

S. Donaldson, C.A. Christie, and M.M. Mark. (2009) *What Counts as Credible Evidence in Applied Research and Evaluation Practice?* Thousand Oaks, CA: Sage. 265 pages.

Reviewed by Kenneth Watson

“In God we trust—All others must have credible evidence.”

The first issue of the *Canadian Journal of Program Evaluation* contained the report of an interview with Donald T. Campbell on the topic of credible evidence (Watson, 1986). Campbell already had 20 years behind him of being an eminent guru and proponent of experimental and quasi-experimental design for evaluations (Campbell & Stanley, 1963). That’s a long time ago, but in 2011 the old battles about what is acceptable evaluation evidence still rage. Is experimental evidence the “gold standard,” or are other more messy methods equally insightful?

It is easy to illustrate the heat that is generated by this debate. For instance, in 2003 the U.S. Department of Education, Institute of Education Sciences, declared that it would favour experimental designs and some types of quasi-experimental evaluation designs over other methods in its granting competitions (Federal Register, RIN 1890-ZA00). In response, the American Evaluation Association, at its 2003 annual meeting, released a statement that such a preference for experimental and quasi-evaluation designs would be a “political, ethical, and financial disaster.” The AEA did not, however, claim that it would be subversive or fattening, although some have raised those possibilities.

In 2006 the Stauffer Symposium Series at Claremont Graduate University in California held a conference on “What Counts as Credible Evidence in Applied Research and Evaluation Practice?” which, three years later, resulted in an edited compilation of papers with the same overall title. The editors included a past-president of the American Evaluation Association and two eminent professors and evaluators. The book has two core sections:

- “Experimental Approaches as the Route to Credible Evidence”—100 pages, four articles, one of which argues strongly for high-quality impact evaluations, that is experimental and

quasi-experimental designs; and three others that address the idea of experimental approaches but are skeptical.

- “Non-experimental Approaches for Building Credible Evidence”—80 pages, five articles, all arguing in one way or another that experiments (randomized field trials) and quasi-experiments are either not needed or are not sufficient alone.

In addition the book has an introduction and a conclusion, of about 70 pages together, in which the editors attempt to frame and elucidate the debate. Stewart Donaldson (ed.) contributes “In Search of the Blueprint for an Evidence-Based Global Society” and an epilogue entitled “A Practitioner’s Guide for Gathering Credible Evidence in the Evidence-Based Global Society.” Christina Christie (ed.) contributed “Social Enquiry Paradigms as a Frame for the Debate on Credible Evidence”. Melvin Mark (ed.) contributed the conclusion, “Credible Evidence: Changing the Terms of the Debate.”

Mark notes the complete lack of consensus in the evaluation community on the topic. He believes that the fundamental disagreement is based on even more fundamental disagreements:

disparate views about what constitutes credible evidence are predicated on divergent, often implicit, assumptions ... and beliefs about the ... purpose of applied social research and evaluation ... and quite different assumptions about the relative ease or difficulty of getting actionable research findings about program effects. (p. 215)

In short, what is the purpose of program evaluation and how difficult is it to achieve that purpose? Mark thinks that we should focus on these more fundamental questions, and, if we did, “future discussions might reduce some disagreements, clarify the rationale that underlies other disagreements, and perhaps even facilitate the injection of credible evidence into the debate itself” (p. 216).

Mark warns that “readers who are familiar with the debate only from this volume, which contains statements from a selected set of relatively thoughtful commentators” (p. 218), will not get the full flavour of the rhetorical excesses to which the debate has been prone.

In summary, is this book useful to Canadian evaluators who are interested in the debate about the necessity of experimental or quasi-experimental impact evaluations? Probably. Its design is a good one, containing papers for, papers against, and papers offering a synthesis

and summing up. However, the case for experimental and quasi-experimental evaluations does not really get equal air time. Most of the contributors put forward arguments about why non-experimental methods are valuable (maybe more valuable), and the defenders of the more rigorous methods have an air of defensiveness that perhaps was sparked by the symposium that generated the book, which in turn was sparked by the firestorm of angry response that greeted the U.S. Department of Education's initiative three years previously. The book is a useful contribution to the debate, but, I would argue, not a balanced one.

REFERENCES

- Campbell, D. T., & Stanley, J.C. (1963). *Experimental and quasi-experimental design for research*. Chicago, IL: Rand McNally.
- Watson, K. F. (1986). Programs, experiments and other evaluations: The key role of replicable treatments. *Canadian Journal of Program Evaluation*, 1(1), 83–86.