

Review of Evaluation: The International Journal of Theory, Research and Practice: Volumes 11(2), 11(3), and 11(4)

Daniela C. Schröter

Western Michigan University—The Evaluation Center

The following contribution provides a summary of Volumes 11(2), 11(3), and 11(4) of *Evaluation: The International Journal of Theory, Research and Practice* (2005). The central motif of issue 11(2) is learning, that is, learning from theories of change, “accountability (f)or learning,” learning aided by cluster evaluation, facets of organizational learning, and learning from the use of certain research designs. Issue 11(3) centers on issues and approaches of (primarily policy) evaluation in different countries (Belgium, the United Kingdom, Norway, and Japan) and contexts (development evaluation, educational evaluation, social service evaluation). Issue 11(4) stresses evaluation designs, methodological approaches, evaluator competencies, and problems of bottom-up versus top-down evaluation. The following briefly summarized the contributions of each volume.

Volume 11(2)

The first two articles stem from authors from the United Kingdom and deal with the trade-offs between linear simplicity and non-linear complexity of theories in evaluation. First, Rick Davies presents the second of a series of articles published in *Evaluation*. He builds on his 2004 contribution which provided a “methodological approach to representing theories of organizational change” (Schröter, 2004, p. 116) from hierarchies through heterarchies to business process maps. His recent article focuses on the reciprocal nature of change by moving away from considering only unidirectional change, since “most change is a two-way process” (p. 134). As such, he promotes networks and their interconnectivity on various scales, including all levels inside and among organizations and specifically recommends the use of networks over logical frameworks because they can (i) depict hierarchical structures, (ii) take a personalized view by using actors as starting points, (iii) be more altruistic, (iv) recognize limitations of individuals, and (v) identify individuals who have certain information. Davies suggests that there may be a third contribution that will provide guidelines “detailing how a network perspective can be systematically operationalized within the design, monitoring and evaluation stages of development programs” (p. 147).

Second, Mhairi Mackenzie and Avril Blamey discuss theory-based evaluation with a particular focus on (i) the issue of articulating theories of change (ToC) prospectively and (ii) key strengths of theory-based evaluation in social service settings suggested by Connel and Kubisch (1998): (a) increased precision in program planning, (b) assistance in making choices regarding evaluation questions and methods, and (c) decrease of intricacies with making causal claims. Based on

case studies, the authors conclude that the ToC approach has utility, but that the strength identified for theory-based evaluation would not be limited to that approach and moreover, would be highly dependent on the quality of the ToC revealed. Thus, many shortcomings inherent in theory-based evaluation are confirmed, e. g., time extensiveness, problems with potential misspecifications, avoidance of unanticipated outcomes, questionable causal claims, and confusion of evaluator roles (c.f., Coryn, 2005a & b).

In the third contribution, Markku Lehtonen from France investigates the extent of success of the Organization for Economic Cooperation and Development (OECD) in developing the Environmental Performance Review (EPR) program by pooling two divergent evaluation approaches; (1) accountability-oriented models derived from the movement toward “New Public Management” and (2) models stressing the importance of learning that grew out of the “search for new policy instruments for managing complexity, uncertainty, and plurality of values in the pursuit of sustainable development” (p. 169). As argued, these two approaches are subsumed under the three primary goals of the EPR. Thus, learning would be inherent in the goals of supporting national governments in conducting evaluation and discourse among member countries, while accountability constitutes an explicitly stated goal. The influence of evaluation would be inherent in three potential outcomes: (1) decisions and actions, (2) shared understanding, and (3) change in legitimacy of actions. Lehtonen goes on by construing the concepts of learning from evaluations and accountability and then specifically discusses peer review and its potential “in terms of simultaneously promoting both learning and accountability notably in international contexts” (p. 184).

Philip Potter centers his article on the utility of cluster evaluation to aid inter-project and inter-policy learning and communication as exemplified on the EU Equal program in Germany. Potter starts out by discussing contextual factors of evaluands suited for cluster evaluation, i.e., stakeholders on multiple levels and sites in varying contexts and specifically contrasts cluster evaluation with multi-site evaluation, meta-analysis, and critical review. Then he describes the evaluand in context and argues that by encouraging the use of interactive and dialogue-intensive techniques of communication and audience-specific methods of dissemination, cluster evaluation would promote communicative learning processes among program participants on the case and policy level. Due to the potential of cluster evaluation to “inform knowledge generation and management in organizational settings involving theme-related networks of heterogeneous – and autonomous – local interventions with multiple goals” (p. 203), its principles and techniques would be specifically beneficial in evaluating “European programs with high heterogeneity and complexity” (ibid).

Henrik Schaumburg-Müller examines how evaluation findings are organized and used to contribute to organizational learning and potential changes in policy and operation for implementing foreign aid initiatives on cases of the Danish national aid agency Danida and the OECD. He concludes that both rational and learning organization perspectives may only have limited importance since multiple processes, mechanisms, and sources of information may influence the actual acceptance and corresponding implementation of evaluation results. In essence, interests of key stakeholders in organizations may be divergent from evaluation results, and political influences among other criteria may overpower findings from evaluations and determine decision making.

David H. Greenberg and Stephen Morris critically examine “advantages and disadvantages of social experimentation in the context of a specific random allocation demonstration that is the largest yet undertaken in the United Kingdom—the Employment Retention and Advancement (ERA) Demonstration” (p. 224). The authors conclude that in social policy intervention settings, such as the one exemplified, experimental designs are superior to studies using pre/post designs, matched observations, or comparison groups at no increased cost or time while assuring greater validity in almost all instances. However, they also argue that non-experimental designs may be more feasible in some circumstances and provide information that is not to be ascertained through random allocation (e.g., about relative effectiveness of program components). Finally and most importantly, the use of experimental designs may not be ethical in some cases.

Volume 11(3)

In the first article of Volume 11(3), Frédéric Varone, Steve Jacob, and Lieven De Winter discuss “the institutionalization of policy evaluation in Belgium” (p. 253) and point out three characteristics which thwart the development of an evaluation culture in the country. These are (i) partitocracy, meaning that policy making processes are determined by political parties, (ii) confined parliamentary control of the executive, and (iii) federalism. To overcome these barriers and build policy evaluation capacity in Belgium, the authors suggest (a) the development of an epistemic evaluation and research community, (b) invigorating parliamentary and supervisory roles, (c) the initiation of common efforts at multiple levels, and (d) the instigation of discussion among political parties.

Second, Rick Davis revisits “realistic evaluation” and highlights limitations of the approach in its implementation in local English Best Value Reviews. The author specifically assessed five hypotheses and suggests: (i) In realistic evaluation contexts are often not sufficiently assessed and inadequately specified, thus cannot be generalized. However, inventories of “basic elements in widely defined contexts” could contribute to realistic evaluation efforts. (ii) Mechanisms are often under-researched and thus skewed in nature. (iii) Unanticipated goals and outcomes and related mechanisms are often neglected in black box evaluations, thus, ignoring causal links. (iv) Models may be simplified or be constrained by causal lock-in. (v) The evaluand’s boundaries and locus needs to be clearly specified. For further information, a summary and critique, of *Realistic Evaluation*, see Coryn’s (2005) contribution in JMDE 03.

Monica Rolfsen and Hans Torvatn provide insight into communicative processes experienced during a formative evaluation in Norway and suggest five strategies for successful communication and related trust building. (i) Expectation for the evaluation team must be clarified. (ii) Diversity should be utilized as a foundation for innovation, meaning that diverging values and backgrounds should be maximized to embrace ideas not to be thought of otherwise. This requires listening skills and attempting to understand the other. (iii) Good (trustful) communication processes through a “shared language” enhance understanding. (iv) Rolfsen and Torvatn suggest the use of “functional argumentation,” a method that uses the evaluand’s own expressions (e.g., goals) as a basis for discussion. This method is contrasted to normative argumentation which is based on the evaluators’ values and thought to be less likely to generate buy-in. (v) Finally, the authors stress the

importance of continual evaluator-client communication in the form of feedback and dialogue.

Robert Picciotto examines the coherence of policy in the international development context to enhance “understanding of ‘synergies’ and ‘trade-offs’ between the different policies that affect poverty reduction in developing countries” (Stern, p.252). After providing a synopsis of the rationale for policy coherence for development, Picciotto reveals how the concept is rooted in: the European Union, the United Nations, International Financial Institutions, NGOs, and the OECD. Thereafter, he suggests concepts and techniques for evaluating policy coherence under specific consideration of evaluability, method, governance, and the involvement of developing countries.

Jenny Owen, Tina Cook, and Elizabeth Jones discuss the participatory evaluation of the “Early Excellence Initiative,” an educational program in the United Kingdom targeted at families with young children. As a result of the evaluation and specific consideration of interactions between national and local-level evaluators as well as program staff (practitioners), the authors observed two major shifts in the evaluation. First, data collection and reporting requirements of evaluations were moving from more participatory-oriented mechanisms toward accountability-oriented performance management systems. Second, participation of local practitioners shifted to a more marginalized role in the evaluation process. These shifts toward evaluation for accountability purposes only would hinder evaluation that supports learning. Thus, the authors suggest a framework that does not only look at what works but also why it works.

Yuriko Sato introduces a theory-based policy evaluation method by adapting a goal attainment model according to a Project Design Matrix to achieve a Policy Evaluation Matrix (PEM) with which she investigates the impact, effectiveness, efficiency, and relevance of Japan's Foreign Student Policy toward Thailand (from 1954 to 2001). As a result of implementing her proposed method, Sato elucidates the following advantages and weaknesses: (i) The PEM has utility for ascertaining program or policy theory as it contains a wide range of influencing elements. It is also a useful tool for communication among varying stakeholders. (ii) As with most theory-driven evaluation approaches, the matrix fails to illuminate unintended effect. (iii) PEMs have been proven to support the measurement of impact, effectiveness, efficiency, and relevance, though the assessment of sustainability has still to be investigated. (iv) Finally the author suggests that the evaluation example can be perceived as a model Official Development Assistance evaluation.

Volume 11(4)

Issue 11(4) begins with a contribution by June Lennie who provides an account of implementing the Learning, Evaluation, Action, and Reflection for New Technologies, Empowerment and Rural Sustainability (LEARNERS) framework to elucidate empowering and disempowering impacts on rural communities in Australia. Strengths and weaknesses of the capacity building process were derived through ongoing metaevaluation and a combination of participatory evaluation, participatory action research, and feminist theories. Impacts were grouped into social, technological, political, and psychological empowering and disempowering effects. Reasons for disempowering outcomes suggested, include power inequities,

differing levels of knowledge and understanding, and conflicting value schemes of participants and evaluators among others.

Laurie Stevahn and Jean King describe means for constructive conflict management in program evaluation. Based on conflict strategies theory and constructive conflict resolution theory, the authors derive a set of evaluator competencies to resolve conflict constructively. The competencies include (i) cooperative goals structuring skills and (ii) integrative negotiation skills. The first skill set comprises abilities to create “positive interdependence in social situations and participatory evaluation tasks” (p. 420) common to all evaluation regardless of the degree of participation. The second emphasizes a set of skills that can be sub-grouped under the domain of interpersonal competencies and which lead to win-win situations for all participants, because “all disputants aim to maximize everyone’s outcomes” (p. 422). The utility of these competencies for evaluators and evaluation can be seen in enhanced mutual understanding and communication between evaluators and stakeholders as well as in potential reduction of evaluation anxiety.

Sandy Oliver, Angela Harden, Rebecca Rees, Jonathan Shepherd, Ginny Brundon, Jo Garcia, and Ann Oakley provide a framework for conducting systematic reviews using a mixed methods approach “to move beyond the study of ‘what works’ to questions concerning what works for whom, in what context and why” (p. 441). The framework is intended to improve information for policy makers and others who require syntheses from research-based publications.

Hanne Foss Hansen’s contribution promotes two key considerations to be weighted for choosing an appropriate evaluation design: (i) evaluation purpose, (ii)

characteristics of the evaluand, and (iii) the problem addressed by the evaluand. The author differentiates between two purposes of evaluation, i.e., formative and evaluative, in Hansen's words "learning" and "control." If the purpose of the evaluation would be "control," results-based evaluation models to assess program goal attainment would be more appropriate. If learning is the primary evaluation purpose, then process models would be adequate. Moreover, these formative efforts would be more participatory than the summative evaluation of goal attainment. With respect to the evaluand's characteristics, Hansen suggest to considering (a) potential alternatives within the realm of the characteristics (i.e., "possibility reasoning," p. 453) and (b) what justifies a certain design's use within the context of the evaluand ("legitimacy and justice reasoning," *ibid.*). She argues that one should use economic evaluation models (i.e., forms of cost analysis) only if a complete program theory and program objectives are available. If only the objectives are clearly stated, goal attainment models would be superior. If the evaluand's program theory is imperfect and objectives are not clear, or if only the objectives are unclear, the author suggests comparative evaluation models. Moreover, the selection of an evaluation model should be considered in context, that is, under consideration of the organizational milieu and model of state (i.e., sovereign, autonomous, negotiated, or responsive). As an alternative to this legitimacy and justice reasoning, Hansen offers "change reasoning," through which evaluation models can be altered for the different models of state. The problem the evaluand intends to address leads Hansen to promote the development of program theory to be assessed through theory-driven evaluation models, in the European context "realist" evaluation. The author stresses that theory-driven approaches easily become unrealistic given the problems of misspecification in addition to the numbers of variables that realistically can be dealt with. Hansen concludes with a

discussion of other factors that influence decisions about evaluation models used, namely what can be negotiated between evaluator and client and what are the evaluator's competencies (i.e., people do what they know).

The last two contributions to issue 11(4) focus on top-down versus bottom-up approaches to evaluation in very different contexts. Enrique Reboloso, Baltasar Fernández-Ramírez, and Pilar Cantón discuss both approaches in education contexts in Spain, a university administration and an integrated infant and primary school. The bottom-up self-evaluation approach implemented in the university setting yielded results, in that participants recognized shortcomings and need for change. Though not anticipated or expected, the top-down research-driven approach did not yield change. Colin Jacobs discusses the “success” of top-down and bottom-up approaches in the context of development evaluation, more specifically the evaluation of the Comprehensive Development Framework (CDF) in Kyrgyz Republic. Although, the World Bank and other donors promote knowledge exchange and learning across all 48 pilot countries in which the framework is implemented, the complexity and variation of programs and projects implemented raise doubt about their evaluability. Both top-down and bottom-up approaches to evaluating the CDF resulted in difficulties. The first was thwarted by problems in utilizing and measuring indicators suggested on the CDF's web site. The second was challenged by regional and cultural aspects. Jacobs concludes by calling for a system which addresses the needs of the people and builds capacity.

References

Coryn, C. L. S. (2005). Revisiting realistic evaluation [Review of the book *Realistic evaluation*]. *Journal of MultiDisciplinary Evaluation*, 3, 143-149.

Coryn, C. L. S. (2005). Practical program evaluation: Assessing and improving planning, implementation, and effectiveness [Review of the book *Practical program evaluation: Assessing and improving planning, implementation, and effectiveness*]. *American Journal of Evaluation*, 26(3), 405-407.

Davies, R. (2004). Scale, Complexity and the Representation of Theories of Change. *Evaluation: The International Journal of Theory, Research and Practice*, 10(1), 101-121.

Schröter, D. C. (2004). Evaluation: The International Journal of Theory, Research and Practice, 2003-2004 [review]. *Journal of MultiDisciplinary Evaluation*, 1, 114-120.