

*This chapter examines the conceptual relationships between the research strategy of critical multiplism and the methods of meta-analysis and generalizability theory for quantitative data synthesis.*

## Critical Multiplism, Meta-analysis, and Generalization: An Integrative Commentary

*Aurelio José Figueredo*

In Chapter Two of this volume, Shadish describes and defends the research strategy of critical multiplism. Critical multiplism can be characterized as the strategy that systematically extends and applies the principles of multiple operationalism (Campbell and Fiske, 1959) to all components of the research enterprise. Just as we deem that any single measure of a hypothetical construct is fallible and subject to systematic method bias, we must also acknowledge that any single research method or procedure (collectively called *research tactics*) is equally limited. Thus, multiple operationalism can be viewed as the application of critical multiplism to the problem of fallible measurement. Shadish identifies several components of the research process to which multiplism can and should be applied: question formation, theory or model selection, research design, data analysis, and interpretation of results.

The application of critical multiplism to certain components of the research process can be identified with certain familiar methodological doctrines. For example, in question formation, critical multiplism can be seen as addressing both the threat of confirmation bias (Greenwald, Pratkanis, Leippe, and Baumgardner, 1983) and certain aspects of the problem of incommensurability (Feyerabend, 1975). In theory or model selection, critical multiplism can be equated with the method of multiple working hypotheses (Chamberlin, 1890) or strong inference (Platt, 1964). In research design, critical multiplism can be seen as the premeditated equivalent of the progressive effect of competing scientific research programs

(Lakatos, 1970, 1978) or rival research traditions (Laudan, 1977). Such historical continuities make critical multiplism more, rather than less, useful as a broad integrative framework for contemporary methodology. The fact that the proposed unified system contains many familiar elements should increase confidence, not invite dismissal.

Shadish notes that critical multiplism can be implemented at one or both of two levels: the individual level and the institutional level. At the individual level, the individual researcher must take the trouble to implement critical multiplist strategies in any given study, so that the validity of his or her inferences from the study will be strengthened. At the institutional level, the population of researchers constituting any given portion of the scientific community must foster multiplicity in their research tactics, so that valid inferences can be drawn from the scientific literature as a whole if not necessarily from individual studies. At the institutional level, both the individual study and the individual researcher are considered fallible and biased instruments of which nothing more can be expected. Shadish reviews the various social psychological, motivational, and economic constraints on the practice of critical multiplism at the individual level and concludes that, for the present, it can be recommended only tentatively at that level. Although he acknowledges that institutional barriers can also exist, he nevertheless strongly endorses the immediate implementation of critical multiplism at the institutional level. Unproductive pseudoproblems and polemics, such as the current dispute between qualitative and quantitative approaches in program evaluation, would thus be rendered moot (Sechrest, 1992).

Given the literature documenting the intellectual vices of confirmation bias and the consequences of paradigmatic tunnel vision in the history of science (Greenwald, Pratkanis, Leippe, and Baumgardner, 1983), a stronger case can and should be made for the implementation of critical multiplism at the individual level. While the pettiness and egomania of certain researchers may hamper this effort, there is no reason to believe that at least some of us are not capable of rising above these limitations and aspiring to this worthy ideal (Chamberlin, 1890). Critical multiplism warns us that we may be biased in various ways as individual researchers, but it also counsels us that such bias is no justification for either intentionally cultivating our limitations or tolerating those failings in others. A consistent meliorist position would be to do what we can both as individuals and as parts of a collective.

Nevertheless, it is clearly easier for the implementation of critical multiplism to start at the institutional level. There are two alternative ways of accomplishing this: proactively or retroactively. Shadish argues for a proactive position, as exemplified by the expression *planned critical multiplism* (Shadish, 1986). He contrasts this with a mindless multiplism that introduces methodological heterogeneities uncritically at the indi-

vidual level for no apparent reason and with no particular purpose in mind. However, there is one way in which multiplism can be both critical and retroactive at the institutional level, and that is in the field of meta-analysis. In Chapter Three of this volume, Cordray, although not explicitly addressing the problems of critical multiplism, describes how modern techniques of meta-analysis can be used for these purposes.

Meta-analysis represents an emerging technology for the quantitative review of research literature. The purpose of meta-analysis is twofold: the data synthesis of homogeneous results and the causal analysis of discrepant results. Data synthesis of homogeneous results involves first testing for homogeneity of effect sizes, then pooling results across studies. The main functions of data synthesis are the estimation of population effect sizes and the enhancement of statistical power by the pooling of data to detect these effects. Data synthesis has one major problem: It often rejects statistical homogeneity, an indication of discrepant results. As with all null hypothesis testing, it is usually no more than a matter of achieving sufficient statistical power to obtain a statistically significant result. Thus, that the pooling of results in data synthesis has the twin objectives of estimating parameters of population effect sizes and testing for the significance of homogeneous effects across studies may be self-defeating.

When discrepant results are obtained, causal analysis looks carefully at the designs of the different studies. These design features represent the "local molar conditions" (Campbell, 1986, p. 69) under which the discrepant results were obtained. Causal analysis is accomplished by structural modeling of the reported effects. The function of causal analysis is to determine whether results can be generalized across study conditions by capitalizing on the "heterogeneity of irrelevancies" (Cook, 1993, p. 50) between individual studies. Thus, meta-analysis can be used retroactively to implement critical multiplism at the institutional level by the quantitative analysis of research results (General Accounting Office, 1992). As Cordray points out in Chapter Three, this meta-analytic enterprise is observational in nature, whether or not the individual studies sampled are experimental. Thus, meta-analysis can be thought of as an exercise in retroactive, a posteriori, critical multiplism at the institutional level.

Cordray reviews the various vulnerabilities to which meta-analysis is subject as a nonexperimental procedure by recapitulating some of the threats to primary analysis of nonexperimental data and applying the same logic to the planning of meta-analysis. These threats to validity include variable errors, such as sampling errors, processing errors, and coding errors; systematic errors relating to sampling issues, such as small-sample bias and frame bias; systematic errors relating to nonsampling issues, such as noncoverage of subgroups and the influence of inclusion rules on generalization; and nonresponse and missing data, such as field errors and processing errors of this type. He finds that each type of error experienced

in primary nonexperimental research can be duplicated at the meta-analytic level and that each has been acknowledged as a potential threat in the meta-analytic literature. However, Cordray is repeatedly drawn to the errors that often underlie the primary analyses on which the meta-analysis is ultimately based and that it may merely be reflecting. Thus, although meta-analysts can, of course, make errors of their own, it is possible that many of the major threats to meta-analysis consist either of repeating or of not compensating for the known limitations of the primary research.

Thus, the validity of meta-analytic results is contingent on the implementation of critical multiplist strategies at the institutional level, the alternative being a meta-analytic strategy of mindless multiplism that aggregates data without regard to the conditions under which they were obtained. Furthermore, this corrective action is often necessitated by the inadequate implementation of a priori critical multiplist strategies at the level of individual researchers. Any task or component of the research process left homogeneous at the meta-analytic level constitutes a potential threat to the validity of the meta-analysis. This fact implies that critical multiplist strategies of data synthesis at the institutional level are a requirement.

There is one other way in which meta-analysis can go beyond mere synthesis of the published results of primary data analyses. Typically, meta-analytic models are directed to the systematic variations in a single effect across multiple studies. Thus, although there may be multiple meta-analytic predictors, the criterial effect size remains essentially univariate. However, instead of limiting the scope of the meta-analysis to a single effect, we can extend meta-analytic techniques to reconstruct the broad pattern of relationships among the relevant variables. Thus, we can synthesize a correlation matrix from a sample of studies reporting either different sets or individual estimates of bivariate correlations. Each correlational element in the synthesized matrix represents an estimate of population effect size obtained by pooling across multiple studies, but no single study needs to report data on all the requisite relationships. We can then construct a structural equations model for this system of relationships at the population level that would not have been possible if we had used only the data from an individual study. Thus, the whole body of literature can yield more than the simple sum (or even the weighted average) of its parts, qualitatively as well as quantitatively, while the individual study can constitute a single piece of a larger mosaic (Schmidt, 1992).

For example, many studies of treatment compliance report the various predictors of compliance, whereas others report only the various effects of compliance on treatment outcomes. The problem with this piecemeal approach is that many predictors of poor compliance are also likely to predict poor treatment outcomes. Thus, from the results of any single study, we usually cannot discriminate the direct effects of such predictors

on outcomes from the indirect effects mediated through poor compliance. In fact, some studies (Coronary Drug Project Research Group, 1980; Yeaton, 1990) have shown that the significant predictors of compliance are systematically associated with poor treatment outcomes even in the control (that is, placebo) condition, where compliance simply cannot be materially causal. By meta-analytically synthesizing the full correlation matrix needed, we might be able to construct the requisite path model to estimate and test separately both the direct and the indirect effects of such predictors on treatment outcomes. Thus, the complex role of such important constructs in evaluation research as treatment compliance can be understood in the context of other relevant causal influences. This type of application represents yet another way in which meta-analysis can be used to implement critical multiplist strategies retroactively at the institutional level.

However, both Cordray and Shadish go beyond the uses of meta-analysis as a merely retroactive implementation of critical multiplism. They both propose that, by assessing the limitations of the scientific literature as a whole, we can plan future research strategically to compensate for specific gaps that we have identified in the collective empirical findings of a field. Thus, an individual study can be designed to foster critical multiplism at the population level by preferentially utilizing alternatives to the mainstream procedures. Rather than deviating capriciously for the sake of mere novelty, the individual researcher can strategically plan a study a posteriori to compensate for gaps in collective knowledge that quantitative data synthesis has revealed. Thus, individual researchers can practice a kind of affirmative action for underutilized tactical elements that promise to answer important empirical questions in the field. The scientific value of an individual study is therefore better assessed in relation to what it adds to the cumulative literature than in relation to what it might add by itself. Although these guiding principles may appear to be intuitively obvious, they do not now dominate attitudes toward research design in the social and behavioral sciences. Currently, any substantial deviation in research tactics has to be explicitly justified in peer review, however plausible it may be.

Taken together, Chapters Two and Three go a long way toward identifying major challenges and charting a promising new course for both qualitative and quantitative research methodology in the coming years. However, there are ways in which we can go beyond them to explore future directions for this general line of inquiry. For example, the literature has failed to address some additional threats to the external validity of meta-analytic models. These threats include violations of independence between replications, as when a meta-analytic sample contains multiple studies by a single researcher, by multiple researchers from a single "laboratory," and by multiple "laboratories" within a single research paradigm. Another,

related threat that the literature has not sufficiently explored is the consequences of correlated design features across studies. This condition produces multicollinearity of meta-analytic predictors, and it therefore introduces uncertainty or ambiguity of causal inference—a threat to the internal validity of the meta-analytic structural model. Moreover, because the retroactive critical multiplism of meta-analysis depends on the naturally occurring heterogeneity of irrelevancies in the research literature, failure to examine systematic homogeneities in the primary data critically constitutes a major threat to the entire meta-analytic enterprise in relation to the generalizability of causal inference.

Figueredo and Scott (1992) have proposed that, in addition to conventional meta-analysis, which they distinguish as effects meta-analysis, we develop a parallel technology for contents meta-analysis. Effects meta-analysis would continue to serve the function of causal analysis of discrepant results by creating structural models for variable effects, but it would otherwise remain subject to the limitations of the dominant research paradigms. In contrast, contents meta-analysis would be analogous to the content analysis of text, focusing as it would purely on the selection of methods, not on the results of individual studies. Appropriate techniques would include either exploratory or confirmatory factor analysis of correlated design features developing measurement models for what could be described as research paradigm constructs. Although many authors (Feyerabend, 1975; Laudan, 1977; Lakatos, 1970, 1978) have tried to define precisely what is meant by *research paradigm* since Kuhn (1970) first formulated the notion, the idea that individual scientific tactics (as Shadish understands the term in Chapter Two) are typically not independent but form discriminable constellations of related elements is common to most conceptions. An empirical approach to the study of these hypothetical metascientific constructs is to look for the stochastically predicted clusters of intercorrelated tactical elements meta-analytically in the methodological contents of published research. Although this multivariate operationalization of a research paradigm does not exhaust the concept, it does appear to be at least consistent with most of the variant formulations of the way in which research paradigms operate, and it should therefore be serviceable for the basic purposes of measurement.

Moreover, by combining the techniques of contents and effects meta-analysis, we could relate the methodological contents of studies to the magnitudes of the effects reported. Thus, the paradigm constructs would serve as latent predictors for structural models, providing both data reduction for the meta-analytic model predictors (which often exceed our usable sample size) and the statistical control of any spurious multicollinearity between them. We would accomplish this task by developing what we could call *meta-analytic factor analytic structural equation models* (Scott, Figueredo, and Hendrix, 1992; Bentler, 1989). Thus, by using these two

meta-analytic techniques in combination, we could determine empirically not only whether research paradigms existed as theoretically specified but what their parameters were as defined by their constituent tactical elements and the biasing impact that their systematic perturbations of effect size could have on our collective results. By providing a firmer evidential foundation for the concept of a research paradigm, this empirical approach might also help to prevent or at least limit the increasing trivialization of a useful metascientific concept by profligate and inflated uses that refer it to relatively minor variations in research tactics, vaguely defined ideological "movements," or transient methodological fads.

Another future direction that I think is promising is in the complementary applications of generalizability theory and meta-analysis in program evaluation (Figueredo, Scott, and McKnight, 1993). Although generalizability theory (Cronbach, Gleser, Nanda, and Rajaratnam, 1972) and meta-analytic studies (Hedges and Olkin, 1985) can both be represented mathematically as variance component models within the general framework of hierarchical linear models (Bryk and Raudenbush, 1992), differences in emphasis produce a gap. Current meta-analytic models emphasize the structural explanation of the systematic divergences between studies where heterogeneities exist, whereas data synthesis can only pool statistically homogeneous results. Because they consider different studies to be multiple levels of an additional facet, generalizability models, which also provide information about these differences, enable us to quantify the degree of relative convergence over the sampled levels of any facet and thus even between partially discrepant studies. For that reason, we might even be able to pool moderately heterogeneous results, with caution, if we use associated generalizability coefficients computed across the specified structural dimensions of difference to qualify our results. Just as we routinely report sample means and qualify them by associated standard deviations, we might be able to report more global estimates of effect size and qualify them by estimates of the systematic variability that can be expected across studies. Thus, the basic difference between these two approaches can be seen as an emphasis on parameter estimation rather than hypothesis testing in the structural models. Therefore, by the complementary application of generalizability theory, meta-analysis can avoid the intellectual cul-de-sac into which institutionalized Fisherian null hypothesis testing the social sciences has often led it (Meehl, 1978).

One other interesting contrast that might produce added complications in model specification, estimation, and testing dovetails with the concerns of contents meta-analysis. Currently, most generalizability models are developed for the results of single-researcher or single-research-group experiments. If we consider researchers as another possible facet for generalizability analysis, as the logic of critical multiplism implies that we should, the researcher facet is typically either fully crossed with all other

facets or constant across them. In a meta-analytic application of generalizability analysis, the researcher facet would be at least partially confounded with other facets of research design, because individual researchers may idiosyncratically study different aspects of the same general problem. Even if individual researchers are represented by multiple studies in the meta-analytic sample, multiple studies conducted by the same researchers are likely to share many design features due either to individual expertise or to theoretical orientation.

Thus, a hierarchical analytical strategy analogous to a split-plot design might be appropriate. Under such a strategy, systematic interindividual differences would be statistically controlled by a subjects factor for individual researchers. The variable design features of multiple studies by individual researchers could be represented by one or more within-subjects factors, and any larger classes or groupings of researchers, such as those associated with theoretical orientations or research paradigms, could be represented by one or more between-subjects factors. The resulting problems in model specification, estimation, and testing would be the high correlations that we might expect between the researcher factors and the study characteristics factors. The finding of Shadish and Sweeney (1991) that behavioral therapists disproportionately tend to use behavioral measures of outcome provides a good example. Another possible solution is to include only one study per individual researcher in the meta-analytic sample and not even attempt to discriminate between researchers and associated design features. However, this would merely mask the confound and not solve the problem. This problem is related to the problem posed by research paradigms that can be addressed by contents (as opposed to effects) meta-analysis. Contents meta-analysis studies the constellations of correlated design features independently of the outcome effects.

In conclusion, meta-analysis can ideally serve as a retroactive or a posteriori application of critical multiplism at the institutional level, pending a more widespread application of critical multiplism at the individual level. However, to accomplish this ambitious task, meta-analysis must eventually address the related problems of nonindependence of individual studies and correlated design features produced by the purported existence of internally homogeneous research paradigms. Complementary techniques that would explicitly include such broad metascientific concerns within the scope of meta-analysis have been proposed in this chapter. Such tactical extensions could function in the way in which certain correlational techniques were joined with existing experimental techniques to create the field of quasi-experimentation (Cook and Campbell, 1979; Cordray, 1986). Instead of merely criticizing the potential limitations of meta-analysis, as we used to criticize confounded or "flawed" experiments before quantitative quasi-experimental techniques were developed, we could then control directly for specifiable threats.

## References

- Bentler, P. M. *EQS: Structural Equations Program Manual*. Los Angeles: BMDP Statistical Software, 1989.
- Bryk, A. S., and Raudenbush, S. W. *Hierarchical Linear Models: Applications and Data Analysis Methods*. Newbury Park, Calif.: Sage, 1992.
- Campbell, D. T. "Relabeling Internal and External Validity for Applied Social Scientists." In W.M.K. Trochim (ed.), *Advances in Quasi-Experimental Design and Analysis*. New Directions for Program Evaluation, no. 31. San Francisco: Jossey-Bass, 1986.
- Campbell, D. T., and Fiske, D. W. "Convergent and Discriminant Validation by the Multitrait-Multimethod Matrix." *Psychological Bulletin*, 1959, 56, 81-105.
- Chamberlin, T. C. "The Method of Multiple Working Hypotheses." *Science*, 1890, 15, 92.
- Cook, T. D. "A Quasi-Sampling Theory of the Generalization of Causal Relations." In L. B. Sechrest and A. G. Scott (eds.), *Understanding Causes and Generalizing About Them*. New Directions for Program Evaluation, no. 57. San Francisco: Jossey-Bass, 1993.
- Cook, T. D., and Campbell, D. T. *Quasi-experimentation: Design and Analysis Issues for Field Settings*. Boston: Houghton Mifflin, 1979.
- Cordray, D. S. "Quasi-experimental Analysis: A Mixture of Methods and Judgement." In W.M.K. Trochim (ed.), *Advances in Quasi-experimental Design and Analysis*. New Directions for Program Evaluation, no. 31. San Francisco: Jossey-Bass, 1986.
- Coronary Drug Project Research Group. "Influence of Adherence to Treatment and Response of Cholesterol on Mortality in the Coronary Drug Project." *New England Journal of Medicine*, 1980, 303, 1038-1041.
- Cronbach, L. J., Gleser, G., Nanda, H., and Rajaratnam, N. *The Dependability of Behavioral Measurements: Theory of Generalizability for Scores and Profiles*. New York: Wiley, 1972.
- Feyerabend, P. *Against Method: Outline of an Anarchistic Theory of Knowledge*. London: Verso, 1975.
- Figueredo, A. J., and Scott, A. G. "An Example of Contents Meta-analysis and Its Use in Higher Education Retention Studies." Paper presented at the American Evaluation Association Annual Meeting, Seattle, 1992.
- Figueredo, A. J., Scott, A. G., and McKnight, P. E. "Complementary Applications of Generalizability Theory and Meta-analysis in Program Evaluation." Paper presented at the Canadian Evaluation Society Conference, Banff, Alberta, 1993.
- General Accounting Office. *Cross Design Synthesis: A New Strategy for Medical Effectiveness Research*. GAO/PEMD 92-18. Washington, D.C.: GAO, 1992.
- Greenwald, A. G., Pratkanis, A. R., Leippe, M. R., and Baumgardner, M. H. "Under What Conditions Does Theory Obstruct Research Progress?" *Psychological Review*, 1983, 93, 216-229.
- Hedges, L. V., and Olkin, I. *Statistical Methods for Meta-analysis*. New York: Academic Press, 1985.
- Kuhn, T. S. *The Structure of Scientific Revolutions*. (2nd ed.) Chicago: University of Chicago Press, 1970.
- Lakatos, I. "Falsification and the Methodology of Scientific Research Programs." In I. Lakatos and A. Musgrave (eds.), *Criticism and the Growth of Knowledge*. Cambridge, England: Cambridge University Press, 1970.
- Lakatos, I. *The Methodology of Scientific Research Programs*. Cambridge, England: Cambridge University Press, 1978.
- Laudan, L. *Progress and Its Problems: Towards a Theory of Scientific Growth*. Berkeley: University of California Press, 1977.
- Meehl, P. E. "Theoretical Risks and Tabular Asterisks: Sir Karl, Sir Ronald, and the Slow Progress of Soft Psychology." *Journal of Consulting and Clinical Psychology*, 1978, 46, 806-834.

- Platt, J. R. "Strong Inference." *Science*, 1964, 146, 347-353.
- Schmidt, F. L. "What Do Data Really Mean? Research Findings, Meta-analysis, and Cumulative Knowledge in Psychology." *American Psychologist*, 1992, 47, 1173-1181.
- Scott, A. G., Figueredo, A. J., and Hendrix, J. "Meta-analysis of Research in Higher Education Retention Studies." Paper presented at the American Evaluation Association Annual Meeting, Seattle, 1992.
- Sechrest, L. "Roots: Back to Our First Generations." *Evaluation Practice*, 1992, 13, 1-7.
- Shadish, W. R. "Planned Critical Multiplism: Some Elaborations." *Behavioral Assessment*, 1986, 8, 75-103.
- Shadish, W. R., and Sweeney, R. B. "Mediators and Moderators in Meta-analysis: There's a Reason We Don't Let Dodo Birds Tell Us Which Psychotherapies Should Have Prizes." *Journal of Consulting and Clinical Psychology*, 1991, 59, 883-893.
- Yeaton, W. H. "Causal Power: Strengthening Causal Claims Using No-Difference Findings." In L. Sechrest, E. Perrin, and J. Bunker (eds.), *Research Methodology: Strengthening Causal Interpretations of Nonexperimental Data*. Washington, D.C.: Agency for Health Care Policy and Research, U.S. Department of Health and Human Services, 1990.

AURELIO JOSÉ FIGUEREDO is associate professor of psychology at the University of Arizona.