Book Review

Katherine Jenny Thompson¹

Paul J. Lavrakes, Michael W. Traugott, Courtney Kennedy, Allyson L. Holbrook, Edith D. de Leeuw, and Brady West. *Experimental Methods in Survey Research: Techniques that Combine Random Sampling with Random Assignment*. 2019, Wiley, ISBN: 978-1-119-08374-0, 544 pages.

Qualitative and quantitative research are entrenched in survey methodology. Certainly, qualitative research tools that use small datasets such as focus groups and onsite visits of businesses provide invaluable information in assessing *construct validity* in survey instrument design, as does usability testing. However, these tools are limited, in the sense that they rely on observational data, are purposively gathered, and should not be used for any form of causal inference. At the opposite end of the spectrum, large amounts of data can be easily gathered via the internet using probability or nonprobability samples such as online convenience panels; the resulting statistics are time-dependent measures, generally considered to be externally *valid*. Again, however, these vehicles do not allow for cause-and-effect analysis.

Experimental Methods in Survey Research is an extensive collection of papers gathered together for one express purpose: to convince survey researchers to utilize experimental designs combining random assignment of subjects to treatment within random (probability) samples. *Internal validity* is attained via the random assignment of treatments, using blocking in the assignment process to control for “known” factors. Using a probability sample combined with successful recruitment methods should achieve *external validity*, assuming that the sampling frame covers the target population (i.e., has low *coverage error*) and that the response mechanism is *ignorable* (i.e., respondents are a random subsample). In assembling and organizing these chapters, the editors “embraced both the Campbell and Stanley validity framework” for research studies (statistical conclusion validity, construct validity, external validity, and internal validity) as well as the key components of the total survey error (TSE) framework, specifically coverage error, sampling error, nonresponse error, and measurement error. Thanks to the extensive usage of computer-assisted technologies in survey collection and the prevalence of internet access in the majority of target populations, the editors argue that there is little or no cost increase in embedding true experiments within ongoing programs, and therefore such tests should be carefully planned and executed on a wide scale in common practice.

The book is divided into nine topical areas, comprising twenty-three chapters (the first chapter is an introduction). Each chapter covers one experimental research topic, providing a literature review and presenting at least one case study. Not perhaps

¹ U.S. Census Bureau, Economic Statistical Methods Division, 4600 Silver Hill Road, Washington, D.C., 20233, U.S.A. Email: Katherine.j.thompson@census.gov.

Acknowledgments: Any views expressed are those of the author and not necessarily those of the U.S. Census Bureau.

unexpectedly, the majority of the material covers issues related to household surveys; establishment surveys are limited to one chapter. Given this framework, there is a high percentage of real estate dedicated to attaining appropriate within-household coverage, and potential interviewer effects in terms of participation, response (unit and item), and experiment implementation. An equally high percentage of real estate is dedicated to assessing mode effects. In many of the presented studies, satisficing is a major quality concern, and response and completion rates are the primary measurement of treatment effects.

There are compelling arguments for embedding designed experiments into ongoing programs, especially probability surveys. However, there are other considerations that are highlighted in the book. Coverage of the broad population might not be an issue, but assessment of subdomain-specific effects might be. For example, comparisons between gender-specific or age-specific estimates could be made difficult by imbalanced samples. Interviewer effects can impact field experiments: is it better to assign one treatment per interviewer or all treatments to all interviewers? The scope of the experiment is important. Survey managers are unlikely to implement an experiment that might lower the overall response rate or noticeably affect the quality of the final estimates. For these practical reasons, embedded experiments might be confined to a portion of the survey sample such as small businesses or farms in an establishment survey. Carryover effects need to be considered with experiments embedded in longitudinal surveys; introducing or removing a treatment from a longitudinal panel could likewise have effects on response and/or quality and could lead to confounding. Telephone or internet collection could reduce or avoid interviewer effects. However, telephone-only and internet-only sample designs can be subject to frame coverage imbalances. Furthermore, response and completion rates for internet surveys tend to be very low. Each of these topics is discussed in detail in the context of experiments embedded in a variety of surveys.

As an overview of quantitative research applied to survey methodology, the book is a success. The variety of topics is comprehensive, and the literature overviews are generally very informative. Indeed, the authors provide thorough and understandable discussions of their study problem(s), the experimental designs, and the tested treatments. As a survey statistician – not a survey methodologist – I found the background sections useful; I imagine that they would be likewise invaluable to the novice survey methodologist. The inevitable compromises between implementing (statistically optimal) fully factorial designs and the practically feasible partially factorial designs are thoughtfully presented.

On the other hand, I had difficulty digesting other aspects of the material. First, low response rates and completion rates are endemic throughout the presented case studies, with response rates often well below 50%, generally closer to 25% with one (admittedly extreme) case study examining online panel data with a cumulative response rate of 6.1%. I found it difficult to justify the external validity of such experiments, without making heroic assumptions about nonresponse bias – that cannot be validated in practice. Many of the presented studies offer inconclusive results and attempt to limit their analyses, but there were chapters that presented large tables of contrast tests performed within the same experiment highlighting significant differences verging on p-hacking. The majority of studies accounted for complex designs in their analyses and made use of reweighting to correct for coverage; one chapter explicitly states that nonresponse effects in probability

surveys are “ignorable or addressable through poststratification.” However, technical explanations were fairly limited, and most chapters provided very little concrete guidance in implementation for the novice survey methodologist or survey statistician and only one chapter providing formulae.

Ultimately, the book provides a snapshot of the *current* state of experimental design in probability samples. It is not a textbook, and it does not attempt to provide extensive historic context (for example, the numerous CATI/CAPI survey tests administered by the Census Bureau in the 1990s). In structure, it resembles a dedicated special issue in a survey methodology journal, with thoughtful introductions for each topic provided by the editors. It is well-organized and for the most part, well-written. The discussions of the tested hypotheses are explicitly laid out and grounded in theory. That said, there are severe drawbacks in presenting experimental results from studies plagued by low response rates, even when corrective weighting and appropriate statistical test modifications are employed, especially when analyzed from a total survey error perspective. While advocating for more experiments embedded in surveys, the book also illustrates the challenges in doing it well.