DOCUMENT RESUME

ED 372 120 TM 021 963

AUTHOR Grover, Burton L.
TITLE Philosophical Inquiry into Meta-Analysis.
PUB DATE Oct 93
NOTE 14p.; Paper presented to the Northwest Philosophy of Education Society (Vancouver, British Columbia, Canada, October 1993).
PUB TYPE Viewpoints (Opinion/Position Papers, Essays, etc.) (120) -- Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Effect Size; Hypothesis Testing; Integrated Activities; Literature Reviews; *Meta Analysis; *Philosophy; *Research Methodology; *Statistical Analysis
IDENTIFIERS Criterion Variables; *Philosophers

ABSTRACT

A search of the ERIC database and a review of the literature suggests that meta-analysis is ignored by philosophers, a situation that is regrettable but remediable. Meta-analysis is a method by which one attempts to integrate findings quantitatively from several research studies related to a common general topic. Philosophers should certainly pay attention to meta-analysis if their task is to investigate knowledge claims and assess their significance. Three areas in particular are fertile ground for philosophers. One is the importance of the questions considered by meta-analysis. Another is the matter of generalization to a population. A third area for philosophers to consider is variation in criterion variables and parsimony. Many have been excited about the potential of meta-analysis to make sense of a mass of confusing contradictory studies and to reach new conclusions where none seemed logically possible. While results of some meta-analyses encourage this excitement, disagreements among methodologists can be disconcerting. Better technical expertise may resolve such problems, but it is also possible that philosophical consideration will give more direction to these efforts. (Contains 9 references.) (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
This document has been reproduced as received from the person or organization originating it.
Minor changes have been made to improve reproduction quality.
Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
BURTON L. GROVER

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

ED 372 120

PHILOSOPHICAL INQUIRY INTO META-ANALYSIS

Burton L. Grover
Western Washington University
Bellingham, WA

Presented to the Northwest Philosophy of Education Society
Vancouver, B.C., October 1993

TM021963

BEST COPY AVAILABLE

TM

PHILOSOPHICAL INQUIRY INTO META-ANALYSIS

Presented to the Northwest
Philosophy of Education Society
Vancouver, B.C., October 1993

Burton L. Grover
Western Wash. Univ.
Bellingham, Wash

The Educational Resource Information Center (ERIC), as you are probably well aware, is a comprehensive system for accessing and cataloging any literature from journals and unpublished papers which is remotely associated with education. The system is designed to enable investigators to find a manageable amount of literature pertaining to either a narrowly focused topic of interest or something cross-indexed to two or more topics. Every piece of literature in the ERIC system is given several topic descriptors to help in this searching.

On the CD-ROM for ERIC disk which covers the period from 1982 until June of this year, the indicator "philosophy" has 8467 listings. The narrower topic "educational philosophy" has 3799. The indicator "meta-analysis" yields 889. Despite the numbers in both these categories, an attempt to find anything in ERIC indexed to both will be fruitless. Such a cross-indexing yields a grand total of two references--one an article about the place of didactics in the curriculum in Scandinavia and the other from a business journal which surveyed management plans and noted the absence of specific philosophies behind each. The ERIC system is by no means perfect, but even so the cross-indexed search should give some picture of the degree to which philosophic inquiry has been directed at the well-established and growing field of meta-analysis. Using other philosophy oriented descriptors, such as

TM021963

logical positivism, is no more successful in finding a connection with meta-analysis (perhaps in this instance not surprising given how old obituaries are for logical positivism). A manual search through the table of contents of educational philosophical journals will probably confirm the inattention to meta-analysis.

Meta-analysis appears to be something completely ignored by philosophers. It is proposed here that this is an oversight by philosophers which is serious but remediable. (It seems that specialists in the methodology of meta-analysis often argue philosophically but pay little attention to those who are labelled as philosophers.)

The situation is reminiscient of C. P. Snow's (1960) two cultures wherein the literary community and the scientific community have little to do and are largely ignorant of the other. That was the situation Snow perceived and lamented over thirty years ago; the gap today between philosophy and meta-analysis appears seems like a current replay.

I do not wish to pretend to be as perceptive as Snow was nor as expert in both communities. Strictly speaking, I am neither a statistician nor a philosopher. Somewhere along the line I seem to have fallen into a crack between the two. Yet viewing both sides from this position, one can look for bridges or other lines of communication between them. None are seen. Neither are thrown rocks, at least none that are noticed. One see groups which simply seem to ignore each other.

Being between the two is not always a comfortable feeling. One can try to develop a survival strategy that includes trying to

gain credibility by saying "epistemological" a few times and hoping that technical jargon like "the test for homogeneity of effect size" has intimidation value. If the strategy is partly successful, idle curiosity may also lead one to decide which side is more easily intimidated by terms from the other.

If philosophers are generally ignorant about meta-analysis, an obvious initial question is, "What is it?" Meta-analysis is a method by which one attempts to quantitatively integrate findings from several empirical research studies related to some way to a common general topic. The term was officially coined in 1976 by Gene Glass in his presidential address to the American Educational Research Association as he described Mary Smith's and his effort to evaluate the effects of psychotherapy (Glass, 1976).

Since that time, the number of meta-analyses reported has grown steadily. Its use is not limited to education; the data bases of PsychINFO and Medline list an equal or greater number, and the greatest interest in its use may be in the health sciences.

Should philosophers pay attention to meta-analysis? Certainly, if their task is to investigate knowledge claims and assess their significance. If meta-analysts have anything left after they struggle to report their quantitative procedures, they make all sorts of claims about relationships between constructs. As with all empirical research, these constructs are operationally defined in terms of some method of measurement or operation, and the conclusions of meta-analyses may well include statements about the validity of these definitions.

How should philosophers approach meta-analysis? Gingerly, perhaps, given their past avoidance. Perhaps it would be best to

grope tentatively and non-judgementally, searching for issues and questions most deserving of greater attention. This intention of this paper is not so much to propose questions for philosophers to ask but to try to outline some areas which may be fertile ground for philosophical inquiry. Three such areas are provided and discussed briefly.

QUESTION AREA #1:THE IMPORTANCE OF THE QUESTIONS CONSIDERED BY META-ANALYSTS.

The question of importance can be asked of any research study, whether quantitative, qualitative, historical, or of any other type. The quick and easy answer is that some are and some are not. The procedure of meta-analysis or any empirical research inquiry is essentially neutral in regard to the significance of the issue studied. Further, the assignment of value to the topics studied can vary depending on whoever assigns the values; what may be trivial to some can be very important to others.

Is this response to the question too easy? Perhaps, but an alternative is not obvious. Maybe a better route would be to ask if tnere are any limiting factors which keep meta-analysts away from significant issues. The claim that meta-analysis is not appropriate for theory-testing could provide a clue, especially if theory-building and testing is given high priority in social and health sciences. So might the disdain of qualitative researchers for a process which cannot use their work.

A look at a sampling of topics or problems reported might be instructive. Meta-analyses reported in the Psychological Bulletin and The Review of Educational Research, the flagship journals for

reviews in psychology and education, from 1986 to 1992 included those addressing the following issues:

Psychological Bulletin
Psychological Predictors of Heart Disease
Cognitive-Behavior Therapy and Maladapting Children
Parental Divorce and Well-Being of Children
Gender and Leadership Style
Subliminally Activated Fantasies
Physical Attractiveness Stereotype
Psychological Effects of Military Service in Vietnam

The Review of Educational Research
Coaching for the Scholastic Aptitude Test
Mathematics and the Gender Gap
Effectiveness of Mastery Learning Programs
Student Self-Assessment in Higher Education
Effects of Vocabulary Instruction

Very likely each of us could make some judgement about the comparative value of at least some of the topics. I would think that a study of mastery learning is more important than coaching for the SAT and psychological predictors of heart disease than subliminally activated fantasies. Others might judge differently, but it is likely they could make a judgement. Yet could we articulate the criteria by which these judgments could be made, and to what extent would there be agreement or disagreement over them?

The fact that there could be more than one criterion could be complicating. A meta-analysis conducted on the effectiveness of using practice questions for instruction found many studies with conflicting results (Bangert-Drowns et al. 1991). A striking finding of further analysis was that much of the contradiction could be explained by a single "moderator" variable. Effectiveness was highly related to whether correct answers were made available to the students beforehand; studies which withheld access and corrective

feedback until subjects had attempted answers showed positive results while others did not. The issue of using practice questions for instruction does not seem to have the same sense of urgency as peace in Bosnia, yet teachers at most any level are likely to make decisions about practice questions more often than about Bosnia. Is frequency of applicfation a valid criterion?

Another aspect of the general question of subject matter importance is how to assign substantive value to numerical results. The basic result of most meta-analyses is an average effect size. An effect size is how much the average of the experimental group exceeds the control group when put on a standard scale. At least one is computed for each study in a meta-analysis, and they are then averaged across all studies. Harris Cooper (Wachter and Straf 1990) discusses some of the difficulties and disagreements researchers have had about making judgements of "large" and whether such judgements are contingent upon the behavioral science involved or other factors. A guideline sometimes proposed is that an effect size of .2 be considered "small." In that vein Harris reports a noted research reviewer deemed .3 to be "small" from a meta-analysis of the effects of desegregation on Black achievement scores. Is that judgement warranted? Possibly, but Harris points out that few if any are sure in these judgements and implies that anyone can get into the act. An effect size of .3 would indicate that approximately 62% of the experimental group would have scores (on whatever criterion measure used in the study) higher than the average member of the control group. Just how, if at all, can this be judged to be big or small?

QUESTION AREA #2: THE MATTER OF GENERALIZING TO A POPULATION:

All but the most nominalist of meta-analysts generalize or imply a generalization from their data to a larger population. This is implicit for any statistical study unless it is limited strictly to descriptive statistics, that is, if any statistical significance tests or probability estimates are reported. The distinctive thing about meta-analysis is that the data points are not scores on individual subjects but on complete studies. In other words, generalizations are not made to populations of subjects but to populations whose members are complete studies, each of which involves several subjects under certain specified conditions. This type of generalization involves a higher level of abstraction.

A few words about the concept of population may be appropriate here. A population is quite abstract and in a sense quite limited. For instance, a report of a sample survey refers--or should refer--not to a population of persons in the full meaning of the word person but to a set of responses to a specific question as asked and interpreted by a specific set of interviewers during a specific time period under specific circumstances. Once these limitations about defining the population are accepted, the task of a sample survey is to find a small representative sample from which it is reasonably safe to generalize to a larger population.

For an experimental study which involved random assignment of subjects to treatments and a statistical significance test, the strict interpretation of the generalization stemming from the significance test is not to other subjects or even to the same

subjects at other times or circumstances but to a hypothetical population of all other possible ways the subjects could have been divided (qualifications are in order here if we were to consider a study which utilizes a random-effects analysis of variance model, but that is outside the scope of this paper).

The full import of generalizing to other studies is not quickly grasped, at least it hasn't been for me. The generalization is at a higher level of abstraction. It is to all other hypothetical studies for a defined problem area which conceivably could have been conducted but were not. This includes studies which use markedly different measures of a criterion variable, such as self-reports, scores on a personality inventory, misbehavior referrals, or observations of play with dolls in studies of television and violence.

Questions to be posed for meta-analysis possibly might concern whether the studies included in a meta-analysis are in fact representative sample of that which could be studied. Would studies that have not been done for reasons of frugality or ignorance or whatever have yielded markedly different results? This includes the question which overlaps the next question area, whether the criterion measures used by the included studies are a fair sample of the domain of conceivable outcomes and whether the matter of generalizing across this dimension is at all appropriate.

This matter is made more difficult by the imperfections and limitations of individual studies. Here it is assumed that all studies are in some way imperfect and limited. To the extent that imperfections of one study cancel out the imperfections of

another, meta-analysis can cope (if the right "moderator variables" are identified and coded correctly). Also meta-analysis can handle the possibility that only studies with positive results are published while those showing no positive results languish hidden in some file-drawer. Yet if there are conceivable studies that have never been done but differ in certain key features, how can one generalize to those members of the population? If one cannot, close attention needs to be paid to how the findings are interpreted.

Generalizing from a sample to a population, in fact the concept of population itself, seems to involve some basic ontological and epistemological assumptions that philosophers might well try to uncover, make explicit, and then analyze.

QUESTION AREA NUMBER 3: VARIATION IN CRITERION VARIABLES AND PARSIMONY.

Most all meta-analyses include studies which use different measures of their outcomes. If the meta-analyses did not, there usually would not be enough studies in a grouping to integrate. This was in fact a distinctive feature of meta-analysis when it was first described.

To ward off potential criticism of meta-analysis for comparing apples and oranges, Glass (1976) stated that it was the business of program evaluators to make this comparison. Early meta-analyses showed how this could be done, particularly in the health sciences. In a meta-analysis of psychological treatments for asthma, different studies used criteria as different as remission of symptoms, psychiatrists' rating of improvement, use of drugs,

number of emergency room visits, responses on a Rorschach, and forced lung capacity as measures of effectiveness. In addition the time period used for follow-up evaluation varied from 0 to 120 weeks.

Does integration of different measures of something which is very general in order to find an overall effect size make sense? Relief of asthma probably is sufficiently restricted in definition, but what about when one tries to assess the effects of something like class size? Can cognitive and affective outcomes be combined? The meta-analysis of class-size studies, involving over 100 different comparisons, integrated data from standardized achievement tests, pupil attitude, teacher satisfaction, pupil-teacher interaction, and observations of teaching behavior. Could this diverse group also be meaningfully combined with any long-term social effects, should any investigator have attempted to look at those effects? (Both meta-analyses described in part in Glass, McGaw, and Smith, 1981)

If it is found that overall effect size varies systematically with criterion measures, the route most meta-analysts would probably take is to report the results as such, namely that the effect size is larger when certain criterion measures are used than when other measures are used (after adjusting for varying reliability of different measures in some meta-analyses). That seems simple and appropriate enough to do, but then is there any point in discussing overall effect size?

This question reveals a basic difference in belief and aims among meta-analysis methodologists. Some look for an basic and underlying complexity. Others, notably Hunter and Schmidt (1990),

argue strenuously for parsimony and describe methods by which variations in effect size can be ascribed to "artifactual" or chance factors, including such as low reliability of measures. Raudenbush (1991) suggests this reflects a difference between using reviews as preludes to further inquiry or using them as guides for practice. Research users, he suggests, are tired of funding more studies in a well-researched area and are saying, "It's time to sum up or shut up."

The simplicity versus complexity question is probably as old as philosophy itself. Philosophers who know the question could very well have insights on this issue as it applies to meta-analysis.

Many have been very excited about the potential of meta-analysis to make sense of a mass of confusing contradictory studies and to reach new conclusions where none seemed logically possible. Results of some meta-analyses encourage this excitement. Yet disagreements among methodologists can be disconcerting, and optimism is dampened when different meta-analyses on the same issue reach drastically different conclusions (Abrami, Cohen, and d'Apollonia, 1988). Maybe better technical expertise alone will resolve these problems, but it is also possible that philosophical consideration will give more meaning and direction to these efforts.

REFERENCES

Abrami, P C , Cohen, P A , & d'Apollonia(1988). Implementational problems in meta-analysis *Review of Educational Research, 58(2). 151-179.*

Bangert-Drowns, R L , C C Kulik, J A Kulik, and M.T. Morgan(1991). The Instructional Effect of Feedback in Test-Like Events. *Review of Educational Research, 61(2). 213-238.*

Cooper, H. M.(1989). *Integrating Research: A Guide for Literature Reviews(2nd ed.).* Newbury Park, Calif Sage

Glass, G. V (1976). Primary, secondary, and meta-analysis of research. *Educational Researcher, 5(10),* 3-8.

Glass, G V, McGaw, B , & Smith, M L (1981). *Meta-Analysis in Social Research.* Beverly Hills, Calif Sage

Hunter, J E & Schmidt, F L (1990). *Methods of Meta-Analysis: Correcting Error and Bias in Research Findings.* Newbury Park, Calif. Sage.

Raudenbush, S W (1991) Book review of *Methods of Meta-Analysis* by J. E. Hunter & F L Schmidt *Educational Researcher,* 20(7), 33-37.

Snow, C P (1960) *The Two Cultures and the Scientific Revolution.* Cambridge, Eng.: University Press, 51 pp

Wachter, W W and M L Straf, editors (1990) *The Future of Meta-Analysis.* New York Russell Sage Foundation