

## Research in education

NO. 20 NOVEMBER 1978

Methodological problems in measuring teaching effectiveness Rolf Schulmeister	page 1
Participating and observing in curriculum evaluation: an agenda for discussion David Smetherham	11
Teachers' definitions of boys and girls: some consequences David Hartley	23
Contrasting organisational climates—sociology or social psychology? Ray Gallop	37
Task orientation as a predictor of academic attainment for Diploma in Management Studies students John H. Cox	53
Research notes	
Staff-room conversation and the curriculum I. A. Rodger	63
An example from the Banbury Enquiry of a three-dimensional contingency table for the chi-square test K. E. Selkirk	71
Book reviews	77

EDITORS A. D. Edwards, Senior Lecturer in Education, and James Rushton,  
Deputy Director for College Studies, University of Manchester

REVIEWS EDITOR D. H. Hargreaves, Reader in Education, University of  
Manchester

EDITORIAL BOARD Professor W. H. G. Armytage Sheffield University,  
Professor M. Chazan University College of Swansea, T. Christie Manchester  
University, Professor H. Entwistle Concordia University, Professor E. Hoyle  
Bristol University, Professor P. W. Musgrave Monash University,  
Professor Alec Ross Lancaster University, Professor D. Swift Open Uni-  
versity, Professor Philip H. Taylor Birmingham University, Professor W.  
Taylor London University, Professor D. Wright Leicester University

PUBLICITY Mrs A. Griffin

MANCHESTER UNIVERSITY PRESS □ ISSN 0034 5237

## Methodological problems in measuring teaching effectiveness

Rolf Schulmeister  
University of Hamburg

**Summary.** This paper does not argue for more reliability in using traditional evaluation methods, but for abandoning some of the basic starting points of empiricist methods. Objectivity of measurement, in the manner of the natural sciences, cannot be increased by greater effort and precision, because this principle is antagonistic to the laws of social, symbolically mediated interaction. It is argued that comparative evaluation, or teacher assessment, is not important. Our aim should be to develop evaluational methods for innovational settings. Most traditional methods of evaluation assess the individual, not the system, and tend to gather the subjective opinion of the individuals. As a result, conclusions tend to be based on present views in ignorance of alternatives. In consequence it is a rather conservative approach which fixes the status quo of the system and immobilises it against more radical innovations. Much more fruitful would be the abandonment of the attempt at quasi-scientific objectivity and its replacement by on-going evaluation of the complex interactions in the innovation process via 'action research'.

### INTRODUCTION

The question of the effectiveness of different methods of teaching cannot be answered simply by comparing the effects of the methods concerned. This is true in spite of the fact that simple comparisons of one kind or another are very common in research on teaching.

Attempts to form hypotheses or nomological statements concerning the relative effectiveness of instructional methods presuppose that an experimentally controlled domain of objects and events can be clearly demarcated. Thus, in the educational case, a class of phenomena must be determined which are to be subsumed under the concept of successful attainment: such phenomena might include, for example, students' performance on certain tests, their future advancement in general, their relative success in later occupational careers, attainment of cognitive or affective objectives, or the learner's mnemonic, practical, critical, or creative abilities. It may be observed that at present there is no consensus, either on the desirable objectives, or, for that matter, on the criteria defining success in learning. Yet even if such consensus could be taken for granted, we would still have to know the detailed educational history of those whom we expect to per-

form in the specified manner. In other words, what are the preconditions of learning on the learner's part? Does his past experience with learning suffice for certain kinds of learning, or for the level of achievement demanded? Or must we seek first to create prerequisite conditions for learning in the student? Can the students be regarded as being sufficiently motivated towards the task required, or is it necessary to provide for special motivating conditions in the learning situation? Is it possible to speak of the learner, or must we distinguish between different kinds of learners? etc.

### THE DIMENSIONS OF THE PROBLEM

Generally speaking as regards the interaction of the process of learning, the learning situation, and the success in learning, one has to discriminate three dimensions of the problem (Schmalohr):

**Prerequisites on the part of the learner** which constitute his ability to learn and which influence learning positively or negatively, e.g. the students' different socialisation, their individual motivational structure, and their previous learning experiences (Beard and Bligh, 1969). The effects of these factors, presumably, are not just to be aggregated in a simple additive way. Their relationship is not simply a linear one, as has of necessity to be assumed in experimental research; factors also differ in intensity and level (McKeachie, 1967).

**The constituent components of the learning situation or system itself.** These consist of its objectives, content, and form. The persons and organisations involved in its operation, its communication and interactional structures (which are defined by other factors), the 'social' climate or atmosphere, the rewards and sanctions provided by the system, the kinds of processes which are facilitated, and last, but not least, the paradigms of learning, on the basis of which teaching methods are designed (Weinert, 1970).

**The intended and unintended effects of the learning situation on the student's cognitive and affective behaviour and on different kinds and different levels of objectives.** Which type of learning situation is better is a question that can be answered only if the objectives are known, and are known to be valid (Schmalohr; Heckhausen, 1970). This is also true for the additional learning resources offered in a learning situation and their interaction with the objectives (Riedel, 1968).

The problem of teaching effectiveness therefore is a question which has been put in much too simplified a manner; instead we must question the relationship between the prerequisites of learning, the learning system, and the outcome. While learning situations and their success depend on the prerequisites on the part of the student, they themselves determine which motivations and attitudes are stimulated and which kind of objectives actually will be operative. Thus the interdependence between motivation components of the learning system, and theoretical criteria—and not the learning

situation or instructional method alone—determines the observable effectiveness of the teaching/learning process. We must think of the actual learning process in terms of a complex interaction of many different systemic and individual, objective and subjective factors. And this relationship, it can be presumed, is not a linear one.

The extreme complexity of the problem discussed here essentially accounts for the fact that empirical research on the effectiveness of teaching has not, so far, produced general and unambiguous results: 'In empirical research on teaching it is not possible to assess methods of instruction as to their effects in an isolated way' (Roth, 1971). Being confronted with a vast number of interdependent variables, most researchers have tried to find dominant interactions of major variables, but so far only a small number has successfully been controlled, and only in some rare cases. And it may be presumed that progress in this field is not dependent on the number of trials and the addition of single results, but—because of the nature of social phenomena—consists in a progressive consensus relying on normative assumptions, hypothetical questions, heuristic models, or conjectural interpretations. (A similar critique, at a more general level, of educational research has been presented by Kitwood, 1976.)

### PRECONDITIONS FOR RESEARCH

Attempts to formulate general statements about the effectiveness of the teaching/learning process have not only to define the domain of objects and events considered to be relevant, but also to:

- (a) specify the nature of the relationships postulated between the different variables, that is, between teaching methods and criteria of success in learning (e.g. linear regression, causal relation, or nomological-probabilistic relationship);
- (b) specify the conditions which allow interactions of the different variables to be observed at all levels and which put restrictions on the possible generalisability of observations (e.g. temporal and spatial conditions, special features of the learners, preconditions of the social system).

It follows therefore that differences in student performance cannot causally be attributed to different kinds of learning situations, or immediately traced back to differences between the instructional methods used. The converse is also true: different learning situations often do not show significant differences in student performance. If relevant variables are left out of consideration, totally misleading interpretations of results can hardly be avoided, as the following examples show.

Grading may influence student performance; if the grading system is changed, or grading is stopped, student performance may change and this can be accounted for only if the relationship between grading and student attitudes is known, otherwise results at the beginning and at the end of the course, or results from different courses, are no longer comparable (Roth, 1971).

If relevant variables are observed in isolation, the postulated effects usually are not significant. They gain significance only when their interaction with other variables is studied. Indeed, if these circumstances are not properly taken into account, there seems to be no difference in effectiveness between any teaching methods (Meister, 1974)! If the important role the interaction of variables plays in comparative research in education is acknowledged, the difficulty then apparent is to decide when and where to stop involving new variables and analysing new possibilities for interactions between variables. Because there does not exist a reasonable theory of social interaction, including progress in learning, this problem cannot be decided on a general basis. If cognitive aspects are taken to be central in measuring teaching effectiveness, the relevant social and emotional basis of cognitive learning processes are left out of consideration despite the fact that the latter account for the larger part of variance in student performance (Weinert, 1970). Up to the present, experimental designs cannot claim to have a rational reason (i.e. one which is legitimate with respect to the methodical control of the testing situation) to neglect the affective influence on learning. Since, however, most experiments do just this, they are open to the objection that their seeming results are fallacious. The fallacy lies in the delusory assumption that cognitive performance could in a methodically convincing fashion be abstracted from the social and emotional dimensions of the learning process. Research on the correlative relationship between, for example, intelligence and cognitive attainment has produced clear evidence substantiating the influence of non-cognitive conditions on learning. Only very few of these conditions have so far been taken into account in published studies: those which have, have been concerned with 'effects of achievement motivation, of levels of aspiration, of different kinds of anxiety, and the dimension extraversion-introversion' (Weinert, 1970).

Out of a multitude of available studies, the majority can be disregarded on grounds of lack of validity and reliability. This is especially true of research which does not even attempt to find out about effects of teaching systems independent of opinions and values of their respondent. Interviews and interrogations may yield valuable clues, but in the last resort they should be regarded rather sceptically. For instance, students in the study by Joyce and Weatherall (1959) maintained that lectures were the most useful method of teaching, whereas other interrogations resulted in opposite evidence according to the subjective views of the questioned subjects (see Beard and Bligh, 1971, 26). In most cases, enquiry techniques do not consider the context-dependency of subjectively biased statements; interrogated subjects are likely to be ignorant about the range of possible alternatives, a fact that severely restricts the generalisability of their statements. Finally, the design of questionnaires induces expectations on part of the interviewed subject which, of course, interact with the other variables in the interviewing situation; the resulting distortions are often not even recognised as such.

More reliable than enquiry techniques are studies testing for effectiveness of teaching/learning situations by means of learner-independent measures.

But even here we may object that these studies restrict the number and nature of possible results, owing to the selection of variables, which becomes necessary for methodological reasons. The selection, being influenced by normative and implicit assumptions, often leads to results which cannot be accepted. Thus, for example, a comparative evaluation of a didactic and of a small-group-based learning situation, solely based on a cognitive test, will obtain but a small part of the totality of positive effects and will necessarily ignore those effects produced by social learning in small groups. Thus the evaluative part of the experimental design may be responsible when the outcome is non-significant, or unexpected. It is with this point in mind that Nachmann and Opochenky (1958) put the following question: 'Do different methods of instruction, which in many studies produce nonsignificant differences in the amount of knowledge gained, actually produce different effects which only remain concealed as a consequence of the kind of testing instrument used?' (translated from the German, Meister, 128). The experimental design of comparative studies cannot be exempted from criticism: in most cases the empirical basis is too narrow, and researchers tend to overlook the fact that experimental conditions themselves determine what may be observed at all within the experiment. They suffer from over-generalisation of the domain for which given statements are presumed to be valid, because they do not reflect the self-made experimental barriers.

The points criticised can be illustrated by the otherwise extremely elaborate methodical design of Meister (1974); he compares seminar courses and group studies, giving both as subject matter the same specially prepared text (cognitive objectives and well selected content). One may suppose that in this way the seminar courses are very much facilitated, but he does not prepare the group study activities in a special way, e.g. by supportive group dynamic training. Thus the experimental conditions are defined in a one-sided manner, which produces effects running counter to the presupposed and seemingly fair hypothesis that both forms of teaching/learning situations have the same likelihood of success. Moreover, both situations are evaluated by the same testing instrument developed on the cognitive basis of the prepared seminar text. Thus Meister fails to obtain significant differences and risks tautological interpretations.

This example shows also that verbal descriptions of different instructional methods do not necessarily denote the same thing in various situations: a well prepared lecture which deals with specific well chosen subject matter is quite different from an ordinary lecture on a general subject, delivered without any special preparation; any given group formed by chance is perhaps more different from a study group which is led by a trained tutor than it is from a lecture, especially if the results are measured in terms of cognitive objectives on lower taxonomic levels. If those situations are not prepared according to their characteristic features, large intra-treatment variance may result, e.g. 'group study' and 'group study' may show larger differences between them than between 'group study' and a 'lecture' (McKeachie, 1970).

In order to compare different forms of instruction, identical conditions have to be secured if we are to follow up the standards set by empiricist research ideals. The paradox of empirical research in education lies precisely in the fact that the special effects of a particular learning situation cannot be proved under such circumstances at all. For example, in comparing small-group teaching with a lecture presented in an optimal fashion, if the lecture is measured according to the standards of group teaching, this would mean modifying the lecture contrary to its normal features; conversely, if small-group teaching is presented in a fashion similar to the lecture, this in turn annihilates the predicted advantages of small-group teaching as a social learning activity.

As regards comparative experiments, a compromise must be reached between two somewhat conflicting conditions:

- (a) all instructional methods must be prepared, presented and evaluated in a way that is appropriate to their particular character, so that it should be possible to realise their specific objectives and advantages;
- (b) objectives and subject matter must, to some extent, be kept constant in various situations in order to ensure comparability.

It is obviously not possible strictly to ensure both conditions, i.e. optimal conditions for each method, and identity of objectives and subject matter (Roth, 1971). It should now be clear that, because the conditions are mutually conflicting, linear induction of general laws is impossible. Instead, we must try to gain insightful consensus into the relative advantages of learning situations on a general level. There is no sense in excluding normative criteria by trying to reach complete experimental control, which, if possible at all, will lead only to unique and non-comparable, and therefore uninterpretable and meaningless, results. Experimental studies are dominated by the belief that it is possible to exclude prejudices and normative assumptions: this obviously is an aim which cannot be reached by experimental control, but which is reasonably the case in a communicative discussion about observations on the one hand and objectives on the other.

So far we have discussed methodological criticism directed at empirical research in educational settings. Nevertheless, we do not share the view that the main deficiency of this branch of research is its lack of strict methodological implementation (compared to its strict methodological ideal), as does, for instance, Weinert (1970). Nor is the decisive factor which constitutes the difficulties of empirical educational research, the so-called Hawthorne effect, as is suggested by Schmalohr. The Hawthorne effect, which cannot be controlled perfectly by experiments in the field of social interaction, is just a resultant of the complexity of the underlying phenomenon, i.e. the non-linear interaction of social events which are mediated by communicative behaviour, regulated by intentions and interpretations, and governed by norms and judgements instead of by natural laws. Therefore the question arises: is there any point in insisting on 'scientific' methodological strictness in the face of objects and events which do not conform to the laws of natural science, or to the rules of

stimulated and repetitive behaviour, but follow instead the dialectics of intentional action? Before answering this question, we shall still try to pursue the methodological criticism of empirical research in education at a more general level.

The complexity of social relations forces the researcher, who is fixing the experimental preconditions on the limits of applicability of his work in order to isolate the effect of a defined variable, steadily to increase the number of his independent variables. 'One might speak of a methodically induced, and demanded, progressive increase in the subdivision of experimental preconditions' (Holzkamp, 1972, 20).

In the extreme case, control of the experimental variables results in an artificial situation, the characteristics of which may no longer be transferred back to the typical real-life situation. Thus insistence on methodological strictness leads to meaningless results. If we cut short the process of successive control of experimental conditions at an earlier stage we shall get non-comparable results, since not all the variables are equivalent, not all 'contaminating factors' are excluded, etc.

The tacit assumption prevalent in empirical research on teaching, that through accumulation of experimental studies an inductive approximation to a theory of teaching and learning could be developed, does not in fact stand on solid ground, since the necessary precondition, as in the natural sciences, viz. the invariance of underlying experimental conditions, is not fulfilled in the social sciences.

A glance in the direction of general learning theories suffices to confirm that these theories do not utilise inductive generalisations built upon empirical data, gathered in a multitude of experimental investigations. Instead, theoretical paradigms and abstract models are put to work, simply illustrated by experimental data which fit the prior theoretical assumptions. The paradox is that research in learning and developmental psychology strive to keep up the image of an empirical science, but in reality are forced to fall back on prior theories. As a result, the rational relationship between theory and empirical data remains undiscussed. When general laws of learning are presented it is obvious that empirical data must be provided as supportive evidence. To formulate general insights and principles concerning teaching and learning processes requires hermeneutic reflections on the conceptions and unspoken assumptions of both participants and observers alike. Hardly any measurements of the effectiveness of different teaching methods rate highly in this respect.

Yet, in this usual research process, the researcher may find methodical procedure helpful to the extent that it makes him aware of, and at the same time simplifies, the relationships in reality. He may find especially helpful those methodical procedures which, although possessing less reliability in the sense of the empiricist paradigm, nevertheless exhibit greater affinity to the domain of objects and events called 'social action'. For instance, depth interviewing and participant observation may be less formalised and

reliable, but they allow follow-up of the processes in which teaching and learning effects develop.

Against this kind of methodology there are objections. The major one is that reliability of measurement is reduced, owing to the interdependency of experimenter and evaluator, and the informality of the relationship. Yet, over the last few years, it seems to have been widely accepted that procedures which yield valid information should be preferred to 'reliable', but contentless, methods. This is accepted because the interrelationship between innovation, changes in the normative attitudes of all parties concerned (researchers and subjects), and the outcome of the learning experiment is such that the only meaningful procedure by which we can gain insight into the effects of social innovations in the area of teaching and learning is by an on-going evaluation of the interrelationship itself. Such a concept of evaluation is familiar to many sociologists under the general heading of 'action research'.

#### ACKNOWLEDGEMENT

The author wishes to acknowledge the considerable assistance and advice of Dr M. G. Cornwall, of Brighton Polytechnic, in rewriting and correcting the author's original version of this paper.

#### REFERENCES

- Beard, M., and Bligh, O. (1971), Research into Teaching Methods in Higher Education. London: Society for Research into Higher Education.
- Heckhausen, H. (1970), 'Zur Bedeutung moderner Lernmotivationsforschung für die Hochschuldidaktik', in H. v. Hentig (ed.), Wissenschaftsdidaktik. Die neue Sammlung.
- Holzkamp, K. (1972), Kritische Psychologie. Vorbereitende Arbeiten. Frankfurt am Main.
- Joyce, C. R. B., and Weatherall, M. (1959), 'Effective use of teaching time'. The Lancet.
- Kitwood, T. (1976), 'Educational research and its standing as science'. Studies in Higher Education, I. 1.
- McKeachie, W. (1967), 'Research on teaching at the college and university level', in N. L. Gage (ed.), Handbook of Research on Teaching. Chicago.
- (1970), 'Motivation, Lehrmethoden und Lernen in Hochschulen', in F. Weinert (ed.), Pädagogische Psychologie. Fifth edition, Cologne and Berlin, pp. 159-87.
- Meister, H. (1974), Lehrmethoden, Lernerfolge und Lernvoraussetzungen bei Studenten. Dusseldorf.
- Nachmann, M., and Opochinsky, S. (1958), 'The effects of different teaching methods'. J. Educ. Psychol., XLIX, 245-9.
- Riedel, K. (1968), 'Lehrverfahren', in H. Roth (ed.), Begabung und Lernen. Stuttgart.

- Roth, L. (1971), Effektivität von Unterrichtsmethoden. Empirische Untersuchungen zu Wirkungen der Organisationsformen von Lernbedingungen. Hanover.
- Schmalor, E. (-), Die Wirksamkeit akademischer Lehrveranstaltungen. Bericht über amerikanischer Untersuchungen in Materialien des AHD. Hamburg, pp. 1-26.
- Weinert, F. (1970), 'Analysis and investigation of teaching methods', in Handbook of Research on Teaching, ed. N. L. Gage; in Ingenkamp, K. (ed.), Handbuch der Unterrichtsforschung, II. Weinheim, Berlin and Basel.