THE STUDY OF SAMPLES AND THE STUDY OF CASES

Presidential Address to the Annual Conference of the British Educational Research Association, September, 1979.

by Lawrence Stenhouse, Professor of Education and Director of the Centre for Applied Research in Education, University of East Anglia, Norwich, England.

Let me begin by recalling that last year I offered this conference a definition of research as 'systematic inquiry made public'. Such inquiry is a response - alternative, for example, to prayer or contemplation - to a problem, and it aims to solve the problem by the achievement of understanding. Two kinds of problems of understanding to which systematic inquiry is a possible response are: problems of understanding the world in which we are called upon to act and problems of understanding what we ought to try to do.

In basic research - as opposed to applied research in such a field as education - these two problems are integrated by theory which is created by and addressed to researchers. Theory is such an understanding of the world as enables us to decide how to act as researchers. It structures knowledge in such a way as to let us plan by what research act we shall attempt to advance knowledge. The function of academic or 'pure' theory is to support the planning of research acts.

Educational research has as its overriding aim the support of educational acts - it is not 'pure' but 'applied'. Yet it must also support the planning of research acts in educational settings. Our problem is to find approaches to research which produce theory which is of use both to practitioners of education and to practitioners of educational research and which enables both to act in the light of systematic intelligence.

Now, I could attempt to build on this foundation an analysis of the relation of theory and practice in educational research; but that is not my purpose today. I want to hunt a dichotomy, which I perceive not merely as a logical distinction, but also as embodied in the social transactions of the educational research community, which as a consequence, may allegorically be throught of as a two-headed animal.

The two heads are constantly disagreeing but the terms of the dispute change. As the chameleon changes its colours and Proteus his shape, so this dichotomy changes its verbiage. The tete-a-tete distinction is now between 'quantitative' and 'qualitative', now between 'psycho-statistical' and 'ethnographic' and now again between 'positivist' and 'humanist'.

The product model is opposed to the process model, the conceptually abstracted to the naturalistic approach. Each of these categorisations reflects differences of value or taste: the formal garden on the one hand, the cunning simulation of nature on the other. Each too reflects individual abilities and disables one of the dichotomis heads; for words fail the psycho-statistician and the ethnographer doesn't count. These underlying differences of temperament and training are by no means irrelevant. They reflect real determinants of style.

But my purpose is to conjure the dichotomy to chop logic; to seek a formulation that will establish some logical - rather than stylistic or psychological - relationships across the cloven heads as a basis for reflective discourse rather than competitive banter. The problem is to get our dichotomy's heads equipped to speak to each other.

In this spirit I'll try asserting that the most important distinction in educational research at this moment is that between the study of samples and the study of cases.

Scientific experimenters in such a field as chemistry use samples - of such substances as zinc or hydrochoric acid or ammonium - and generalize the results from a laboratory experiment to a defined 'target population'. The basis of their procedure is control: the control of the purity of the substances. Pure samples represent accurately the behaviour of all possible samples of the pure substances subjected to the same processes.

In the life sciences control of purity cannot often be established. Instead it seems one must represent in the sample the range of variation in the population to which one must generalize. It would seem that the sample needs to be judged because it cannot be tested or controlled for purity.

It was the striking virtue of Ronald Fisher's Design of Experiments (1935) that he recognized and presented so clearly the idea that in experimental situations where we aspire to predictive generalization random sampling is to be preferred to judgemental sampling, the reason for this being that the error in judgemental sampling is inaccessible to estimate whereas the error in random sampling can be calculated by an application of the mathematics of probability. The so-called 'psycho-statistical paradigm' in educational research was founded upon this insight.

Now, although Einstein rebelled against the use of probabilistic results to support theory in physics, nevertheless we might well concede that at any particular time in the progress of a pure science, the discrimination between competing theories might be a matter of probability and such a situation could well guide research acts. Fisher, however, was working not in a pure science, but in the applied science of agriculture, where the purpose of experiment was to choose among alternative cultures of crops or animals. A decision as to how best to act agriculturally was at stake.

The criterion for this decision was gross crop yield, and the assumption was that of consistent treatment within the field. That is, treatment could not be differentiated from plant to plant on an individual or sub-group basis.

This experimental procedure was adapted to education, sometimes with good results. But there were always limitations. While in agriculture it is normally accepted that the fate of individual seeds of corn or individual battery hens - does not matter unless it makes the relation of investment to gross yield unfavourable, in education the fate of individual students is generally held to be an appropriate concern. Further, while in agriculture it was assumed that the expense of differentiating treatments between individuals could only be justified in clinical cases on expensive animals, in education it was widely believed that the treatment of students should be a differential response to diagnostic assessments of their needs.

Experiments concerned to guide the choice of curricula or teaching methods threw into relief the limitations of an approach dependent on sampling and the application of the statistics of probability.

At first the reactions to the exposure of these limitations treated the problem as technical. This is true of the classic papers by Campbell and Stanley (1963), Bracht and Glass (1968) and Snow (1974). To an extent it is residually true of Walker and Schaffarzick's (1974) consideration of the criteria of yield.

However, it became clear that the problem went deeper than this. Statistically significant preferences for one treatment as opposed to another generally meant that in a substantial minority of cases - as many as 40% it could be - the treatment which showed better overall was in fact worse. What was sauce for the goose proved not to be sauce for the gander!

From this observation came the pursuit - through mathematical analysis - of trait-treatment - or as Cronbach called them 'aptitude-treatment' interactions. The hope was that it might be possible so to define the properties of individuals and institutions which interacted with the treatment to differentiate the outcomes, that the application of a schedule or test to an institution or individual could tell us immediately what to do in that case.

Thus an anxious child of nine in a formal authoritarian school in Lancashire and a relaxed child of seven in an informal free-disciplined school in the West Riding might require different treatment and yet by feeding in data about all the interacting variables we should be able to prescribe at least the outlines of the two differentiated treatments. And note that it is the researcher who offers that prescription without depending on the judgement of the teachers involved in the context and with the children in question.

Of course, something can be said by way of such guidance, but the hope that this might come close to a recommendatory prescription has not been justified, as was signalled by an important paper from Cronbach (1975).

To make refined judgements about what educational action to take in particular cases lodged in particular contexts, we need much more information than can at present be reduced to indices and we need to present our conclusions in a way that feeds the judgement of the actors in the situation, a way that educates them rather than briefs them.

To gather about the subjects of our study evidence sufficiently rich to support the kind of discussion from which judgements can be made as conjectures and then subjected to refutation or confirmation in the light of evidence, this is a far more laborious and extensive task than gathering data on samples by schedules or tests. We are closer to the historian or to the clinician in style. It is as if we were committing ourselves to studying a sample of 1500 children in 100 schools by means of individual tests and clinical histories of each child and detailed historical and sociological studies of each school. We should need, let us say, 100 researchers working in parallel, and even then we should have difficulty in reducing the evidence they gathered to data which was comparable and quantifiable. And the problems of working in such intimate detail with schools and children make any approach to random sampling virtually impossible.

A natural strategy in this situation is to set about the patient cumulation of studies of cases. We need to go back to careful direct observation of educational institutions and processes. This is a laborious business and we must work cumulatively. It is the fate, too, of the archaeologist. Digging takes time. His equivalent of the sampling problem is the decision which site to dig next. And this decision can be made wisely only in the light of the cumulated work of other archaeologists.

I believe that the description of cases and the analytic categorization of samples are complementary and necessary approaches in educational research, and it is high time that the superficial stylistic differences between their proponents were recognized as impediments to good sense in the research community. There may of course be gulfs of value and commitment within that community, but I believe that the important party divisions exist on both sides of the sample study/case study fence.

I do not want on this occasion to give detailed consideration to the methodology of the study of cases - it is an area in which colleagues and I are at present working with the support of an S.S.R.C. grant and, as we do so it becomes clear that others here and elsewhere abroad are engaged in parallel activity. But there are one or two selected points I want to make.

The first is that there is an acute need for attention to be paid to quantitative aspects of case study. It seems to me quite clear that descriptive case studies should not confine themselves to words. What indices might best be gathered to describe a school and to locate it within the population of schools? Some - number of pupils, teachers, library books - are obvious. But what of site value, average travelling distance for sports fixtures, geographical distribution of former pupils, age of textbooks and other possibilities? And what testing and measurement might we do in a case-study, what, for example, is the potential of time series analysis? What would Bennett's data on teacher styles and pupil progress look like if it were organized as 37 classroom case studies rather than as a sample? I very much hope that we shall find attention being given to such problems in the next years.

A second problem is the achievement of an understanding of the relative status of observation and interview in case study, and of what is going on in each.

It seems to me that in talking of observation in educational case study the analogy with participant observation has too readily been drawn. Can one participate in a school, whose roles are functionally defined in the same way that one can participate in a community which is a setting for living in a much broader sense?

The observer who relies on or gives priority to his own perceptions and the interviewer who gives priority to the perceptions of others both gather evidence. I use the term evidence to contrast with the alternative term data. Data are standardized - and attenuated - at the point at which it is gathered in order to make them comparable: evidence is not comparable except by virtue of a critical process. Such a process is comparable with the critical methods of the historian. The point is well made by comparing Wragg's Rediguide on Interviewing (1978), which is concerned with structured interviews whose intention is to gather data (i.e. information made as comparable as it can be by the process through which it is gathered), with Ronald Blyths oral history interviews behind Akenfield (1969) whose intention is to gather evidence (i.e. information whose comparability is later to be established by critical process).

The problem of field research in case study is to gather evidence in such a way as to make it accessible to subsequent critical assessment, to internal and external criticism and to triangulation. Here it seems to me that interview is at present better placed than observation in spite of the sociologist's prejudice in the other direction. This is because historians have over many years developed critical tools for dealing with evidence of the kind provided by the voices of participants. But we have as yet a long way to go in developing parallel critical techniques to discount the biases and distortions which may arise from an observer's attachment - not to self-esteem and self-interest in their common forms - but to a theoretical stance.

Such problems as these point to the need for us to establish conventions for the conduct and reporting of fieldwork in case studies, which attempt to secure a basis for verification and for cumulation. Such conventions should never be thought of as binding: they offer either guidance or a position against which to rebel. They are themselves hypothetical. But they must address the two closely related questions: how can a reader verify a case study? And how can a reader who is a researcher use another's case study as a contribution to his own work? Think back to my analogy with archaeology.

These questions - which have yet to be resolved even in a hypothetical form - seem to me to show great promise as growing points of our thinking. They are also fundamental to the proper supervision of students at doctoral and masters levels. They are entitled to advice about them.

My own view has recently come to be that all fieldwork should yield a case record - of observational fieldnotes or of interview transcripts - which serves as an evidential base to underwrite a descriptive case study. Such case records might be made available either in print or possibly in microfiche,pouched in the cover of a dissertation, and these should be footnoted as an historian footnotes his sources. The reader will follow into the record his doubts or his interests.

It is an exciting possibility that current interest in the careful study of cases might produce a national archive of such case records. If we had such an archive now, we could understand in much greater intimacy and depth the recent history of our schools.

Of course one of the problems is that educational researchers have for the most part become impatient for results. Few relish the discipline of the historian who takes years in archives before he can move from the paper or monograph to the large canvas. In fact, however, the problems are - as any good historian knows - less daunting than they appear to social scientists. The fact is that you spend less time reading theory and more time reading primary sources!

But I fear my prejudices are showing!

Behind all descriptive case study of the kind we are discussing there lies another and extremely serious problem: access to data in terms of the rights of the subjects who are studied. I believe that this should be an easier problem in a professional field like education than in a personal and private field, but a great deal of work needs to be done to make good that belief. In particular perhaps theory and discussion which emerges from such work needs to be fully accessible to teachers. It is absurd that educational research should make itself less accessible to teachers than political history is to politicians, or than, say, the Webb's studies of trade unionism are to trade unionists.

It would be useful to our field if the attempt to assess critically the process of case study in respect of its collection of evidence or data, its problems of verification and cumulation and its responsibility to address educators led to a similar review of the processes of sample study. Recent revelations about Cyril Burt might suggest the need for a check up.

Such monitoring of standards is necessarily in all fields of educational research a responsibility of the research community. It is not one for the enterprise of individuals. At present a group of colleagues and I are seeking a framework within which such a communal enterprise can be undertaken by those who feel some responsibility for the conduct of case study. It is an activity of the research community in which I hope B.E.R.A. will play an important part; for an educational research association exists to support precisely the kind of discussion in the research community that is needed.

REFERENCES:

BLYTH, Ronald (1969) Akenfield London: Allen Lane, Penguin Press.

BRACHT, G.H. & GLASS, G.V. (1968) The external validity of comparative experiments in education and the social sciences. American Educational Research Journal 5, 437-474

CAMPBELL, Donald T. & STANLEY, Julian C. (1963) Experimental and quasi-experimental designs for research on teaching. In Handbook of Research on Teaching edited N.L. Gage. Chicago: Rand McNally

CRONBACH, Lee J. (1975) Beyond the two disciplines of scientific psychology. The American Psychologist

FISHER, Ronald A. (1935) The Design of Experiments Edinburgh Oliver & Boyd.

SNOW, Richard E. (1974) Representative and quasi-representative designs for research on teaching Review of Educational Research 44,3, 265-291

WALKER, Decker F. & SHAFFARZICK, Jon (1974) Comparing curricula Review of Educational Research 44,1, 83-111.

WRAGG, E.C. (1978) Conducting and Analysing Interviews. Rediguide No. 11 Nottingham: University of Nottingham, School of Education.