British Educational Research Journal, Vol 11, Issue 1, Pages 45-49, 1985

The Implications of Meta-analysis for Educational Research

C.T. FITZ-GIBBON, University of Newcastle-upon-Tyne
(now *University of Durham*)

ABSTRACT *Meta-analysis, as developed over the last few years by Glass and others, is a quantitative method for synthesising research results. Its use is illustrated here by means of examples ranging from irreproachable to dubious. Being simple to use and easily understood, meta-analysis will undoubtedly become popular and this increasing use may well bring about some notable changes. The well-controlled, small-scale experiment is likely to become more important and better use will be made of many existing research reports, reports which can now be dusted off and incorporated into meta-analyses. Because meta-analysis focuses on how much difference something makes (the magnitude of an effect) and not on whether or not the difference was statistically significant at a pre-specified level, its use encourages a more scientific approach to the interpretation of quantitative results. It also offers some hope that we might eventually have a clearer idea of the conditions under which research findings can be generalised. Progress in this direction will require mutual support between quantitative and qualitative research methods.*

Meta-analysis is a method of synthesising research results. In a meta-analysis the effects of variables are examined in terms of the 'Effect Size', i.e. in terms of *how much* difference they make, rather than only in terms of whether or not the effects are statistically significant at some (frequently arbitrary) level such as 0.05. The kinds of data for which it is most suitable are data arising from true experiments, i.e. from interventions with experimental and control groups, although in their important book Glass, McGaw & Smith (1981) also deal with extracting Effect Sizes from correlational studies.

The emphasis on the size or magnitude of an effect, while fundamental in the physical sciences, represents a very significant shift of emphasis in social science research. A number of possible implications are considered below.

British Educational Research Journal, Vol 11, No 1, 1985, pages 45-49

The Value of Small Experiments if Properly Enhanced

Experimentation as a research method has faced severe difficulties, not the least of which has been the problem of the *scale* of an experiment. Small-scale experiments can be tightly controlled and thoroughly evaluated but, lacking large numbers, they often lack statistical power, carry little weight and have uncertain generalisability. Large-scale experiments, on the other hand, are almost impossible to implement and, since they are difficult to monitor and control, the interpretation of their results is often in doubt.

There is good reason to believe that educational research will need to proceed with considerable reliance on the use of small experiments. One reason for this is an important principle in implementing field experiments or in conducting action research: the experimental programme should be conducted in a way which is as congruent as possible with the way in which the programme would be implemented if it were widely adopted. Since most instruction is delivered in classrooms, to small groups of pupils, field experiments should be run using classrooms, with their small groups of pupils. Thus experiments will essentially be small scale, using small numbers of subjects, not the thousands of survey research. Moreover, since classrooms have their own effects (Greene, 1980; Gustafsson, 1978), it will often be desirable to randomly assign half a class to one method and half to another method. Inevitably this means small numbers and therefore an individual experiment of low power. The use of meta-analysis techniques to synthesise findings from multiple replications allows these small experiments to contribute to the accumulation of knowledge in the social sciences in the way that innumerable small experiments have contributed to knowledge in the physical sciences.

This enhanced value for the small experiment has its own implications:

- Humble, small-scale research reports which have simply been gathering dust may now become useful.

- Small-scale research conducted by individual students and lecturers will be valuable since meta-analysis provides a way of coordinating results drawn from many studies without having to coordinate the studies themselves.

- For historians, a whole new genre of studies is created – the study of how effect sizes vary over time, relating this to historical changes.

None of this is meant to imply that we are about to enter an era of easy, small scale, controlled experiments. There is nothing easy about experiments. Obtaining cooperation for an experiment is nowhere near as easy as obtaining permission to conduct a passive observational study. Nor is it easy to obtain adequate control. And like other research, experiments often throw up more questions than one started with: they generate new hypotheses as well as testing old ones. Indeed Bronfenbrenner drew attention to the idea that experiments should not be seen as the end-of-the-line hypothesis testing stage of research but as an immensely valuable means of *generating* hypotheses. He records, 'Dearborn's dictum', the injunction: "if you want to understand something, try to change it" and continued,

> What I am urging is a reversal of our usual procedure of relegating the contrived experiment to the last stage of the research. Instead, I advocate moving it right up front to the beginning ….. (Bronfenbrenner, 1975).

With the advent of meta-analysis, there is now a better chance that the small experiment will be recognised as making an important contribution to the aggregation of knowledge in the social sciences.

Statistical Significance may Finally be Dethroned

Because, with Effect Sizes, it becomes easier to concentrate on the educational significance of a finding rather than trying to assess its importance by its statistical significance (the unlikelihood of its occurrence under a null hypothesis) we may finally see statistical significance kept in its place as just one of many possible threats to internal validity. Anyone who doubts that this change is long overdue is entreated to read Carver's excellent article, 'The case against statistical significance testing' (Carver, 1978).

Research Councils might Consider a new Procedure for Targeted Research

The argument was advanced in the first of these two articles that the application of meta-analysis to replications (i.e. to experiments testing essentially the same constructs) was irreproachable. When there are important areas of debate with policy implications, it would be an appropriate strategy for research councils to:
(1) Fund a design project which would identify key variables about which information was needed and publish a blueprint for the kinds of experiments which would be needed to elucidate the debate.
(2) Fund multiple replications by making grants to a dozen or so sites all of which would conduct experiments which followed the blueprint but differed in their context variables.
(3) Convene a conference at which the implementations and results would be presented and discussed, thus including the 'thick description' and awareness of complexities which elegant statistical treatments must never overshadow.

Publishing Policies may Change

As meta-analysis becomes more important, editors of journals should assess submissions in the light of their possible usefulness for future meta-analyses. This means that sample size and statistical significance would be less important than clear descriptions of how the variables were operationalised and the adequacy of the design. Editors might also wish to insist that when a short report is published the author undertakes to make a longer, more detailed, more descriptive, report available for subsequent use by meta-analysers.

Another change in publishing policies will be that new types of quantitative information will become *de rigueur* in research reports. It used to be the case that statistical significance was the only required piece of information in a research report. As a result, a considerable portion of the work of those developing MAs has been concerned with trying to retrieve basic information which was not reported, as, for example, retrieving estimates of the control group standard deviation when only a t-test and means had been reported. As the research community becomes aware of the need to make their work accessible for later meta-analyses, fuller details will be required and the reporting of Effect Sizes will become common practice.

Generalisability and Sampling may be better Understood

Glass *et al.* used the term 'study characteristics' for context variables. Whatever one calls them, their selection is as problematic as the choice of independent and dependent variables in primary studies: a non-explicit mixture of tradition, prior evidence and intuition, which probably should be more often related to explicit theories (cf. Hargreaves, 1981; Fitz-Gibbon & Morris, 1975).

Campbell & Stanley (1966, p. 17) noted that generalisability rested essentially on vague notions of the 'stickiness' of nature. Results would generalise better closer to home than

further away, better with a more similar sample than with a more different sample and so on. But similar or different on what variables and what are the variables which make a region or a group homogeneous and generalisable-to? Analyses of context variables via meta-analysis might begin to suggest some general principles of generalisation as well as to direct researchers in particular fields to the particular variables they should specify in describing their sample, treatment, location and personnel.

Any description of procedures, location, personnel, is only a partial description. What kind of information is important, and what is unimportant to the outcomes of the experiment? Bertrand Russell warned that the major characteristics of an epoch may well be the very aspects of life on which nobody comments. Similarly, a crucial aspect of an experiment might be an aspect which seemed too obvious for comment to the experimenter. Hence there is a need for 'thick description' within which might be buried aspects which only later appear important. Qualitative accounts should always accompany quantitative work.

Meta-analysis and Qualitative Research – a new need for mutual support

The value of qualitative accounts, such as ethnographies, for recording context variables has just been mentioned. Another instance of mutual support between qualitative and quantitative work occurs when quantitative analyses support experiential evidence. When we find the experience of teachers, qualitative, anecdotal and descriptive data pointing to the same conclusions as those arrived at by quantitative research we shall have more faith both in the conclusions and in *all* the methods employed to reach those conclusions. To take one example, a meta-analysis has demonstrated peer tutoring to be exceptionally effective (Hartley, 1977), and this finding fits entirely with the experience of many teachers, as documented in books (e.g. Gartner, Kohler and Riessman, 1971; Allen, 1976), articles (Ebersole, 1972; Bond, 1982) and surveys (Fitz-Gibbon, 1977). When Thelen wrote,

> Cross-age tutoring works. I can think of no other innovation which has been so consistently perceived as successful (Thelen, 1969, p. 230),

he was presumably referring to his own perceptions, experiences, anecdotal evidence, etc. That the judgement of this informed impression has now received support from meta-analyses of more than 200 studies (Hartley, 1977; Cohen, Kulik and Kulik, 1980) is an example of the kind of convergence between experiential and experimental data which makes us hopeful that social science is proving cumulative and informative; that social science is indeed a science. It may prove to be an untidy and difficult science, like meteorology or medicine, but this is no excuse for our not trying the scientific method a little more thoroughly.

It may encourage readers who feel grave misgivings about social science *qua* science to know that Glass, the main proponent of meta-analysis, is no thoughtless, naïve or dogmatic positivist. In an article in 1979, for example, he considered the possibly unpredictable nature of educational problems. Of meta-analysis, Glass *et al.* concluded cautiously, "The approach we call meta-analysis seems to be too plainly reasonable to be false in any simple sense. Whether it will be useful is a different matter" (Glass, McGaw & Smith, 1981, p. 231).

Correspondence: Professor C.T. Fitz-Gibbon, Curriculum, Evaluation and Management Centre, Mountjoy Research Centre 4, University of Durham, Durham DH1 3UZ, England.

Note

This is the second of two articles on meta-analysis. The procedures of meta-analysis were described and illustrated in the first article, which also dealt with some commonly raised objections. This article explores implications that meta-analysis may have for the conduct of research in the social sciences.

REFERENCES

Allen, V. (Ed.) (1976) *Children as Teachers: theory and research on tutoring* (London, Academic Press).

Bond, J. (1982) Pupil tutoring: the educational conjuring trick, *Educational Review*, 34(3), pp. 241-252.

Bronfenbrenner, U. (1975) The ecology of human development in retrospect and prospect. Invited address at the final plenary session of the conference on Ecology Factors in Human Development, University of Surrey, Guildford, 13-17July.

Campbell, D.T. & Stanley, J.C. (1966) *Experimental and quasi-experimental designs for research* (Chicago, Rand McNally).

Carver, R.P. (1978) The case against statistical significance testing, *Harvard Educational Review*, 48(3), pp. 378-399.

Cohen, P.A., Kulik, J.A. & Kulik, C.C. (1982) Educational outcomes of tutoring: a meta-analysis of findings, *American Educational Research Journal*, 19(2), pp. 237-248.

Ebersole, E.H. (1972) The Soto pupil-team program for reading: an experiment in structural tutoring, *Improving Human Performance*, 1(4), pp. 39-42.

Fitz-Gibbon, C.T. (1977) *A Survey of Tutoring Projects* (Los Angeles, Center for the Study of Evaluation).

Fitz-Gibbon, C.T. & Morris, L.M. (1975) Theory based evaluation, *Evaluation Comment*, 5(1), pp. 1-4.

Gartner, A., Kohler, M. & Riessman, F. (1971) *Children teach Children: learning by teaching* (New York, Harper Row).

Glass, G.V. (1979) Policy for the unpredictable (uncertainty, research and policy), *Educational Researcher*, 8(9), pp. 12-14.

Glass, G.V., McGaw, B. & Smith, M.L. (1981) *Meta-analysis in Social Research* (London, Sage).

Glass, G.V. & Smith, M.L. (1979) Meta-analysis of research in class size and achievement, *Educational Evaluation and Policy Analysis*, 1, pp. 2-16.

Greene, J.C. (1980) Individual and teacher/class effects in aptitude treatment studies, *American Educational Research Journal*, 17(3), pp. 291-302.

Gustafsson, J.E. (1978) A note on class effects in Aptitude X Treatment Interactions, *Journal of Educational Psychology*, 70(2), pp. 142-146.

Hargreaves, D.H. (1981) Schooling for delinquency, in: Barton, L. & Walker, S. (Eds) *Schools, Teachers and Teaching* (London, Falmer).

Hartley, S.S. (1977) Meta-analysis of the effects of individually paced instruction in mathematics, *doctoral dissertation*, University of Colorado.

Thelen, H.A. (1969) Tutoring by students, *The School Review*77, pp. 229 -244.