

BARRY MCGAW

2. PROSPECTS FOR THE EMPIRICAL TRADITION IN EDUCATIONAL RESEARCH

Presidential Address to AARE, Brisbane, 13 November 1976

The Australian Association for Research in Education, in the time since its founding conference in 1970, has become the principal organisation of educational researchers in Australia. The last few years, in particular, have seen its membership and attendances at its annual conference grow substantially. Yet it has remained an organisation somewhat dominated by those educational researchers whose roots are in the empirical research tradition. The majority of papers read at its conferences reflect this tradition. Five of its first six presidents would identify with this tradition.

This 1976 conference, through its thematic emphasis on the implications of research methods for practice, has provided an explicit invitation for all to display their methodological wares and to point to their substantive fruits. Although such invitations to reflection often produce protracted sessions of 'navel contemplation' in which a succession of speakers bewails the poverty of present research results, the inadequacy of present research methodology or the complexity of educational phenomena, the hope that many had for this conference was that it would provide the grounds for a constructive, and broadly based, review of potential contributions and developments. Whether the conference achieves this will obviously be for the participants to judge, but the detailed program at least gave promise of it.

In this address, I make no pretence of judging the contributions of traditions other than the one I know myself. What I have chosen to offer is an analysis of the prospects for the empirical tradition in educational research. I do this, not to re-affirm the dominance of this tradition within our organisation, but to facilitate a review of it by all members, whatever their own disciplinary or methodological affiliation might be. In fact, my objective is even more limited than this might suggest. The particular empirical research tradition on which I want to focus is that of educational psychology. My intention is to review some of the major shifts in emphasis in that tradition, and to offer an analysis of its future prospects.

THE ORIGINAL DREAMS

The original dreams of educational psychologists were grand enough! I can distinctly recall how, while studying chemistry and psychology as an undergraduate, I was

impressed by the grandness of the theoretical vision reflected in Hilgard's (1956) text on theories of learning. Hilgard documented the increasing complexity and completeness of the theoretical formulations in this area of psychology in a way that seemed to give promise of matching, in kind, the formulations that I was encountering in the physical sciences. As one having his first experience of the behavioural sciences, I was impressed by the prospect of elaborate theories with which to account for the phenomena of human behaviour and the prospect of applications in the field of educational practice to which I was already committed (both by inclination and by contractual obligation). The logical positivism of the physical sciences was being clearly expressed as a methodology of the behavioural sciences.

My hope, as a fresh young undergraduate in 1960, was certainly consistent with the vision of the early educational psychologists but, as I was later to discover, by 1960 it was already outdated. My subsequent more detailed experiences with research on learning from prose materials (McGaw & Grotelueschen, 1972) or 'connected discourse' as it is called by those who wish to make their activities seem more esoteric, and with research on cognitive processing (Anderson & McGaw, 1973), have made it clear how much more modest the vision of learning theorists is now than it was when Hull was erecting his model (Hilgard, 1956, pp. 121-184). The information processing model that Professor Gagne offered this morning did not seem to me to be based on a claim for comprehensiveness of the type that characterised these earlier models.

After the original investigations of William James (1890), and E.L. Thorndike and Woodworth's subsequent series of investigations to determine the bases on which the effects of one learning activity might transfer beneficially to another, Thorndike (1906) summarised the results in a book entitled *Principles of teaching*. In 1913, he published his three-volume *Educational psychology* in which he argued that the function of educational psychology was to provide 'knowledge of the original nature of man and of the laws of modifiability of learning, in the case of intellect, character and skill' (1913, vol. 1, p. 1).

Much of the original *Handbook of research on teaching* (Gage, 1963) reflects a similar view of the manner in which empirical research could be expected to provide an understanding of educational processes and, ultimately, a basis for more systematic and effective educational practice.

There have, of course, been few who would claim that the grand vision had yet been fulfilled. Cronbach (1957), in his presidential address to the American Psychological Association, attributed the failure to fulfil the early theoretical promise, both in educational psychology and more generally in psychology as a whole, to inadequate methodology. In particular, he pointed to the artificial distinctions that had been established between the experimental and the correlational traditions within psychological research and showed how the separation of these two traditions had restricted the theoretical vision of those within each.

For the experimenters, he said, individual differences had become only a source of annoyance, to be controlled or, at least, minimised. For them, it was the environmental characteristics, the treatments, that must be manipulated so that they

could understand how the treatments operate, on the average. For the correlationists, individual differences had become the stuff of life. The treatment conditions had to be stabilised, or the effects of their consistencies statistically removed, so that they could understand the nature and the extent of individual differences. The experimentalist was offering to education the prospect of re-shaping the environment to fit best the average child. The correlationist was offering to education the prospect of better identification and selection of individuals for existing environmental conditions.

In proposing a synthesis of these traditions through a focus on aptitude-treatment interactions, Cronbach sought to integrate the experimenter's attention to manipulations of treatment with the correlationist's attention to the multi-variability of responses. For education, the prospect of this broadened research would be an ability to determine, for each child, the best situation and, for each situation, the most appropriate children. Thus the concept of aptitude-treatment interactions entered the literature and became the focus of a good deal of attention.

Cronbach's (1957) analysis was in no sense a retreat from the ultimate vision of theoretical formulations with which to understand human behaviour and on which to base educational practice. His paper was a significant attempt to identify more precisely the crucial theoretical questions to be addressed and to propose the methodology with which to answer them all within a logical-positivist view of reality. The answers to be obtained from investigations of aptitude-treatment interactions were still expected to provide laws, conceived in positivist terms, and conforming to the pattern of laws which have been developed in the natural sciences.

THE RETREAT TO SMALLER VISIONS

A substantial effort has been invested, since 1957, in response to Cronbach's call for investigations of aptitude-treatment interactions. At the same time, however, there have been two other responses to the evidence that empirical research in educational psychology had failed to have any substantial impact on educational practice. Both developments appear to represent a retreat from the grand visions of pervasive theoretical formulations and theoretically based practice. One reflects a withdrawal from any short-term concerns about practice and a focus on limited theoretical formulations. The other reflects a withdrawal from concerns about theory and a focus on the task of directly improving practice. The first approach is not immediately practical; the second is atheoretical.

In some ways, the first approach represents little change from the situation that Cronbach criticised in 1957. It is characterised by research in which the scope is consciously limited, with a few conditions being controlled and the rest being ignored. It continues to be a search for main effects, for treatments that work for all individuals under certain restricted conditions.

The research on learning from prose materials, commenced by Rothkopf (1966) at Bell Laboratories, can serve as a useful illustrative example. Rothkopf's attention was restricted to learning from prose and, more particularly, to the effects

on learning of questions inserted in the prose at intervals. Although he interpreted his results in terms of the general principles of operant conditioning, his purpose was not the construction of an elaborate theory of learning. His purpose was to understand the effects on learning of additions to routine prose materials that might control the attention of readers. A large number of studies has followed from Rothkopf's original work, many of them done by him and his associates at Bell Laboratories. Each study has sought to elaborate a small aspect of the limited theory. There is no expectation of grand theoretical propositions flowing from these investigations. At best, they are hoped to provide a piece in the jigsaw. There is certainly no expectation of any short-term impact on educational practice. Lest what I say may seem to denigrate these researchers, let me acknowledge that I have myself contributed to this line of research (McGaw & Grotelueschen, 1972).

In a seminar with graduate students at the 1971 conference of the American Educational Research Association, Rothkopf provided an intriguing insight into his motivation in developing this line of research. He described a conversation with a physicist at Bell Laboratories in which, when Rothkopf complained about the complexity of the psychological phenomena with which he had to deal, the physicist claimed to be no better off. The physicist pointed to the impossibility of predicting which way a rain drop would splash when it hit a roadway but to the success with which physical behaviour in vacuum tubes can be predicted. The secret, he said, was to work with contrived and idealised situations. Rothkopf's selection of prose material at his vacuum tube may have made this an unfortunate analogy!

Note the contrast between Cronbach's exhortations and Rothkopf's practice. Whereas Cronbach (1957) was arguing for the simultaneous consideration of multiple treatment conditions and multiple measures of individual differences to accommodate, within a single research design, as much of the complexity of the natural phenomena as possible, Rothkopf was seeking more contrived situations within which to simulate idealised conditions. But both anticipated the ultimate development of positivistic laws of human behaviour.

To put that another way, Cronbach was seeking more complex versions of the traditional empirical research paradigms, whereas Rothkopf and many others were seeking more controlled situations within which to apply these traditional paradigms. Each anticipated, however, from their research endeavours, the production of general laws of human behaviour.

Still others, however, reject the traditional paradigms as impotent or even inappropriate. Stake (1967), analysing the tasks involved in the evaluation of an educational activity, proposed the collection of a much broader range of data than is usually gathered in empirical studies, while at the same time acknowledging that his strategy could not be expected to yield enduring generalisations of the type that the empirical research methods were supposed to yield.

EDUCATIONAL EVALUATION AS ATHEORETICAL PRAGMATISM

The reactions of Stake (1967) and others in the late sixties to the traditional empirical methods seemed to stem clearly from disenchantment with the products of the methods. On the one hand, the theory yielded by the rigorous but restricted research of those working within self-imposed constraints seemed to offer theoretical formulations, but only ones which were trivial in the face of the complexities of real educational activities. On the other hand, the more elaborate methodology and the broader focus of studies of aptitude-treatment interaction seemed to have been little more successful in establishing a theoretical base on which to deal with educational problems. What was proposed was a complete rejection of the view of behavioural research underlying both approaches. The range of legitimate data was to be potentially unlimited, but its interpretation was to depend, not upon some arbitrarily determined rules of statistical inference based upon a set of assumptions that could not be satisfied, but upon the insight and judgement of the evaluator.

This approach has won a substantial number of supporters. A new division of the American Educational Research Association has been formed to accommodate it. It breeds a kind of pragmatism that focuses directly upon educational practice and seeks a basis for the further development of that practice. It makes no pretence at hypothesis formulation and no claims for generalisability of the findings. Improvements to practice in one place may stimulate improvements in another but only to the extent that one person's apparently good idea is accepted by another who chooses to capitalise on it by amending his own practice. There is no expectation of laws that describe enduring relationships emerging from these sorts of investigations.

All this is not to imply that evaluations produce little information. In fact, they produce much more information than do traditional applications of the empirical research strategies. A major problem is, of course, to establish bases for evaluating the information provided through the multiple perspectives used. The evaluator seeks a fullness of description that goes well beyond the data provided by the typical controlled experiment which seeks to determine whether some alternative treatments are consistently discriminable from one another in terms of some particular outcome measures. Evaluations seek to describe, or portray (Stake, 1975), the prior conditions of an educational program, the detailed processes that occur in its operation and its outcome, both intended and unintended (Scriven, 1973). Some approaches are more concerned than others with formal procedures for establishing relations among conditions within the particular program. Whereas Stufflebeam *et al.* (1971) sought to describe contextual and input conditions and processes and outcomes, Stake (1967) proposed a search for contingencies among such variables. All approaches draw upon many perceptions of the program. None places the evaluator in the role of the empirical researcher gathering his 'facts'.

Those writers who demanded these much fuller descriptions, much more detailed observations and much more use of insight and judgement unfettered by statistical decision-rules have demanded more but, in some respects, offered less. Their claims are less pretentious if judged in terms of what they see to be the ultimate products of their enterprise. They have no vision of powerful, explanatory

theoretical models towards which to aim. The focus of their attention is in developing a more pragmatic strategy for improving (or changing) current practices or even just judging current practices. Their models are models of evaluation not models of education. In fact, they are less even than that. They are really just strategies for evaluation, offering general approaches to the evaluation task often without much detailed direction for any would-be evaluator. The details of particular evaluations are held to be situation specific in a way that allows of no description of a generalisable and detailed methodology.

In contrast with the traditional commitment, in educational research, to the development of theory, the evaluators have abandoned the theoretical vision not because it seemed too difficult but because it seemed unhelpful in the face of pressing problems of practice. It is not clear from much of their writing whether this reflects an expedient view that theoretical models are too remote to be worth pursuing at this stage, or a more theoretical stance in which such models are held to be ultimately unattainable or inappropriate. Whichever is the case, there is a general rejection by evaluators of strategies that claim to accumulate wisdom in the form of generalisable formulations of theoretical relationships among educational variables.

THE FUTURE PROSPECTS

In the last 10–15 years it seems to me that educational psychology has been moving in the three fundamentally different directions that I have outlined. Some researchers have employed the more powerful methodology to which Cronbach pointed in the hope of building broad theories, others have concentrated on more restricted aspects of theory in the expectation of subsequent integration with other components, while a third group has abandoned any focus on theory in favour of a more immediate atheoretical attack on the problems of practice.

We are now, in the mid-seventies, in a position to make some judgements about the relative successes of these approaches.

With respect to the research on aptitude–treatment interactions a good deal of evidence is now in and has been extensively reviewed (e.g. Bracht, 1970; Cronbach & Snow, 1969). Cronbach (1975), in his address on receipt of the Distinguished Scientific Contribution Award at the 1974 conference of the American Psychological Association, reflected on the present state of the research activity he had stimulated seventeen years earlier and reached some intriguing conclusions. In his earlier analysis he had said that the results of the relatively independent traditions of research on the effects of treatment differences, on the one hand, and the nature of individual differences, on the other, were confused because they failed to take account of the interactions between individuals and treatments. In his more recent reflection, he pointed to some studies that show clear-cut interactions of the type predicted.

For example, experimental comparisons of didactic teaching methods and problem-solving methods, which had been inconsistent in the earlier research, had

now shown the problem-solving methods to be superior with a middle-class clientele but the didactic methods to be superior with lower-class children.

Cronbach (1975) also pointed to inconsistencies in the studies of such interaction, inconsistencies that he suggested are themselves due to higher order interactions. To illustrate this with a hypothetical example, suppose that the teaching method by social class interaction just described, that is, the differential effectiveness of teaching methods with different social classes, actually interacted with the content of the instruction. For example, the superiority of didactic methods with lower class children may appear with science instruction but not social science. Investigators unaware of this crucial content variable would be faced only with the inconsistency with which they have been interacting between teaching method and social class. The problems are how to anticipate the factors that will cause such higher-order interactions and how to collect enough data with which to study them.

The position that Cronbach took on this in 1975 is intriguing. Remember that, in 1957, when confronted with inconsistencies in the results of studies of overall treatment effects, he proposed a new set of questions about interactions and a more powerful methodology with which to deal with them. In 1975, with evidence that the important questions are even more complex, he proposed no further stepping up of the methodology. Instead, he proposed a scaling down of our expectations of educational research.

Because of the problems of anticipating potential sources of higher order interactions, Cronbach argued for a careful attention to the detailed conditions of each study and a complete description of them. Exceptions to any generalisation could then provide the clue to the sources of higher order interactions. But how are these exceptions to be fitted into any pattern? Once a potentially interacting factor has been identified, it is possible to design a study in which the higher order interactions can be investigated empirically but there are two problems. Firstly, there is no basis for estimating the likely significance of an interaction from an observed exceptional case and so no way of knowing how much value there may be in pursuing it. Secondly, the amount of data required to study the interaction may be prohibitive. What we need is a more effective way of synthesising the results that suggest the need for this type of further investigation.

One potentially powerful approach has recently been developed. In his presidential address at the American Educational Research Association earlier this year, Gene Glass (1976) outlined an empirical approach to what he called meta-analysis. Meta-analysis, in contrast with primary analysis of original data or secondary analysis of existing data, is an analysis of analyses. Typically, meta-analysis has been done by literature review. It often proceeds like literary criticism. It occasionally takes a more empirical form when the number of related studies with significant and insignificant results are counted to see which outcome is dominant. Such a voting procedure usually reveals some studies for and some against and produces a general feeling of despair about the whole enterprise of empirical research in education. Glass suggested that these approaches to synthesis are so inadequate that we end up 'knowing less than we have proven' (1976, p. 46)

because we fail to put it all together, so to speak. What Glass proposed is a more thoroughgoing empirical approach to synthesis. He described how related studies could be assigned numerical descriptors that indicate both design characteristics and the magnitude of the effects revealed in their data. Relatively complex analyses of variance could then reveal the strength of any overall conclusions as well as the presence of interactions of the type that Cronbach felt could not be studied directly. Glass's (1976) procedure seems to me to hold promise of carrying us beyond an impasse that Cronbach (1975) revealed.

To illustrate his approach Glass took 1000 studies from the confused literature on the outcomes of psychotherapy and counselling (much of which literature suggests that 'nothing' is as good as 'counselling', that is that counselling is no better than nothing). His complete analysis, based on 375 studies, quantifies differences between types of counselling in ways that none of the individual studies attempted. It is an elegant empirical synthesis of the confusing results of a large number of empirical studies, which finally offers the therapists grounds for confidence in their activities denied them by the classical literature reviewers. It estimates the magnitude of the superiority of therapy overall and provides empirical comparisons of the efficacy of different approaches to therapy.

Now where does that leave us? Can we find in Glass's empirical approach a procedure for confronting the questions that Cronbach thought were answerable? In part, I believe, we can. But Cronbach pointed up another problem that may, indeed, be more serious. He suggested that time itself may be an important factor with any of our data. Results obtained at one time may simply not recur at another time because of changed circumstances. Our conclusions may be less enduring than we have hoped for, our laws less stable. Indeed, it may be inappropriate to think of them as laws. Cronbach himself in 1975 lowered his sights to a short-term empiricism, a set of relationships among explanatory concepts that are generalisations only in the sense that they serve as working hypotheses at a particular time. He has rejected the model of the physical sciences, which, he believes, has 'fixed our eyes on an inappropriate goal' (1975, p. 126).

We would do well to be chastened by Cronbach's experiences and to be more modest in our expectations. But we may be able to go further than he suggests. We may not break free from our time-bound generalisations but, with Glass's (1976) meta-analysis, we may be able to broaden the basis of our generalisations within that frame.

That then appears to be where the research on aptitude-treatment interactions has taken us and where it might lead, but what of the research directed to more limited theories and the pragmatic attention to practice? The controlled studies of limited psychological phenomena may build cumulatively on one another in a way that research on more complex aspects of educational practice cannot. They may give rise to a better understanding of particular aspects of human behaviour but I have no expectation of a thoroughgoing science of human behaviour and I see no evidence of an empirically based educational theory constructed upon what psychology presently offers or will offer. To believe that educational phenomena can be understood in terms of psychological phenomena alone is to indulge in a

reductionism that does no justice to the educational phenomena. To understand learning or development, or even their interaction, for example, is not to understand teaching. It is with respect to the psychological phenomena themselves that Cronbach (1975) no longer hopes for more than short-run empiricism. With respect to educational phenomena, it is unrealistic to hope for more!

Kallos and Lundgren (1975) argue that education ought to be considered lawful only in the sense of being rule-governed. They suggest that we should seek to establish the contextual conditions in which participants opt for certain sets of rules rather than others. We ought not to presume that there are underlying laws, they saw, expressed in terms of hypothetical, psychological constructs, with which to account for teaching behaviour, but only that there are external system constraints that influence conscious decisions made by teachers. Without necessarily going as far as this in rejecting completely the potential value of positivistic laws, it does seem clear that we need to distinguish educational research from psychological research and that we need to recognise the limits within which our methodology can operate, indeed, the limits imposed by our methodology.

As for the atheoretical attention that has been given to practice, there seems to be no basis for teaching general conclusions about its value. Scriven (1969) and Stufflebeam (1975) have discussed, in some detail, a basis for meta-evaluation, but I know of no documentation that reports the effectiveness of various approaches to evaluation in influencing practice. The commitment to evaluation seems still to rest on a general hope that it will achieve what research has not. Whether revealing to participants in an educational activity the perceptions of other participants, and perhaps even their own, will lead to changes judged by any or all of them to be of value is to my mind completely unknown.

THE ULTIMATE GOAL AND THE IMMEDIATE TASKS

Presumably, there is unanimity about the ultimate goal of empirical research in education being the improvement of educational practice. But what are its prospects? It seems to me that fundamental psychological research should not be conducted in the name of educational research and that whether it yields enduring or, at least, useful generalisations about psychological phenomena ought to be the only criterion by which it should be judged. Those pursuing atomistic psychological research with this in view, however, should consider seriously the caution issued by Cronbach (1975) before, or as, they press relentlessly on in their pursuit of models fashioned on those of the natural sciences.

For those who are concerned with educational practice, I believe a legitimate case can be made for several courses of action. Some may choose only one; others may choose simultaneous advances along more than one for their own professional satisfaction. The remoteness of much of the more theoretical research from practice and the pressing nature of many practical problems justifies, in my view, serious attention to the empirical evaluation of practice, but I harbour a concern that it offers no ultimate promise of progress. Unless the wisdom from isolated experiences can be accumulated each new set of practical decisions will be made

from as much ignorance as the previous. Empirical investigations of the influence of contextual constraints on conscious choices take us a step further by helping those educators in a position to manipulate the context but, unless these educators know which choices are to be preferred, they in turn will be manipulating in ignorance. In that context, I mean choices that could be made on the basis of their utility in achieving goals, not normative choices that must be made among goals.

As much as this research on the effects of constraints on choices, we need research on the effects of choices. We do not need to achieve that our conclusions will be immutable over time to justify such research. What we need is sufficient fidelity to the complexity of the context for us to be able to isolate the conditions crucial to the effect. We need not demand that every research study control or systematically vary the universe of conditions but we should demand that each documents the conditions in which it was conducted. With this sort of detail, we can apply the insight and judgement of which Cronbach (1975) spoke as well as providing the basis for empirical syntheses of the type that Glass (1976) provided.

I believe that there are regularities in the phenomena we study. Too often we seem to see only the complexity and retreat either to contrived situations from which we hope to have removed the complexity or abandon altogether our faith in the regularity and pursue the specific as though they were nothing more. And if all that sounds like an attempt to legitimate what I have been doing for the last eight years, it probably is!

REFERENCES

- Anderson, R. C., & McGaw, B. (1973). On the representation of the meaning of general terms. *Journal of Experimental Psychology*, *101*, 301-306.
- Bracht, G. H. (1970). Experimental factors related to aptitude treatment interactions. *Review of Educational Research*, *40*, 627-645.
- Cronbach, L. J. (1957). The two disciplines of scientific psychology. *American Psychologist*, *12*, 671-684.
- Cronbach, L. J. (1975). Beyond the two disciplines of scientific psychology. *American Psychologist*, *30*(2), 116-127.
- Cronbach, L. J., & Snow, R. E. (1969). *Final report: Individual differences in learning ability as a function of instructional variables*. Stanford, CA: Stanford University.
- Gage, N. L. (Ed.). (1963). *Handbook of research on teaching*. Chicago: Rand McNally.
- Glass, G. V. (1976). *Primary, secondary and meta-analysis of research*. Paper presented at the annual conference of the American Educational Research Association, San Francisco, April.
- Hilgard, E. R. (1956). *Theories of learning* (2nd ed.). New York: Appleton-Century-Crofts.
- James, W. (1890). *The principles of psychology*. New York: Holt, Rinehard and Winston.
- Kallos, D., & Lundgren, U. P. (1975). Educational psychology: Its scope and limits. *British Journal of Educational Psychology*, *45*, 111-121.
- McGaw, B., & Grotelueschen, A. G. (1972). The direction of the effect of questions in prose material. *Journal of Educational Psychology*, *63*, 580-588.
- Rothkopf, E. Z. (1966). Learning from written instructive materials: An exploration of the control of inspection behaviour by test-like events. *American Educational Research Journal*, *3*, 241-249.
- Scriven, M. (1969). An introduction to meta-evaluation. *Educational Product Report*, *2*, 36-38.
- Scriven, M. (1973). Goal-free evaluation. In E. R. House (Ed.), *School evaluation: The politics and process* (pp. 319-328). Berkeley, CA: McCutchan.
- Stake, R. E. (1967). The countenance of educational evaluation. *Teachers College Record*, *68*, 523-540.

- Stake, R. E. (1975). *Program evaluation, particularly responsive evaluation*. Occasional Paper No. 5. Kalamazoo, MI: Evaluation Center, Western Michigan University.
- Stufflebeam, D. L. (1975). *Meta-evaluation*. Occasional Paper No. 3. Kalamazoo, MI: Evaluation Center, Western Michigan University.
- Stufflebeam, D. L., et al. (1971). *Educational evaluation and decision making*. Ithaca, IL: Peacock.
- Thorndike, E. L. (1906). *Principles of teaching*. New York: A.G. Seiler.
- Thorndike, E. L. (1913). *Educational psychology* (Vol. I, II, III). New York: Columbia University Press.