

Situated Research Design and Methodological Choices in Formative Program Evaluation

Jonathan Supovitz
Associate Professor, Graduate School of Education
Co-Director, Consortium for Policy Research in Education
University of Pennsylvania

Executive Summary

A distinctive component of design-based implementation research is that program developers and researchers join together to develop and refine programs that are intended to improve practice in particular contexts. Even though they share the same goal, the two development partners bring different perspectives and have different roles to play in terms of learning and feedback to inform the program's progression. This chapter focuses on the research partners' choices of research design and methods to inform a program's design in use.

In this chapter I make three particular points. First, I describe the rough trajectory that programs typically follow as they evolve, which I call the intervention development curve. The typical trajectory can be characterized by early instability as program developers work out their program's theory of action, followed by increasing stability, then more planned adjustments as they test hypotheses to improve their program design. In some ways this trajectory is a best-case scenario, because relatively few programs survive long enough to reach maturity; some get stuck in early phases of constant tinkering, while others become obsolete due to either internal or external circumstances. This trajectory is also not time dependent because programs could iterate back and forth along the development curve due to a variety of factors.

My second point in the chapter is that the choices of research method and design need to take into consideration where programs are along the intervention development curve. Early on in implementation, in order to maximize learning opportunities, practitioners and designers of interventions need lots of rich and varied feedback because their theory of action may not yet be well aligned with their theory in use. This means a diverse and multi-method approach to data collection to understand how – by what processes and mechanisms - a program influences (or not) its target audience. Because the program is young, its developers are learning as they go and

they are making adjustments as they learn from their own experiences, from evaluation data, and from the contexts they are operating within.

Thus, methodological considerations, and the questions that precede them, also need to match where interventions are on the development curve. Researchers have a range of available methodological techniques, and we should choose the methods that provide the most timely and nuanced feedback possible. Interventions at different places along the intervention development curve will need feedback at different speeds and of different types. For example, very young interventions will need quicker feedback than ones in their adolescence because their adjustment cycles are likely to be shorter. Of course, this is dependent on the scope and timeline of reforms as well.

By contrast, interventions in later phases of development, with more established processes, can benefit from more carefully detailed research methods that focus on particular questions of interest which require different methods of inquiry. For example, particular aspects of a program may emerge as areas of particularly high variation in implementation and developers may want to understand what is it about this aspect of the program that seems difficult to implement in some contexts, while not in others. This might require customized observations or more targeted interviews. Or developers might have a hypothesis about the introduction of a particular adjustment of their model and may seek to test this particular approach within the more regular implementation of their design using a particularly focused set of survey questions for participants. Questions like this, based on accrued knowledge, may contribute to methodological choices at middle and later stages of the intervention maturation process.

Research design considerations also are best made with the intervention development curve in mind. Typically, researchers choose research designs based upon criteria of feasibility and rigor. That is, a design is chosen based on the circumstances – what is feasible in a particular situation – or researchers seek to adopt the most rigorous design, regardless of the stability and maturity of the intervention. Particularly in today’s world of ‘evidence-based research’, where experimental research carries the most caché, there is a lot of incentive to prematurely move towards designs that are inappropriate for the developmental phase of the intervention.

DBIR considerations might affect our decisions about design choices. The traditional application of the Campbell and Stanley typology would call for the highest application of research design to identify program impacts. Cook (2002) and others argue that evaluation should always aspire towards experiments, which they contend are the most accurate and cheapest (in the long run) choices. This is because, they argue, we get closest to an accurate assessment of program effects with an experimental approach.

However, taking into consideration where programs are on the intervention development curve might lead to different conclusions. In the early stages of a program’s development, when a program is being continually refined and the intervention itself is not stable, then quick data-based feedback is highly valued and should drive the choice of research design. In these circumstances, given typical tradeoffs between design complexity, the time it takes to set up a study, the availability of an excess pool from which to select a comparison group, and the time for analysis, early studies are likely to be pre- or quasi- experimental. Since programs are likely in flux in their early phases, the goal is not to gather stable and robust measures of program impacts, but to understand how programs are playing out. In these cases non-equivalent samples, or even samples of convenience allow for counterfactuals that can provide rough approximations

of the likelihood of program effects, are sufficient. When randomization is easily feasible in these early trials, it still is a preferable approach because it allows for more reliable learning opportunities, but the tradeoffs against expediency are unlikely to be warranted given the intervention development process.

Experimental designs can be useful during the development process, once programs begin to stabilize, if their purpose is to provide evidence of how a program works, rather than just if it works. An experimental situation has the advantage of providing stronger information back to program developers about how the program works – that is, if there are differences in intermediate outcomes. This is because the program design team can have more confidence that the intervention is producing these differences rather than worrying that they are misreading differences that are caused by non-equivalent subjects or situations. But the over-riding purpose of experiments in these stages is to inform program improvement, not provide stable measures of ultimate impact. Thus formative experimentation has a different purpose than summative experimentation.

My third point is that, as programs develop, the situation of implementation plays an important role in the evolution of interventions, enabling and constraining how the intervention matures, and we need to reconsider the role of situational influences in formative evaluation research. I distinguish situation from context because we associate context with the background of a program in its development, but I contend that program and situation are inextricably entwined and cannot be disentangled. The evidence from over 30 years of implementation research suggests that situation plays a regular and influential role in intervention implementation and impacts (Berman & McLaughlin, 1976; Supovitz & Weinbaum, 2008).

This influence has important implications for program design, both within particular settings and across settings. By increasing research attention to the interactions between situation and program, we can understand how situations are likely to adjust design and potentially incorporate this learning into the design itself. Therefore, situation should be treated as more than an ancillary factor in considerations of research design and methodological choices. This can be done by incorporating knowledge of where interventions are on the intervention development curve into design and methodological research choices, which encourages consideration of situation, because this, in and of itself, is part of the situation of a design. Further, regular application of qualitative data collection and analysis that attends to the interactions between situation and program can reveal the ways that the context is facilitating or impinging on the design. And finally, quantitative analyses should pay more attention to developing variables of situation to incorporate into statistical models. Together, mixed method design and analyses that consider situation as an important factor in understanding implementation and impacts are more likely to reveal the ways that interventions influence their intended audiences.

DBIR allows us to rethink the traditional relationships between practitioners and researchers, the choices of design and methods in consideration of program maturity, and the interactions between programs and their situation. By viewing research as an integral part of a program's development; by making design and methodological choices in consideration of where programs are in their development; and by considering that the situation in which programs evolve as a potential source of change in the nature of the program itself, we alter fundamental perspectives on how research can best contribute to the steady work of building robust programs for educational improvement.