BOOK REVIEWS / COMPTES RENDUS DE LIVRES


Reviewed by Kaireen Chaytor

This book makes an important contribution to evaluation practice in Canada. Stake’s discussion of standards-based and responsive evaluation addresses many issues, particularly approaches to fulfilling external criteria while attempting to reflect local or relevant experiences.

CRITERIAL AND INTERPRETIVE EVALUATION

Evaluations for publicly funded programs usually have external guidelines for the method and scope of the evaluation. The guidelines must be followed even though one of the overriding guidelines is that evaluations are program managed. The Treasury Board Secretariat indicates that “[p]rogram managers should ensure program staff and partners collect performance information … and demonstrate results through timely evaluation activities and reporting” (Treasury Board of Canada, 2005). There is often a tension around the extent to which the evaluation responds to external criteria, or reports on the experience and outcomes of a program. This central tension of evaluation practice is the theme of this book. Even though Stake states that we evaluators tend to be centred on either “standards-based data or responsive data, one or the other, seldom both in equal helpings” (p. 173), he successfully shows the benefits of the two perspectives working together. In other words, the differences do not have to be oppositional. Evaluators can use Stake’s work as they follow Canadian federal government guidelines and report on the implementation of programs.

An example of the use of standards-based and responsive evaluation is found in the work of the Centre for Health Services and Policy Research (Watson, Broemeling, Reid, & Black, 2004). Primary health care, a broad and diverse concept, is brought to a standards-based approach when it follows the guidelines of the Treasury Board Secretariat.

Corresponding author: Kaireen Chaytor, Chaytor Consulting Services, 2979 Oxford Street, Halifax, NS B3L 2W3; <chaytor@hfx.eastlink.ca>
EXPLANATION FOR PRACTICE

Stake brings a unique richness to his explanations of terms and concepts, demonstrating their use for both critical and interpretive evaluation. He argues, for example, that “quality” is what we experience, as we experience it.

This theme is continued in Chapter Two as he explains models, dispositions, and roles. Stake tends not to use the term “model.” He finds it too rigid or too pre-ordinate for evaluation. His discussion of disposition gives insight into what we bring to evaluation — the all-important starting point. Roles are also important to have established at the outset. Understanding what we are doing and why we are doing it is an important matter to be clarified.

Stake’s explanation of standards-based evaluation is not as comprehensive as his other discussions. His reference to standards is based more on program standards than on guidelines for evaluation as a standard. Evaluators are confronted with a variety of standards partly determined by the roles described in Chapter Two.

In Chapter Four Stake explains responsive evaluation, stressing it is not necessarily the same as participatory evaluation. He notes the need to respect multiple realities, the use of issues as conceptual organizers, and the need to find stories that develop an issue. Responsive evaluation can be difficult to understand and use. His explanation is very helpful, with consistent themes of getting deeply acquainted with the actual operation of the program and answering the question: Is this particular program succeeding?

Stake urges the reader to look at the concepts used when data gathering, reminding us that evaluation is mainly the search for quality. Chapter Five is a hodge-podge, but the discussion of data gathering as it pertains to criterial and interpretative evaluation forces the reader to think about both the use of systematic data gathering and the extent to which some of the “answers are going to come out of your own head” (p. 110). In the discussion on analysis and synthesis, Stake wisely asks the reader to approach the topic conceptually and thus provides wise advice. For instance, he notes his fear of simplicity in the analysis, since programs are not simple, and asks evaluators to be clear about whether they are evaluating the whole or the parts of a program, since the whole may be different from the sum of its parts. He challenges evaluators to “tell their readers something of what they are not including” (p. 169), since “people rely on what they infer from their experiences” (p. 175). Audiences do their own thinking and the evaluator should help them do it well.
In keeping with the theme of bringing together the two perspectives, Stake notes that “the client and provider of funds have the final say” (p. 190). He urges evaluators to be clear at the outset, to find a balance between insiders and outsiders (using American Evaluation Association guiding principles), and to help see the perspective of various groups. He again brings out the theme that the evaluator’s job is the identification of program quality. He notes how getting started is very important. The evaluator must recognize and be in a position to develop the issues, since “good issues or research questions are critical to good evaluation studies” (p. 225).

Chapter Nine continues with the theme that evaluation is a search for evidence and a search for quality. Stake invites us to question the use of randomized experiments when studying a single program, whether we are tied to realities of practice as opposed to rationality of science, or are showing the different interpretations of quality and adding resolution where we can. Stake’s summary chapter, “Doing It Right,” brings together many of the qualities of standards-based and responsive evaluation. He identifies the need for an evaluation literature that includes codes and rules and principles, but argues the standards should not be used to oversimplify. His distinction of evaluation when it pertains to personnel matters or to a product is helpful in understanding the scope of program evaluation. His critique of measures of performance is very relevant to criterial and interpretative tensions in the requirement for evaluation frameworks from all federal departments. An interesting example of blending the two perspectives is his suggestion to “follow IRB protections and create your own as well” (p. 268).

Stake ends with his own reflection on evaluation. He wants us to get deeply acquainted with the program. He sees evaluation as searching for quality, then reporting evidence of that quality to others. “Evaluation fundamentally ends with the realization of evaluation quality” (p 174). He recognizes the many perceptions of quality and wants us to think constantly about whether the sum of the quality of the parts represents well the quality of the whole. He suggests “quality was born when it was seen” and “seeing quality is a human construction, whether measured or felt” (p. 286). One short paragraph summarizes well the theme of the book:

In all that we ponder, there will be help from criterial and interpretive thinking. In all that we evaluate, there is need for both standards-based and responsive evaluation. Together they seldom add up, but with their sepa-
rate insights, together they cause us to move more sens-itivity, to think more deeply, to report more cautiously, and to be more committed to the representation of quality. (p. 284)

STYLE, STRENGTHS, AND WEAKNESSES

I have employed a responsive approach in evaluation based on earlier works by Stake, and was also helped over the years by discussions with the author. This work is needed to continue the discussion.

Because of the insights this book provides for practice, I hope it is read widely. However, I have a few concerns about using the book as a teaching text. The extensive use of school examples may not allow the reader to see the application to public-sector management.

Stake need not apologize for his lack of suggestions for technique. His focus on clarifying the starting points is valuable to the reader. He includes numerous suggestions for additional resources. His commentary throughout the book separates building evaluation capacity from either standards-based or responsive evaluations. His position is a little surprising, given his own background as an educator.

The book has several strengths that compete for first place. One is the connection to and use of the evaluation literature, so we see both the parts and the whole very well. Another great strength is the engaging style. A third is Stake’s argument that we can bring standards-based and responsive evaluation together in a meaningful way, a desirable objective for practice.

REFERENCES


Many texts and journal articles present a specific approach to evaluation theory and practice. This text compares current evaluation theories and practices, with contributions from many of evaluation’s leading theorists and practitioners. In the 26 chapters, some authors use a conversational writing style, while others follow a conventional academic writing approach. Alkin’s text is structured around three categories that help the reader understand individual theories and the relationships between theories. An evaluation theory “tree” is used, with each branch representing key theoretical directions: use, methods, and judgement/valuing. He then situates each of the contributors in one of these three branches, with the trunk representing social inquiry, and accountability and control — two foundation elements for evaluation. It is interesting to note that several of the invited contributors criticize the tree metaphor as simplistic and an “inaccurate reflection of influences” (p. 218).

Alkin bookends the text with strong introductory and summary chapters. The text begins with an excellent précis of the main theoretical perspectives of the text contributors. This piece provides the rationale for locating each contributor in one of the three tree branches. With this chapter as context, we then move on to the first branch, “Methods.” George Madaus discusses Ralph Tyler’s many contributions to educational program evaluation, such as complementing basic student testing with consideration for behavioural objectives. William Shadish and Jason Luellen then review Donald Campbell’s perspectives on evaluation research methods. We see that Campbell was much more than a logical positivist and an advocate of experimental methods; indeed, he had an appreciation for cultural relativism.

Thomas Cook continues our examination of evaluation methods, specifically causal generalization, by exploring the impacts of Donald Campbell and Lee Cronbach on Cook’s evaluation research. This was a fairly dense and rather demanding chapter that would appeal to

Corresponding author: Mark Seasons, PhD, Associate Dean (Undergraduate Studies and Educational Liaison), Faculty of Environmental Studies, University of Waterloo, Waterloo, ON, N2L 3G1; <mseasons@fes.uwaterloo.ca>
dedicated methodologists. Robert Boruch then reflects on his career in evaluation, noting the influences of many theorists, including Peter Rossi and William Shadish among others. He is the first of many in this text to offer, in a concise fashion, strategies for effective evaluations.

Peter Rossi follows with reminiscences of his education and career, noting that he sees himself as “a working evaluation researcher,” rather than an “evaluation theorist” (p. 126). Rossi also notes that evaluation is not tied to a specific social science, and that evaluation benefits from cross-disciplinary exchanges of knowledge and practice. He says that “evaluation is a servant to incremental social change” rather than an instrument for social change, which he considers the responsibility of politicians (p. 129). Rossi is followed by Huey Chen, who examines the roots of theory-driven evaluations. Chen explains that an effective evaluation must comprise future directedness, scientific and academic credibility, and a holistic approach. He stresses that the “ideal evaluation achieves both high scientific and high stakeholder credibility” (p. 134), a balance that is often easier sought than achieved. Chen offers a practical taxonomy of program evaluation that is explained in tables and process diagrams. He proposes a contingency approach to evaluation: one’s choice of evaluation methods and strategies must be situational, rather than “universalist.”

Carol Weiss does not see herself as an evaluation theorist. Rather, Weiss thinks of her ideas as “practical beliefs about how to do evaluations, with sensitivity to the context, demands, and constraints of the immediate situation” (p. 153). She wants to see lessons learned from practice — learning by doing. Weiss’s position is clear: evaluation is “not only a research activity but also a political activity” (p. 157). She goes on to point out that while decision-makers rarely make immediate use of evaluations, she has been “impressed with the longer-term percolation of ideas from evaluation into organizational discourse” (p. 161).

The “Methods” section concludes with a chapter by Jennifer Greene, who reviews Lee Cronbach’s perspective on evaluation. Cronbach believed that methods are a necessary but insufficient condition for effective evaluations. He is more interested in positioning evaluation to advance social well-being by using evaluation to facilitate “broad, extensive discourse” (p. 170).
The next section explores the concept of “Valuing,” the assignment of value or making judgements and the role this has for evaluation practice. Michael Scriven argues that evaluation is much more than a bundle of methods and techniques. It is a discipline, part of a group of “trans-disciplines” which contribute to the body of knowledge in their own and complementary disciplines (p. 187). Elliott Eisner then reinforces the common message that quantitative methods are often necessary and useful, but need to be complemented by qualitative approaches. He explains that “number needs a referent. Those referents are often qualities, and those qualities are often extremely difficult to quantify” (p. 200).

Robert Stake follows with an argument for “responsive evaluation,” which means searching for the often hidden explanations for program performance. He advocates use of qualitative methods, such as the case study, and quantitative methods. Ernest House notes that evaluators “cannot conduct evaluations without deciding how to derive criteria (values) for the evaluation, what data collection methods to use, and what to do with the findings” (p. 219). House calls for an approach to evaluation that is inclusive, dialogical, and deliberative, which he sees as mutually reinforcing elements. Lincoln and Guba extend this discussion by reminding us that “the adoption of terms is a fragile, unstable, and highly value-laden enterprise” (p. 226). They acknowledge that the identification of causation is often highly problematic, and anticipate an evaluation practice that is more pluralistic, socially aware and inclusive (p. 239).

The final section concerns “Use” of evaluation — how evaluation is applied in organizational decision-making. Daniel Stufflebeam leads off with a description of the CIPP (context, input, process, and product) evaluation model. This is an integrative approach that has its origins in mid-1960s education program evaluation. The CIPP pulls together the various key elements of a typical evaluation process. In a short article, Joseph Wholey explains his commitment to the use of evaluation to improve program performance. He is especially interested in evaluations that are time- and cost-effective and relevant to stakeholders (p. 268). These important messages are complemented by Wholey’s reflections on a long and distinguished career in evaluation. Michael Patton continues the analysis of evaluation application, noting that “use concerns how real people in the real world apply evaluation findings and experience the evaluation process” (p. 277). This is a highly personal and situation-specific experience, grounded in decision-making reality.
Marvin Alkin echoes Patton’s position by stating that “while an evaluator may have a preferred mode of conducting an evaluation, he or she must be cognizant of the organizational reality” (p. 297). In this sense, the evaluator must work with evaluation clients and stakeholders to design and apply a relevant, credible evaluation process. David Fetterman discusses empowerment evaluation, wherein program participants conduct their own evaluations with an external evaluator as coach. Empowerment in this context involves politics, but more important, personal and psychological empowerment (p. 311). His ideas, once deemed radical, are becoming more commonplace and widespread. A similar theme is pursued by J. Bradley Cousins, who argues that “a pivotal role for evaluators … is to foster informed judgment making on the part of nonevaluator stakeholders” (p. 321). Cousins stresses the need to understand and factor in organizational culture in evaluation process design and management. Jean King suggests that “all evaluation is participatory, because in every case, an evaluator must interact with someone … to frame the evaluation task” (p. 337). Thus, effective communications and interpersonal skills are essential. Much of the evaluator’s work is to create a culture that is receptive to, and uses, evaluations.

Similarly, Hallie Preskill emphasizes that organizational context is a critical determinant of evaluation success. According to Preskill, evaluation should be “ongoing, reflexive, and embedded in organizational practice” (p. 348). John Owen argues that evaluation use is enhanced by “extended up-front negotiations” with key stakeholders, before data are collected and the formal evaluation process begins.

Researchers, students, and practitioners will find this text useful because it explains and sets in context various theoretical perspectives. Many of the messages will not be new to well-informed evaluators. However, the repetition of key themes by many authors (e.g., context is critical, inclusionary processes are healthy, organizational culture must be considered) tells us that these views are well-accepted and should guide evaluation practice.