

Robert MacCoun

Experimental and Quasi-Experimental Designs for Generalized Causal Inference, by William R. Shadish, Thomas D. Cook, and Donald T. Campbell. Boston: Houghton Mifflin, 2001, 623 pp., \$63.16.

I know a social psychologist who carries in his wallet a laminated copy of Campbell and Stanley's (1966) list of threats to the internal validity of causal inferences. The list is not long—surely he has memorized it by now. But I suspect that the card offers epistemological comfort in times of doubt. Though ostensibly about education evaluation, Campbell and Stanley's rarified analysis of causal inference, with its tables of X's, O's, +'s, and -'s, has an almost Platonic timelessness.

Though a dozen years younger, Donald Campbell's sequel with Thomas Cook (Cook and Campbell, 1979) seems more outdated, perhaps because its aspirations were more encompassing. Campbell, who died in 1996, envisioned an "experimenting society" that would make steady progress through systematic trial-and-error experimentation—in both the popular sense of "trying something new" and in the methodological sense of "random assignment to alternative treatments" (e.g., Campbell, 1969). But, as Campbell and Stanley recognized, there are often major political, ethical, or logistical barriers to random assignment, especially when no-treatment control groups or risky interventions are involved. Thus, Cook and Campbell (1979) devoted an entire monograph to quasi-experimentation—the use of designs that lack random assignment to condition, yet creatively use time and circumstance to permit stronger inferences than a simple observational analysis would allow.

Where Campbell and Stanley provided a concise and focused overview of research designs, Cook and Campbell attempted a state-of-the-art treatise, drawing on then-recent work in the philosophy of science, the politics of bureaucracies, and the statistical analysis of quasi-experimental correlational data. Its perspective has been challenged by the growth of econometrics, the Reagan revolution, and the post-modernism "turn" in the academia.

Random assignment is still the "gold standard" for causal inference, but field experiments rarely resolve uncertainties about program effects. Even with high levels of cooperation, randomization is rarely implemented with complete fidelity. Through the vagaries of real life, participants are misassigned or reassigned or they drop out, undermining many of the inferential benefits of the experiment. Ideally, we would like results to speak for themselves, but it seems that program evaluation will always need sophisticated statistical methods.

Since Cook and Campbell's (1979) book was published, there have been major developments in econometrics and statistical inference, including instrumental variable models, multi-level regression analysis, intention-to-treat modeling, the propensity score approach, selection bias modeling, and the computation of identification bounds on effect sizes. (Winship and Morgan [1999] provide a nice overview of these approaches for non-specialists. The relative merit of these approaches is debated in a series of comments by Rosenbaum, Manski, Robins, and Shadish and Cook in the 1999 and 2000 volumes of *Statistical Science*.)

In *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*, Shadish, Cook, and Campbell applaud these developments rather than resisting them. But they argue persuasively that such methods are a second-best solution when rigorous research designs are precluded or undermined. The stronger the design, the more transparent and explicit the results. In the end,

analysts who present complex econometric manipulations are essentially asking non-specialist audiences to “trust me.” Notwithstanding the motto of my own policy school—“speaking truth to power”—few politicians are willing to simply trust that policy analysts have a direct pipeline to Truth, unless of course the results are politically palatable.

Since the Reagan era, conservatives and moderates have been skeptical about the government’s role in social engineering, so they have been most inclined to embrace program evaluation when it shows that “nothing works” (Murray, 1984). Such examples are easy to find, and Shadish, Cook, and Campbell’s book can help explain why. Far too many program evaluations are almost guaranteed to fail because of their small sample size, sloppy measurement, uncontrolled heterogeneity in treatments and samples, and slavish adherence to the 0.05α level in situations where an a priori decision analysis might have emphasized avoiding false negatives.

But rather than demanding better tests of social programming, many liberal academics abandoned the notion of truth altogether. The postmodern critique of knowledge is unassailable if one is an ontological idealist—if such creatures really exist. But for those of us who believe there is a world out there, the fallibility of knowledge hardly implies that all arguments are equally compelling. The postmodern attack on the social sciences targeted a straw person called the “positivist.” But Tom Cook (1985), perhaps tongue in cheek, has provided us with a fashionably unwieldy PoMo label—post-positivist critical multiplism—for the dominant view of most empirical social scientists at least since Campbell and Stanley (1966): The choice of research methods poses inevitable tradeoffs, and we can at best hope to approximate truth through a strategy of triangulation across multiple studies, investigators, and fallible methodologies.

The critical strain of post-modernism is more convincing than the epistemological strain, articulating the ways in which privilege, power, class, gender, and ethnicity infiltrate perceptions and assertions. But, in fact, the Campbellian framework offers powerful tools for rooting out such biases. Indeed, such methods can and have been used to study bias in the research process itself (see MacCoun, 1998).

Rather than simply updating Cook and Campbell (1979), Shadish, Cook, and Campbell rejuvenate the Campbellian framework from top to bottom. They are still forceful in articulating the enormous inferential benefits of careful field experimentation. But they abandon their earlier insistence that “internal validity is the sine qua non” of evaluation research. External validity assessment has always had an “I know it when I see it” element, lacking the kind of explicit analytic framework Campbell and Stanley provided for internal validity. Taking seriously alternative meta-theories of program evaluation (see Shadish, Cook, and Leviton, 1991)—especially the thinking of Lee Cronbach—the authors offer a new “grounded theory of generalized causal inference” with principles for the sampling of causes, effects, persons, and settings.

Our understanding of generalization has benefited enormously from the development and widespread application of modern meta-analytic methods. It has become clear that apparent inconsistencies across studies often reflect differences in statistical power and implementation, differences that render simple effect versus no-effect “box scores” highly misleading. Skepticism about the hazards of merging apples and oranges—what one statistician once called mega-silliness—have largely given way to an appreciation of how meta-analysis can be used to identify important moderators of the magnitude and even the direction of program effects. Indeed, meta-analysis has become an important tool for both theory

development and methodological study. Over the past decade, Shadish has published a series of meta-analytic studies of research designs, demonstrating, for example, that in many domains experimental and quasi-experimental designs yield comparable effect sizes. In *Experimental and Quasi-Experimental Designs*, Shadish and colleagues devote only a single chapter to meta-analysis, but the meta-analytic perspective—focusing on triangulation across heterogeneous sources of evidence—is apparent throughout.

This new book clearly updates the arguments of Campbell and Stanley (1966) and Cook and Campbell (1979). The reader should not think of it as simply a revised edition of its predecessors, but, rather, a major new contribution to the study of policy analysis, and indeed to the broader social scientific enterprise. The policy analysis community has thankfully managed to avoid humanities-style debates about our “canon,” but if we ever start one, this book would be my top candidate. No *JPAM* reader—and no student of a *JPAM* reader—should be without it.

ROBERT MACCOUN is Professor at Goldman School of Public Policy, and at Boalt Hall School of Law, University of California, Berkeley.

REFERENCES

- Campbell, D.T. (1969). Reforms as experiments. *American Psychologist*, 24, 409–429.
- Campbell, D.T., & Stanley, J.C. (1966). *Experimental and quasi-experimental designs for research*. Skokie, IL: Rand McNally.
- Cook, T.D. (1985). Post-positive critical multiplism. In Shotland & Mark (Eds.), *Social science and social policy* (pp. 21–62) Beverly Hills, CA: Sage.
- Cook, T., & Campbell, D. (1979). *Quasi-experimentation: Design and analysis issues for field settings*. Boston: Houghton Mifflin.
- MacCoun, R. (1998). Biases in the interpretation and use of research results. *Annual Review of Psychology*, 49, 259–287.
- Murray, C. (1984). *Losing ground: American social policy, 1950–1980*. New York: Basic Books.
- Shadish, W.R., Cook, T.D., & Leviton, L.C. (1991). *Foundations of program evaluation: Theories of practice*. Newbury Park, CA: Sage.
- Winship, C., & Morgan, S.L. (1999). The estimation of causal effects from observational data. *Annual Review of Sociology*, 25, 659–706.

Frank Levy

The Roaring Nineties: Can Full Employment Be Sustained?, edited by Alan B. Krueger and Robert M. Solow. New York: Russell Sage Foundation and the Century Foundation Press, 2001, 592 pp., \$49.95.

In one sense, the economy’s performance in the second half of the 1990s was as unexpected as it was good. For most of the previous 20 years, the economy had been shackled by slow productivity and labor and product markets that seemed to generate inflation well before reaching full employment. As recently as 1992, the economy was recovering from recession, but observers were describing it as “the jobless recovery.”

Not to worry—between 1995 and 2000, productivity growth rose substantially, unemployment eventually fell as low as 4 percent, and inflation was nearly nonexistent. It was