as well as behavior). But in general more specific traits require a longer period of time to recognise. You cannot deduce that a person is anti-semitic from his hostility to one Jew. It is necessary to see over a period of time that he is always hostile to Jews and not to other people. Whereas a generally hostile manner may be deduced from consistent hostility to all over a more concentrated time-span. On specificity and clarity of concepts, see Robin Barrow *The Philosophy of Schooling* (Brighton: Wheatsheaf, 1981) and *Injustice, Inequality and Ethics* (Brighton: Wheatsheaf, 1982).

[19] Jan Smedslund, "Between the Analytic and the Arbitrary: A Case Study of Psychological Research," *Scandinavian Journal of Psychology*, 20. (1979). Kieran Egan, *Education and Psychology*, (New York: Teachers College, 1983).

[20] Kieran Egan in an unpublished paper.

[21] J.S. Coleman et al., *Equality of Educational Opportunity.* (Washington, D.C.: U.S. Office of Education, 1966). C.S. Jencks, "The Coleman Report: A Reappraisal of the Most Controversial Educational Document of Our Time," *The New York Times Magazine*, Aug. 10, 1969, pp. 12-22, pp. 34-35. Phillip C. Schlechty, *Teaching and Social Behavior: Towards an Organizational Theory of Instruction.* (Boston: Allyn and Bacon, 1976).

[22] Unwarranted by the rules of the game and the very rules that Sanders draws on to criticise research into teacher effectiveness.

[23] Sanders, op. cit. p. 188.

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

[25] Ibid, p. 50.

[26] Richard E. Snow. "Individual Differences and Instructional Theory". *Educational Research*, November, 1977, p. 13.

[27] Lee J. Cronbach, "Beyond the Two Disciplines of Scientific Psychology." *American Psychologist*, February, 1975, p. 116.

[28] Richard E. Snow, op. cit., p. 12.

[29] Ibid, p. 14.

[30] Lee J. Cronbach, p. 119. How could it be otherwise than that one who seeks challenge and welcomes responsibility likes to be challenged and to be given the responsibility of pursuing his own thoughts and projects? And in this context what is *meant* by a "defensive" student is, presumably, one who doesn't wish to be challenged and given responsibility but to be told what to do!

[31] Richard E. Snow, op. cit., p. 12. If Snow insists on calling this theory, so be it. The point is that by his admission we cannot generalise from such studies.

[32] As an example of a necessary truth masquerading as an empirical dicovery I may cite Hilgard's assertion that the proposition that "brighter people can learn things less bright ones cannot" is one empirically established by psychologists.

[33] See further Robin Barrow, *The Philosophy of Schooling*, op. cit.