Feature Review

THE *HANDBOOK OF RESEARCH METHODS IN SOCIAL AND PERSONALITY PSYCHOLOGY:* A Toolbox for Serious Researchers

Phoebe C. Ellsworth and Richard Gonzalez

University of Michigan

Social and personality psychologists have always embraced a creative diversity of methods, ranging from participant observation (Festinger, Riecken, & Schachter, 1956) to psychometric test construction (Osgood, Suci, & Tannenbaum, 1957), using verbal, behavioral, and physiological measures in laboratory and field settings. We have invented new research designs (Campbell & Stanley, 1966) and new statistical analyses (Campbell & Fiske, 1959; Kenny & La Voie, 1984). We have discovered sources of error that are common to all research on humans (Orne, 1962; Rosenthal, 1963), and developed techniques to overcome them. Personality psychologists are most famous for creative contributions to psychometric theory that solve measurement problems inherent in self-report and observer data. Social psychologists are most famous for their ingenious laboratory experiments, in which significant social psychological phenomena are re-created in simplified form in a setting that eliminates confounds and extraneous noise. When familiar methods seemed wrong for the research questions, social and personality psychologists happily devised new ones; methodological elegance and wit were admired nearly as much as theoretical insight. With new questions and new technology, the variety of methodological choices continues to proliferate, and the task of compiling a comprehensive handbook of research methods in social and personality psychology is a daunting one.

Reis and Judd's aim is to "demonstrate . . . the tremendous methodological richness and innovativeness to be found in social psychological research" and "to provide social-personality psychologists with resources for expanding the methodological diversity employed in their research" (pp. xi–xii). On the whole, the book achieves these goals. The range of methods discussed is considerably broader than the coverage of most textbooks on research methods in social and personality psychology; the handbook includes chapters on computer simulation (Hastie and Stasser), physiological measures (Blascovich), surveys (Visser, Krosnick, and Lavrikas), content and narrative analysis (Charles Smith), behavioral and observational coding (Bakeman), techniques for assessing growth and change (Collins and Sayer), and meta-analysis (Johnson and Eagly). The metaphor of the toolbox recurs again and again, along with the praiseworthy exhortation to use multiple methods.

The book is divided into three sections: Design and Inference Considerations, Procedural Possibilities, and Data Analytic Strategies. The design chapters address the usual issues of internal and external validity; replication; experiments, quasi-experiments, and nonexperiments; and holding variables constant versus varying them systematically versus letting them vary randomly. These chapters are clear, intelligent expositions and extensions of the work of Campbell and others, and should be very useful for a student who has decided on a basic procedure and wants to use it most effectively. (Hastie and Stasser's chapter on computer simulation, included in this section, is quite different: It is a step-by-step exposition of how to use the method to test theory.) The section on procedural possibilities is almost exclusively devoted to various kinds of measures, and the section on data analysis focuses on various kinds of statistical procedures. In the sections on procedures and analyses, most of the authors describe a particular kind of measurement or analysis—a particular tool—urge that it be added to the toolbox, and give a more or less detailed description of how to use it.

Handbook of Research Methods in Social and Personality Psychology, edited by H.T. Reis and C.M. Judd. London: Cambridge University Press, 2000. 558 pp. Cloth, $69.95; paper, $34.95.

The editors clearly took some pains to achieve consistency of coverage across the chapters. Most of the chapters describe the relevant threats to internal and external validity, most of the chapters on design and measurement mention data-analytical considerations as well, and most authors make the arguments (albeit fairly abstractly) that (a) what is most important is to choose a method that fits the question and (b) it is better to design and conduct the study right in the first place than to rely on elaborate statistics to patch things up later on. Most of the chapters are admirably broad in their outlook, discussing a range of methods within the assigned domain. The chapters by Bakeman on behavioral observation and by Kerr, Aronoff, and Messé on methods of small-group research are particularly impressive in this regard. Other chapters are not as broad in their coverage. The one by Bargh and Chartrand, billed in the preface as a chapter on cognitive mediation, is narrowly focused on priming and automaticity; Bartholomew, Henderson, and Marcia's chapter on interviewing is heavily focused on a particular type of coded semistructured interview that, as they acknowledge, is labor intensive and appropriate for a particular kind of research question. The book as a whole, despite its title, devotes far more attention to the concerns of social psychologists than to those of personality psychologists. Individual differences are mentioned in several chapters, but are not the major focus of attention, and personality-test construction shows up only in John and Benet-Martínez's chapter on measurement.

FITTING THE METHOD TO THE QUESTION

The editors also say that "a guiding principle in preparing this volume was that theoretical and methodological questions are not independent" (p. xii). That is, a student who is starting out with a question should be able to choose a method that fits it. (And in turn, an apparently methodological obstacle may have important theoretical implications.) This handbook is likely to be very useful to people who

Address correspondence to Phoebe C. Ellsworth, Department of Psychology, University of Michigan, 525 East University, Ann Arbor, MI 48109-1109; e-mail: pce@umich.edu.

already have an idea about what method they want to use and want to learn more about it; it is likely to be less useful to people who know what they want to study and are wondering which method would be best. Very few of the chapters are organized around content (the chapters by Kerr, Aronoff, and Messé and by Kashy and Kenny on groups and the one by Collins and Sayer on growth and change are exceptions), and although a number of the chapters are quite good at telling us which version of a method to use once we have settled on a general type, few help us decide whether the general type is suitable in the first place. Reis and Gable, for example, provide very specific advice about when to use randomly timed beepers or fixed-interval beepers or the occurrence of the event itself in event sampling, but do not discuss in detail the kinds of questions that are right for event sampling or how event sampling compares with other methods. More generally across all chapters, a researcher who gets the message that multiple methods are a good thing, and decides to try something new, will receive little guidance about which new method will best complement the old one.

Of course, one reason that fitting the method to the question is so difficult is that it requires knowing exactly what the question is and what it is not. A heuristic that can clarify the underlying question is to generate a variety of possible answers, or "plausible rival hypotheses" (Campbell & Stanley, 1966), or alternative models, and to design a study that can rule some of them out. A study should pit one hypothesis against another (or others), not just test a single hypothesis. Experiments should be multiple-choice tests, rather than true/false tests. Several of the authors make this important point. Brewer talks lucidly about testing moderators and mediators and about deciding which types of external validity are relevant to one's question; Eliot Smith includes an excellent discussion on designs for testing alternative hypotheses; West, Biesanz, and Pitts remind us that theory and method are intimately related, so our designs and measures need not control for all possible alternatives, but should control only for those that are theoretically plausible; John and Benet-Martínez review how convergent and discriminant validity facilitate construct validation. In general, almost all of the authors argue that there are no absolute standards for good methods, only the relative standard of a method's fit to the question. Old preferences die hard, however, and there are many lapses, with chapters declaring certain methods superior in an absolute sense: Counterbalancing is essential in a within-subjects design; research on group processes should measure both task and socioemotional variables; rating scales are better than true/false questions, and so on. Despite many protestations to the contrary, causal questions are still favored over other kinds of research questions, and methods that can answer causal questions are favored over other methods.

FITTING THE ANALYSES TO THE QUESTION (AS WELL AS TO THE METHOD)

Although multiple methods are desirable and researchers are encouraged to follow Campbell's notion of triangulation (p. xii), the organization of the data-analysis chapters tends to follow a traditional list of Technique A, Technique B, Technique C, and so on. These chapters are relatively comprehensive in their review. John and Benet-Martínez discuss classic psychometric tools and new developments such as generalizability theory and item response theory; McClelland provides valuable tools for checking statistical assumptions and dealing with outliers; Judd illustrates how some common analyses can be placed within the unifying framework of the general linear model; Wegener and Fabrigar supply an assortment of multivariate techniques that can be used with data from nonexperimental designs; Johnson and Eagly provide computational details for the calculation of effect sizes in a variety of designs.

For the most part, the notion that analyses should also fit the research question receives relatively little attention (there are exceptions; e.g., Judd points out that contrasts can be used to test specific research hypotheses). In several chapters, the emphasis appears to be more on merely fitting the analysis to the design than on fitting it to the question. In the context of research, the "fits" relation is not necessarily transitive. If the analysis fits the design, and the design fits the research question, we are not guaranteed that the analysis fits the research question. For example, a repeated measures analysis of variance may fit the design (repeated observations over time, such as in a diary study), and the design may fit the research question (e.g., What is the time course of a psychological process?), but the repeated measures analysis of variance may not be the best way to test specific questions about individual differences in change or transitions across psychological stages (as discussed in the chapters by Collins and Sayer and by Reis and Gable).

A "triangulation" that should occur in empirical research involves the interdependence of the research question, the research design, and the data-analytic technique. All three are necessary pillars in the development of a strong research program and need to be considered simultaneously. All too often, research in social-personality psychology tends to focus on only one, or possibly two, of the three necessary pillars.

METHODOLOGY'S NEW LOOK: MODELING INDIVIDUAL DIFFERENCES AND TEMPORAL CHANGE

There is an interesting theme that imbues several chapters: Observations come to the investigator as collections. In this context, a collection is not necessarily the usual aggregate of subjects assigned to the same experimental condition or the profile of multiple dependent variables from the same subject. Instead, the idea is that data are sometimes collected via a sampling strategy that involves a collection of observations (or subjects), say, in the form of a time series or in the form of a social group such as a married couple, a family, or a classroom. The recognition that sometimes a data set consists of a random sample of collections has led to the methodological advance known as hierarchical linear modeling, which treats statistical parameters as random effects with an underlying distribution and allows those parameters to themselves be treated as dependent variables. But, much as need begets method, surprisingly method can beget need. This technological advance opens doors to new types of research questions.

The handbook includes several chapters in this spirit. For example, the chapter by Collins and Sayer shows how to model change over time for a single subject, essentially by applying the usual linear regression to each subject (here the collection consists of a single subject's repeated scores) and estimating regression parameters for each subject separately in a way that allows proper pooling across subjects. The chapter by Kashy and Kenny shows how to model nonindependence in dyad and group research by defining the interacting group as a collection, essentially fitting the usual statistical model to each socially meaningful aggregate of subjects and then pooling in a statistically appropriate way.

For us, the deep implication of the hierarchical approach is less statistical (in the sense that it is a new data-analytic tool) and more theo-

retical. Hierarchical models attempt to model each unit separately as well as in the context of other relevant cases. It is in this sense that the methodological emphasis on hierarchical linear models leads the researcher to model individual difference and temporal change explicitly—to no longer lump individual variability into the error term.

It may seem odd to include both social and personality research methods in the same handbook volume, especially given the tensions between the two areas over the primacy of the situation versus the person, over differences in research methodology, and over differences in data-analytic techniques. However, aided by new developments in statistical methodology, we may be able to put down our differences, to model psychological processes as involving both the person and the situation, both between-subjects variability and within-subjects variability.

WHAT'S MISSING? REAL SITUATIONS, REAL PEOPLE

A review of Aronson and Carlsmith's (1968) chapter on experimentation in social psychology in the *Handbook of Social Psychology* points up what is missing in this new handbook. Those authors included sections on capturing the phenomenon under investigation and on techniques for turning conceptual independent variables into real events. They thought about the study as a coherent, meaningful whole in which the measures made sense in relation to the manipulations, and both made sense in relation to the rationale that was given to the subjects. The independent variable was not merely a matter of number of levels, fixed versus random factors, holding constant versus varying systematically; it was also a matter of real events being interpreted by real people and affecting their responses. In the present handbook, the independent variables are generally abstractions, and the challenge of re-creating an analogue of a social psychological process never comes through. The handbook has very little to say about what happens before the measurements are taken and the data are analyzed. We are exhorted that careful attention to these stages (avoiding confounds, subject loss, reactivity) is more important than analytic patch-ups, but the emphasis of the book belies the exhortation.

Put another way, the people we study have all but disappeared. We may call them "participants" now, rather than "subjects," but this handbook treats them like disembodied units of observation, not like people. Real people walk into a lab in the psychology department, wear a beeper all day long, are chosen for a family intervention program or a charter school, are called up by a survey researcher. They wonder what the study is about and what it is for. They have preconceptions about what the study is probably about. They will have very different motivations depending on whether the research, or the situation, makes sense and seems important or seems like a waste of time, and depending on whether it is engaging, confusing, or boring. When we first speak to the participants in our studies, what do we say? How do they interpret the situation? What do they think they are supposed to do in order to look "normal," "intelligent," or "politically correct?" Do we want them to alter their responses to create these impressions or not, and either way, how do we get them to do it? What do they think of us? What is actually going on in the minds of our subjects? Is it anything like what we intended?

A few chapters (Visser, Krosnick, and Lavrikas; Bartholomew, Henderson, and Marcia) emphasize the importance of pilot testing, but on the whole the book is silent about the frustrating, exciting, and crucial stage of getting a feel for the research problem. Even the chapters that do talk about pilot testing mostly discuss measurement problems.

Post hoc manipulation checks to assess mediators are discussed as though people have access to intervening psychological processes after the processes have had their effects. Surely this assumption is sometimes questionable. Did dissonance theorists ask their subjects if they had experienced dissonance? Do the psychologists who study priming ask about the meaning of the prime? Of course not. But why not? Because they do not believe it is useful to do so, given the psychological processes they care about. Sometimes manipulation checks can provide validity information, sometimes they cannot.

Reactivity, demand characteristics, and experimenter bias are mentioned in passing (but far less often than new analytical techniques such as hierarchical linear modeling) with no serious consideration of the thoughts and feelings of the people in a study. Our subjects are treated less and less like people and more and more like data points.

CODA

The final two words of Aronson and Carlsmith's 1968 chapter were, "It's fun!" The authors of the present handbook chapters do not seem to be having much fun. Trying out different ways of framing the question, considering alternative hypotheses, choosing appropriate designs, devising procedures and measures that capture the essence of the phenomenon, exploring different ways of looking at the data—these can be exciting enterprises, full of surprises. The chapter by McClelland on "nasty data" conveys a little of this, suggesting that frustrating data can be a source of deeper understanding. Many of the chapters in this handbook contain litanies of dangers, threats to validity, and statistical obstacles, however, and the research enterprise seems like a grim struggle. Although Aronson and Carlsmith viewed social psychological research as a land of opportunity, the present handbook makes it seem more like a minefield. It is understandable why the authors of the handbook took this approach: to prepare researchers for the critiques they will likely receive from journal reviewers. The hope is that researchers will make fewer dumb mistakes after reading this book; the danger is that they might be so intimidated that they stick to old, familiar paper-and-pencil lab experiments in which the subjects are homogeneous, the *n*s are equal, everything can be counterbalanced, and all possible moderators can be measured.

REFERENCES

Aronson, E., & Carlsmith, J.M. (1968). Experimentation in social psychology. In G. Lindzey & E. Aronson (Eds.), *Handbook of social psychology* (2nd ed., Vol. 2, pp. 1–79). Reading, MA: Addison-Wesley.

Campbell, D.T., & Fiske, D.W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin, 56*, 86–105.

Campbell, D.T., & Stanley, J.C. (1966). *Experimental and quasi-experimental designs for research*. Boston: Houghton Mifflin.

Festinger, L., Riecken, H.W., & Schachter, S. (1956). *When prophecy fails*. New York: Harper Torchbooks.

Kenny, D.A., & La Voie, L. (1984). The social relations model. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 18, pp. 141–182). Orlando, FL: Academic Press.

Orne, M.T. (1962). On the social psychology of the psychological experiment: With particular reference to demand characteristics and their implications. *American Psychologist, 17*, 776–783.

Osgood, C.E., Suci, G.J., & Tannenbaum, P.H. (1957). *The measurement of meaning*. Urbana: University of Illinois Press.

Rosenthal, R. (1963). On the social psychology of the psychological experiment: The experimenter's hypothesis as unintended determinant of experimental results. *American Scientist, 51*, 268–283.