How Mixed-Methods Research Can Improve the Policy Relevance of Impact Evaluations

Burt S. Barnow and Sanjay K. Pandey
George Washington University

In general, impact evaluations rely on quantitative methods, while process evaluations use qualitative methods. In recent years, some studies have combined quantitative and qualitative approaches in what are referred to as mixed methods studies. The purpose of this article is to describe how mixed methods studies can improve the value and policy relevance of impact evaluations, particularly with respect to generalizability to other populations, treatments, and locations.

Evaluations of social programs address a variety of questions and use different methods to conduct the research. Two major categories of evaluations are impact evaluations and process evaluations. Gertler et al. (2011) state that "Evaluations are periodic, objective assessments of a planned, ongoing, or completed project, program, or policy. Evaluations are used to answer specific questions, often related to design, implementation, and results" (p. 230). They distinguish between impact evaluations and process evaluations. "An impact evaluation is an evaluation that tries to make a causal link between a program or intervention and a set of outcomes. An impact evaluation tries to answer the question of whether a program is responsible for changes in the outcomes of interest" (p. 230), and a process evaluation is defined as "…an evaluation that tries to establish the level of quality or success of the processes of a program; for example, adequacy of the administrative processes, acceptability of the program benefits, clarity of the information campaign, internal dynamics of implementing organizations, their policy instruments, their service delivery mechanisms, their management practices, and the linkages among these" (p. 233).

The paper first presents alternative definitions of mixed methods research from the literature, and we then present and justify the definition we prefer from Richwine et al. (2022), followed by a brief discussion of the rationale for conducting mixed methods studies. Greene et al. (1989) provide five purposes for conducting mixed methods studies: Triangulation, complementarity, development, initiation, and expansion; the expansion rationale clearly applies to generalizability. Like other types of evaluations, mixed methods evaluations vary in quality, and we next describe what we consider the best practices in mixed methods studies, which are referred to as canonical mixed methods research in Richwine et al. (2022). The paper then describes the limitations of impact evaluations that only use quantitative methods, and we suggest and implement a framework, described below, to improve the value of impact evaluations by including mixed methods research.

Although impact evaluations provide useful support for local causal claims, they often do not provide a reliable basis for generalizing to other policy contexts (Cartwright 2012, Deaton and Cartwright 2018). This is because quantitative studies using nonexperimental methods often make strong untested or untestable assumptions, and even experimental studies employing randomized controlled trials (RCTs) often have limited external validity (Deaton and Cartwright

2018). A number of scholars have made compelling arguments for adopting a mixed methods approach to overcome the limitations of quantitative impact evaluation approaches (e.g., Bamberger et al. 2026; Fetters et al. 2020; White 2013a, 2013b). This paper carries out a cross-disciplinary review (focusing on leading evaluation and public policy journals) to address two questions. First, we explore the underlying assumptions of impact evaluation methods and the ways in which they limit generalizing across policy contexts. Second, we review mixed methods impact evaluation studies that use the qualitative strand to probe and clarify assumptions in a way that makes it possible to increase policy relevance by generalizing across different populations and policy contexts.

For addressing the first question on surfacing assumptions that limit generalizability of impact evaluation, we will carry out a search in Google Scholar and Web of Science and supplement it with a specific search in leading public policy, evaluation, and economics journals. Our plan is to build on the framework used Richwine et al. (2022) for identifying appropriate articles, which is described in detail in the article's online appendix.[i] For this article, we would supplement the public administration journals reviewed in Richwine et al. (2022) with approximately 30 economics and evaluation journals. Based on this review, we will develop a framework that identifies what limits generalizability of impact evaluation methodologies. To address the second question, we will apply what we learn from answering the first question and carry out a targeted search for mixed methods evaluation studies published over the last 5-10 years that offer exemplars of different ways of improving the generalizability and policy relevance of impact evaluation studies. We are somewhat pessimistic about our ability to identify many impact evaluation articles that use mixed methods for addressing generalizability; none of the 179 articles reviewed in Richwine et al. (2022) cited generalizability as a rationale for using mixed methods. As a fallback strategy, we will explore evaluations funded by the U.S. Departments of Labor, Health and Human Services, and Education, which often include process studies as well as impact evaluations.

CONTRIBUTION

Given the continuing interest as well as vexation with the limited generalizability of impact evaluation studies, our study will contribute to a better understanding of barriers to generalizing from impact evaluation studies and offer insights from our exploration and exemplar studies on how some of these barriers can be overcome.

References

Bamberger, M., Tarsilla, M., & Hesse-Biber, S. (2016). Why so many "rigorous" evaluations fail to identify unintended consequences of development programs: How mixed methods can contribute. *Evaluation and Program Planning*, *55*, 155-162.

Cartwright, N. (2012). Presidential address: Will this policy work for you? Predicting effectiveness better: How philosophy helps. *Philosophy of Science*, 79(5), 973-989.

Deaton, A. & Cartwright, N. (2018). Understanding and misunderstanding randomized controlled trials. *Social Science and Medicine*, 210, 2-21.

Fetters, M. D., & Molina-Azorin, J. F. (2020). Utilizing a mixed methods approach for conducting interventional evaluations. *Journal of Mixed Methods Research*, *14*(2), 131-144.

Gertler, P.J., Martinez, S., Premand, P., Raawlings, L.B., & Vermeersch, C.M.J. (2011). I*mpact Evaluation in Practice*. Washington, DC: The World Bank.

Greene, J.C., Caracelli, V.J., & Graham, W. (1989). Toward a conceptual framework for mixed-method evaluation designs. *Educational Evaluation and Policy Analysis*, 11(3), 255-274.

Hendron, K., Luo, Q.E., & Pandey, S.K. (2018). The state of mixed methods research in public administration and public policy. *Public Administration Review*, 78(6): 904-916.

Richwine, C., Luo, Q.E., Thorkildsen, Z., Chong, N.J., Morris, R., Barnow, B.S., & Pandey, S.K. (2022). Defining and assessing the value of canonical mixed methods research designs in public policy and public administration. *Journal of Policy Analysis and Management.* 41(3), 891-920.

White, H. (2013a). The use of mixed methods in randomized control trials. *New Directions for Evaluation,* 2013(138), 61-73.

White, H. (2013b). Achieving high-quality impact evaluation design through mixed methods: The case of infrastructure. In *Impact Evaluation of Infrastructure Interventions* (pp. 137-150). Routledge.

[i] See https://onlinelibrary-wiley-com.proxygw.wrlc.org/action/downloadSupplement?doi=10.1002%2Fpam.22392&file=pam22392-sup-0001-Appendix.pdf