

Chapter Two

How We Got Here: A Summary of the Evaluation Landscape, History, Paradigms, and Balancing Acts

By now you have a good sense of where we stand on evaluation. We take it seriously and follow our philosophy with investments. But where do we stand in the broader landscape? How did we arrive at our particular values and viewpoints about the usefulness and power of evaluation? How did we find our place in the world of evaluation? It takes a short history lesson and some understanding of the art and science of research paradigms to answer these questions.

The Evaluation Landscape

The original mission of program evaluation in the human services and education fields was to assist in improving the quality of social programs. However, for several reasons, program evaluation has come to focus (both implicitly and explicitly) much more on *proving* whether a program or initiative works, rather than on *improving* programs. In our opinion, this has created an imbalance in human service evaluation work—with a heavy emphasis on proving that programs work through the use of quantitative, impact designs, and not enough attention to more naturalistic, qualitative designs aimed at improving programs.

We discuss two reasons for this imbalance:

- the historical context of program evaluation in the U.S.; and
- the influence of the dominant research paradigm on human services evaluation.

Historical Context of Evaluation in Human Services

Although human beings have been attempting to solve social problems using some kind of rationale or evidence (e.g., evaluation) for centuries, program evaluation in the United States began with the ambitious, federally funded social programs of the Great Society initiative during the mid- to late-1960s. Resources poured into these programs, but the complex problems they were attempting to address did not disappear. The public grew more cautious, and there was increasing pressure to provide evidence of the effectiveness of specific initiatives in order to allocate limited resources.

During this period, “systematic evaluation [was] increasingly sought to guide operations, to assure legislators and planners that they [were] proceeding on sound lines and to make services responsive to their public” (Cronbach et al., 1980, pg. 12). One lesson we learned from the significant investments made in the 1960s and ’70s was that we didn’t have the resources to solve all of our social problems. We needed to target our investments. But to do this effectively, we needed a basis for deciding where and how to invest. “Program evaluation as a distinct field of professional practice was born of two lessons...: First, the realization that there is not enough money to do all the things that need doing; and second, even if there were enough money, it takes more than money to solve complex human and social problems. *As not everything can be done, there must be a basis for deciding which things are worth doing.* Enter evaluation” (Patton, 1997, p. 11).

Today, we are still influenced by this pressure to demonstrate the effectiveness of our social programs in order to ensure funders, government officials, and the public at large that their investments are worthwhile. In fact, since the years of the Great Society, pressure to demonstrate the worth of social programs has increased. Limited resources, increasingly complex and layered social problems, the changing political climate, and a seeming shift in public opinion about the extent to which government and other institutions should support disadvantaged or vulnerable populations have shifted the balance even further to an almost exclusive focus on accountability (prove it works), versus quality (work to improve).

The Scientific Method as the Dominant Evaluation Paradigm

A second factor leading to an emphasis on proving whether a social program works is the influence of the scientific method on human-services evaluation. When most people think about program evaluation, they think of complex experimental designs with treatment and control groups where evaluators measure the impact of programs based on statistically significant changes in certain outcomes; for example, did the program lead to increases in income, improved school performance, or health-status indicators, etc.?

The scientific method is based on hypothetico-deductive methodology. Simply put, this means that researchers/evaluators test hypotheses about the impact of a social initiative using statistical analysis techniques.

Perhaps because this way of conducting research is dominant in many highly esteemed fields and because it is backed by rigorous and well-developed statistical theories, it might dominate in social, educational, and human-services fields—

members of which often find themselves fighting for legitimacy. In addition, this way of doing research and evaluation is well suited to answering the very questions programs/initiatives have historically been most pressured to address: Are they effective? Do they work?

The hypothetico-deductive, natural science model is designed to explain what happened and *show causal relationships* between certain outcomes and the “treatments” or services aimed at producing these outcomes. If designed and conducted effectively, the experimental or quasi-experimental design can provide important information about the particular impacts of the social program being studied. *Did the academic enrichment program lead to improved grades for students? Or increased attendance? Ultimately, was it effective?* However, many of the criteria necessary to conduct these evaluations limit their usefulness to primarily single intervention programs in fairly controlled environments. The natural science research model is therefore ill equipped to help us understand complex, comprehensive, and collaborative community initiatives.

Balancing the Call to Prove With the Need to Improve

Both of these factors—the historical growth in the pressure to demonstrate effectiveness, and the dominance of a research philosophy or model that is best suited to measure change—may have led many evaluators, practitioners, government officials, and the public at large to think of program evaluation as synonymous with demonstrating effectiveness or “proving” the worth of programs. As a result, conventional evaluations have not addressed issues of process, implementation, and improvement nearly as well. And they may very well be negatively impacting the more complex, comprehensive community initiatives (like many of those you operate in your communities) because these initiatives are often ignored as unevaluatable, or evaluated in traditional ways that do not come close to capturing the complex and often messy ways in which these initiatives effect change (Connell, Kubisch, Schorr, Weiss, 1995; Schorr and Kubisch, 1995).

Clearly, demonstrating effectiveness and measuring impact are important and valuable; yet we believe that it is equally important to focus on gathering and analyzing data which will help us improve our social initiatives. In fact, when the balance is shifted too far to a focus on measuring statistically significant changes in quantifiable outcomes, we miss important parts of the picture. This ultimately hinders our ability to understand the richness and complexity of contemporary human-services programs—especially the

system change reform and comprehensive community initiatives which many of you are attempting to implement.

Following are some of the many consequences of operating within a limited evaluation framework:

Consequence 1. We begin to believe that there is only one way to do evaluation.

Most people (even those trained in research and evaluation methods) don't realize that methods employed, such as an experimental design, are part of larger world views or paradigms about research. These paradigms are based on different assumptions about:

- What is the nature of reality?
- How do we come to know something?
- What should be the relationship between the researcher/evaluator and the participants in the evaluation process?

The dominant research paradigm described above (hypothetico-deductive), derived from medical and other natural science disciplines, is one such paradigm, but there are others. When one research paradigm begins to dominate a field, it becomes easier to forget that other paradigms—which address different goals and questions—also exist.

Patton explains the effect of forgetting paradigms in this way:

The very dominance of the hypothetico-deductive paradigm, with its quantitative, experimental emphasis, appears to have cut off the great majority of its practitioners from serious consideration of any alternative evaluation research paradigm or methods. The label “research” [or evaluation] has come to mean the equivalent of employing the “scientific method” of working within the dominant paradigm (1997, pp. 270-271).

In other words, people begin to believe there is only one right way of doing evaluation.

Consequence 2. We do not ask and examine equally important questions. We have already discussed how the dominant research paradigm is suited for addressing certain impact questions—the very questions that, historically, social programs have been pressured to address. However, while it brings certain aspects into focus, it misses other important dimensions of the program.