Program Evaluation in the Global South: Problems and Prospects

Satlaj Dighe, University of Minnesota
Trupti Sarode, University of Minnesota

Abstract:

In recent years, the number of program evaluations in the Global South has seen an uptick. However, the evaluation discourse in the South is largely dominated by discussions and actions around narrower dimensions such as monitoring and auditing, often driven by donor /funder requirements. Many countries are limited in their capacity to conduct evaluations on their own, instead serving as sites for large experimental and quasi-experimental studies that do not take into account the socio-cultural and political contexts in these settings. Additionally, the emphasis on assessing "impact" leaves program implementers with little information to improve program performance or understand the underlying mechanisms for why their programs work (or not). This perspective paper discusses the gaps and challenges facing evaluation in the Global South and presents recommendations for adopting evaluation approaches that value the complexity of context presented by the international development sector. In addition to utilizing innovative methods to measure "impact", valuing complexity requires intentionality on the part of researchers, policymakers, and practitioners to build, rather than circumvent, local capacity to design and conduct evaluations.

Contact: Satlaj Dighe, dighe004@umn.edu
Trupti Sarode, sarod004@umn.edu

Introduction

Program evaluation – the process of creating and interpreting information to understand the value, worth and merit of public programs – is a useful tool to explain the success and failure of public health, education, and many other social programs. In the international development sector of the Global South , evaluation is a critical resource to understand what works and why. Program and policy evaluation, when implemented properly, helps policymakers and program planners to identify gaps in development efforts, plan interventions, and assess the effectiveness of programs and policies. Evaluation could be a useful tool to understand the distributive impact of development by providing information on how the programs work, and for whom.

In recent years, the number of programs evaluation implemented in the international development sector of the Global South has increased manifold (UNDP, 2019); however, in most developing countries, the focus of evaluation activities has failed to transcend monitoring and auditing of short-term outputs and outcomes (Faulkner, Ayers, & Huq, 2015). Evaluation practice in the international development field has had limited success in serving as a helpful learning resource for implementing organizations (Bamberger, 1989; Kogen, 2018; Shiva Kumar, 2010). This paper intervenes by offering a discussion of the gaps and challenges facing evaluation in the Global South and proceeds to recommend evaluation approaches that advance organizational and sectoral knowledge building. The purpose of the paper is to understand what current practices and structural inequities restrict evaluation's ability to benefit organizations as an effective learning resource and what next steps could be followed to fully realize the potential of evaluation practice as an effective partner in the international development process.

This paper is organized into four sections. The first section reviews definitions of evaluation as relevant to the international development field. Building on this understanding, the second section provides the socio-political context of the Global South evaluation field by discussing the scope and uses of evaluation in the developing world. The third section reviews current dominant evaluation practices in the Global South and the problems involved. The fourth section discusses new prospects for development evaluation by providing concrete recommendations and next steps to address the problems in the Global South evaluations.

Defining Evaluation

Over time evaluation studies scholars, mostly based in the Global North, have sought to standardize definitions of evaluation as a systematic process which determines the merit of an intervention, more recently for its impact on societally important outcomes (Rossi, Lipsey, & Freeman, 2003). Scholarly debates notwithstanding, Patton (2008) offers a fairly comprehensive and accepted definition of evaluation as "the systematic collection of information about the activities, characteristics, and outcomes of a program to make judgments about the program, improve program effectiveness, and inform decisions about future programming" (p. 23). Yet, this standardization is implied, but notably absent, in evaluation as commonly articulated in the international development field.

While the United Nations Evaluation Group defines evaluation as "an assessment, as systematic and impartial as possible, of an activity, project, program, strategy, policy, topic, theme, sector, operational area or institutional performance" (UNDP, 2016, p.2), there is no standard definition or interpretation of evaluation in the international development sector. Indeed, the sector often conflates the terms "evaluation" and "monitoring," using them interchangeably. Michael Bamberger, an accomplished evaluator with the World Bank and other international development organizations (2000), notes:

The term evaluation is used differently by different agencies and authors. Some distinguish between monitoring activities, which are conducted during the project or program implementation to assess the efficiency or effectiveness with which inputs are used to achieve intended outputs, and evaluation activities, which assess the extent to which projects or programs achieved their intended objectives and have produced intended benefits in the target population. In other cases, the term evaluation is used more broadly to cover both of these functions (Bamberger, 2000, p. 96).

In the absence of a clear understanding of evaluation, the activity is interpreted differently by different actors. Thus when a program is "evaluated", it is often unclear to external audiences whether programs are simply audited for financial performance or are evaluated in the sense of the definitions furnished earlier (Hay, 2010). The confusing overlap between evaluation and monitoring has shifted focus from "learning from the program" to "tracking of the program". Thus, instead of using evaluation to understand the merit and value of the program for its stakeholders and broader society, the evaluation activities are narrowly employed to assess whether all the program activities are completed as planned. In addition to lack of clarity on definition and scope of evaluation, several other aspects of development evaluation in the Global South contribute to disproportionate attention to monitoring and tracking. These factors and conditions are discussed in the following section.

Evaluation in the Global South: An Overview

Evaluation practice in the Global South is situated in a markedly different context than that of the Global North. The main difference lies in the scope, ownership, and purpose for which evaluation is implemented. Program evaluation is a widely accepted organizational practice in the Global North (Bamberger, 2000). Most government programs in the Global North require mandatory evaluations to receive further funding. Similarly, nonprofits, foundations, and other entities receiving government and private funding conduct regular evaluation of their work. Evaluation in the Global South, on the other hand, has not gained a foothold except in the international development sector (Bamberger, 2000; Hay, 2010). Evaluation is not a routine and mandatory activity for most government programs in many developing countries (Bamberger, 2000; Mackay, 2002; Mehrotra, 2013). Many still lack institutional structures for developing and managing information systems (Mehrotra, 2013). However, the consequences of governments' absence in the field of evaluation go beyond the problems of information systems and infrastructures. The lack of government interest and involvement in evaluation results in absence of national policies guiding the quality of evaluation (Bamberger, 1991; Mackay, 2002), weak evaluation capacity at national and local level

(Hay, 2010; Shiva Kumar, 2010; Sridharan & De Silva, 2010) and lack of ethical standards governing evaluation practice (Bamberger, 2000).

In many developing countries, evaluation neither is a part of the university curriculum nor recognized as a distinct profession. Trained evaluators are lacking at the national, regional, and local level of development sectors (Shiva Kumar, 2010). Most evaluations in developing countries are carried out by social scientists, experts in their own field of research, though lacking the perspective and training of an evaluator (Shiva Kumar, 2010).

The lack of government's interest in evaluation is rooted in the absence of political will to develop a culture of social accountability (Mackay, 2002; Raimondo, 2018). Broader socio-political conditions such as dictatorships, military regimes, feudal systems, and absent or small middle class tend to compromise society's interest and ability to raise questions of accountability of the public sector (Raimondo, 2018). More often than not, evaluations conducted by international development organizations turn a blind eye towards these structural problems. Gran (1983) claimed that evaluations (of the World Bank) in the Global South emphasize macroeconomic policy issues and inappropriate government policies when the real issues are related to politics, the economy, corruption, and the world system issues he thinks the World Bank is not ready to face. Gran (1983) stated that "[T]he evaluations focus mainly on economic, technical, and financial issues and deliberately (or inadvertently) avoid broader political and economic considerations, which are more important in explaining project failure" (p. 305). This failure to address societal level structural issues related to equity, power asymmetries, and human rights have held the evaluation field back from making significant contribution to long-term social change in the developing world.

Evaluation in the Global South: Current Practices and Problems

 The evaluation field in most developing countries is extremely donor-driven as donors are the primary intended users and evaluation clients (Hay, 2010; Sridharan & De Silva, 2010). As a result, evaluation questions are designed by donors, and the evaluation data respond to donors' information needs (Faulkner, Ayers, & Huq, 2015). Donors' interests primarily result in two types of evaluations: a) accountability focused on listing activities and short-term outputs from the project and b) experimental evaluations seeking to explain the impact of donor contribution in a linear fashion. This paper argues that both types of evaluations have restricted use and thus can stifle evaluation as a learning opportunity, especially in the case of implementing organizations. The two consequences of donor-driven agenda, accountability focus, and bias for experimental (or quasi-experimental) design are discussed at length in the section below.

A Narrow Focus on Accountability

Ensuring accountability is an important mandate for evaluation (Rossi et al., 2003). However, in the international development field, accountability is often narrowly construed as accountability towards donors, thus ignoring wider social accountability towards stakeholders and partnering communities. The main evaluation question for such kind of evaluations is— "are we doing what we said we would do?" The evaluations supported by donor organizations spend most of their resources on monitoring project

implementation and to the production of immediate outputs (Bamberger, 1989). Very few resources are committed to understanding whether policies, programs, or projects achieve their intended impacts and benefits for their target populations (Faulkner et al., 2015).

When observing implementing organizations, it is clear they use evaluations differently than donor organizations. Bamberger (2000) noted that "many (implementing) organizations place emphasis on qualitative evaluations, seeking to understand the process of project implementation and impact of project implementation on intended beneficiaries" (p.335). Donors, on the other hand, are interested in the efficiency of outcome and performance measurements. Since most of the evaluations in the developing countries respond to donor needs, they hold little relevance to implementing organizations. Recent evaluation trends in the international development evaluation such as Results-Based management (RBM) and experimental impact evaluation reflect this donor-centric view of evaluation.

Accountability focused, donor-centric evaluation run a higher risk of producing inaccurate findings. Implementing organizations, by virtue of being closer to the field, are much more aware of field realities that influence the program context. In their absence, evaluations suffer from several inherent biases. In Rural Development: Putting the Last First, Chambers (1983) identified six biases that restrict the capacity of donors to design an evaluation effectively:

> 1) Spatial bias (otherwise known as urban bias): Project staff and researchers do not go far away from the head office.
> 2) Project bias: Agencies plan and evaluate in terms of projects. They show little interest in the majority of the rural poor, who are usually not affected (at least directly) by projects.
> 3) Person bias: Evaluators and development practitioners tend to get their information from elites, males, users, and adopters of new technologies, and people who are active, healthy, and present.
> 4) Dry-season bias: During the rainy season, the development experts make few site visits.
> 5) Diplomatic bias: A combination of politeness, fear, anxiety, and language barriers deter evaluators from speaking to the poor and underprivileged, especially those who are not directly involved with the projects.
> 6) Professional bias: The professionals often talk to the wealthier, more informed, and better-educated farmers because they are able to host the visitors and hold conversations.

Dr. Chambers identified these biases mainly in the context of agriculture and rural development sector; however, these are relevant for other development projects as well.

Methodological Bias

Currently, impact evaluations using experimental design methods are considered the gold standard in the international development sector (Bollen, Paxton, & Morishima, 2005; Picciotto, 2012). However, there is growing attention among evaluation scholars and practitioners to the limitations of "impact measurement" itself

and why "impact measurement," while valuable, might not be the best option to measure a program's success in all situations. For instance, Gugerty and Karlan (2018) argue that a program may not be ready for a randomized-control trial (RCT) and might instead benefit from program improvement techniques such as formative evaluation. Many other scholars support this view and argue for the re-examination of "impact measurement" as the true north when it comes to evaluating a program's success (Scriven, 2008; Cook, 2007). Another conversation that has taken shape recently is around "causality"--there is growing acknowledgment among scholars and practitioners that a range of methods/approaches can test a causal claim. These scholars suggest evaluators need to be perceptive of the context of the evaluation, literate in multiple ways of thinking about causality, and familiar with a range of design/methods to assess causal claims (Gates & Dyson, 2016; Khagram & Thomas, 2010; Patton, 2018).

The skepticism around "successionist" ways to assess causality has been accompanied by growing demand for and emphasis on using complexity approaches to design and conduct evaluations (Gates & Dyson, 2016, p. 35; Walton, 2016). Yet, the dominance of methods like RCTs often precludes the use of complexity approaches even in contexts where these might be relevant. These approaches acknowledging the complexity of causes, context, and systems are also limited by funders' perceived need for simple boilerplate evaluations that provide yes/no answers(Hall, 2004). While they might adequately serve the purposes of accountability, the interest in these less traditional forms of evaluation requires more investigation into their application. Complexity approaches (e.g., developmental evaluation, human-centered design) provide an alternative way of looking at causality/program impact, bolstering organizational learning and innovation, and operationalizing complexity and systems thinking in program design, implementation, and evaluation. Piccotio (2012) points out that methods such as RCTs are best suited for well-established and stable programs. However, programs in the international development field are often too dynamic to meet the criteria of an RCT.

Lack of Evaluation Capacity

In recent years, issues of evaluation capacities have gained traction in the Global North. In comparison, discussions on evaluation capacities in the Global South have lagged behind (Hay, 2010). Sri Lanka, Philippines are among the fewer countries that have taken an interest in the development of country-specific evaluation standards and guidelines (National Evaluation Policy of Sri Lanka, 2018; National Evaluation Policy Framework of The Philippines, 2015). ECB — an intentional approach to building individual and organizational capacity for evaluation — has received attention in the evaluation sector of the North. However, discussions around and efforts to build evaluation capacities in the South have been rather nascent. Recently, the World Bank and the UN have recognized a need to develop evaluation capacity in the Global South. For example, The World Bank has regarded evaluation capacity development as a core principle of its evaluation. The World Bank's evaluation principles highlight its commitment to developing evaluation capacity and leadership within the World Bank and among its implementing partners and clients (The World Bank, 2019) . Similarly, the United Nation's Evaluation Group included the development of national evaluation capacities as one of the core norms of its evaluation practice(UNEG, 2016). However,

while ECB is recognized as an important theme of the development evaluation, the lack of ECB models germane to the countries of the Global South makes ECB initiatives in these countries dependent on the North for the development of ECB strategies, content, and processes(Bamberger, 1991; Hay, 2010).

The Global South needs evaluation capacities that can address specificities of its context —languages, cultures, and socio-economic conditions. Ofir (2013) contended that given the severe power asymmetries and vulnerabilities in developing countries, it is imperative for development evaluators to have an advanced understanding of systemic inequalities and poverty. Evaluators working in the Global South also need to be aware of the development – evaluation interface (Ofir & Shiva Kumar, 2013) — implicit assumptions, values, and frameworks of the field of international development that influence design and results of evaluation (Ofir, 2013; Ofir & Shiva Kumar, 2013; Sridharan & De Silva, 2010).

Cultural competencies are rarely discussed in the context of the Global South despite the fact that many countries in the Global South comprise multilingual, multicultural, and multiethnic societies. Cultural competency requires a thorough knowledge of the people and places for which the evaluation is being conducted, as well as of local history and the culturally derived values of the communities participating in the program. However, more often than not, cultural competency in the international development evaluation is rather simplistically equated with colloquial knowledge of the local language. Many times a national evaluation team member who has basic colloquial knowledge of the language is deemed sufficient to bring in the cultural understanding of the region and communities. Practices such as these perpetuate a strong urban bias in the evaluations (Chambers, 1983; Shiva Kumar, 2010).

Evaluators from the Global South need to build capacity to develop and utilize indigenous evaluation models and frameworks that reflect values and perspectives of their cultures and societies. In the absence of indigenous research, Global South evaluation depends on the North not only for resources but also for ideas, values, methods, and approaches (Hay, 2010; Ofir, 2013). Ofir (2013) noted that capacity building in the South mirrors the training given in the North. She observed that capacity-strengthening efforts in developing countries tend to focus on technical aspects of evaluation within established approaches and frameworks. Although training in these skills is arguably necessary, these skills may not encourage or stimulate deeper questioning of the established or alternative approaches and frameworks. Talking about importance developing of local and indigenous evaluation methods, Ofir (2013) commented that, "this is not about cultural sensitivity, but rather about the fundamental questioning of worldviews, frameworks, and definitions on which evaluation theory and practice—and resultant development—have been built" (p. 585).

The current dominant framework of RBM and experimental impact evaluation side-step the questions of South-specific competencies and indigenous world views of valuing. Today's evaluation field in the South lacks ability of creating knowledge and information that can be useful for program design, planning, and improvement. Along with the methodological rigor and increased capacity to better implement evaluation, the Global South needs new methods and approaches to successfully capture complexities and locally rooted knowledge of the development. The next section discusses approaches that can be relevant in achieving these goals.

Evaluation in the Global South: The Way Forward

Developmental Evaluation and Human-Centered Design as Promising Tools

Developmental Evaluation (DE) is a form of program evaluation that informs and refines innovation, including program development (Patton, 2011). Formative and summative evaluations tend to assume a linear trajectory for programs or changes in knowledge, behavior, and outcomes. In contrast, developmental evaluation responds to the nature of change that is often seen in complex social systems. DE is currently in use in a number of fields where nonprofits play important roles, from agriculture to human services, international development to arts, and education to health. Another technique that has gained salience around addressing complexity and innovation is human-centered design (HCD) – it shares many parallels with developmental evaluation and attends specifically to the user-experiences throughout the program design process. More generally, it involves a cyclical process of observation, prototyping, and testing (Bason, 2017).

Although human-centered design is seemingly focused upon initiation (or program design) and evaluation on assessment after the fact, human-centered design and developmental evaluation share a number of commonalities. Both support rapid-cycle learning among program staff and leadership to bolster learning and innovative program development (Patton, 2010; Patton, McKegg & Wehipeihana, 2015). Program staff and evaluators/designers work as equal partners in the process; while evaluators benefit from a holistic understanding of the program and context for their evaluation efforts, this continuous engagement brings program staff closer to the evaluation thereby building evaluative thinking (or critical thinking) abilities and enhancing their capacity to collect and use data. Additionally, the two-pronged use of developmental evaluation and human-centered design presents a reframing of how evaluators and implementing organizations conceive and conceptualize program impact. Unlike the emphasis of experimental/quasi-experimental techniques around discerning a singular cause-effect relationship for programs and interventions, DE and HCD align with different ways of thinking about causality and program impact—they conceive the world as comprising and operating as a complex system where "the focus is on examining the multiple, interdependent causal variables and nonlinear cyclical feedback processes that affect the structure and dynamical behavior of a system over time"(Gates & Dyson, 2017, p.25).

Theory-driven Evaluation

In recent years, theory-driven evaluations have gained support from evaluators who think that the purpose of the evaluation is not only to assess whether an intervention works or does not work but also to understand how and why it does so. The theory-driven evaluation approach assumes that the design and application of evaluation must be guided by a conceptual framework called program theory (Chen,

2015). Program theory is defined as a set of explicit or implicit assumptions by stakeholders about requisite action to solve a social, educational or health problem and why the problem will respond to this action (Chen, 2015). The development of program theory that builds on experiences and insights of multiple stakeholders is a core component of the theory-driven evaluations. Integrating experiences and assumptions of diverse stakeholders about how the program works is a complex and time-consuming process. However, it is this very process that enhances the evaluation's ability to explain the change caused by the program.

The benefits of using a theory-driven approach are numerous. The program theory helps planners to distinguish between implementation failure (gaps in the execution of the program) and theory failure (gaps in the knowledge about how it is supposed to work). Without a well-designed program theory, it is impossible to understand what aspects of the programs did not work and why. By contrast, impact evaluations conducted using experimental methods are limited to informing policymakers if the program was successful or unsuccessful in leading the desired change. RCTs' strength lies in their ability to explain whether the difference between the treatment and control group is significant. However, by design, they usually fall short of explaining the underlying mechanisms responsible for the success of the program. Additionally, they also cannot provide information on the generalizability of the findings or replicability of the program -- information deemed most useful for policy makers and program planners. The theory-driven evaluation with their careful explanation of context, mechanisms, and the processes of the program is much more successful in providing insights on scaling up or replication of the program in different settings.

Theory-driven evaluation is an umbrella term that includes a plethora of models and frameworks that use program theory as an integral tool of evaluation design. Models and frameworks in the theory-driven evaluation approach range from logic models or log frame analysis that employ a linear approach of explaining program theory to outcome chain modeling, Theory of Change (TOC) approach, or Realist Evaluation (RE) that integrate complexity and uncertainty involved in the social processes(Funnell & Rogers, 2011; Kazi, 2013; Pawson & Tilley, 1997). In recent years, several international organizations such as the World Bank and the United Nations have adopted a theory-driven evaluation approach for their evaluations. However, very few organizations are able to execute theory-driven evaluations in a meaningful manner (Funnell & Rogers, 2011). Thus, awareness, training, and knowledge building on theory-driven evaluation are still lacking in the international development field.

Evaluation Capacity Building

The gaps in evaluation practice cannot be addressed by importing resources, ideas, and evaluators from the Global North. South-based evaluations that blindly follow methods and designs developed in the North and for the North fail to capture complexities and lived realities of the communities in the Global South. The evaluations designed by donors, favor donors' agenda ignoring information and learning needs of local partners (Bamberger, 2000).

The countries of the Global South need to invest in the process of evaluation capacity building (ECB) —the intentional work to continuously create and sustain overall organizational processes that make a quality evaluation and its uses routine (Stockdill,

Baizerman, and Compton 2002, p. 14) — at the individual, organizational, and sectoral level. At the individual level, there is a need for training programs, both short-term and degree level courses in the area of program evaluation. Currently, program evaluation is rarely offered at national universities and colleges (Shiva Kumar, 2010) . Short-term evaluation courses are offered to government employees through select government training courses or to the employees of international organizations through their collaboration with evaluation groups or universities based in the Global North. These arrangements restrict access to evaluation knowledge in implementing organizations or staff of regional and local government offices. In order to tap the real evaluation potentials in the countries from the Global South, evaluation knowledge needs to be built at the regional and local level and in regional languages(Ofir & Shiva Kumar, 2013).

      At an organizational level, ECB can be built with the help of direct and indirect ECB strategies (Labin, Duffy, Meyers, Wandersman, & Lesesne, 2012). Direct strategies include evaluation training, workshops, seminars, evaluation internships, and opportunities to attend external evaluation training. However, ECB experts note that these direct strategies just form the tip of the iceberg of organizational evaluation capacity (Bourgeois & Cousins, 2013; Cousins & Whitmore, 1998; King & Volkov, 2005; Patton, 2008). To promote sustainable evaluation practices, organizations need to initiate a series of changes altering organizational systems, structures, and culture. Examples of ECB structures include a) an organizational-level ECB plan; b) infrastructure to support ECB – a basic monitoring and performance tracking system; capacity to collect, analyze, and report data; organizational databases; c) purposeful socialization of new employees into the organizational evaluation process; and d) peer learning structures for evaluation (King & Volkov, 2005). As in any other capacity building process, ECB also requires commitment in terms of resources. These resources include funding, technology support - investment in database development software and other required technology, hiring technologically competent staff, and knowledge resources - membership in data repositories, access to the research journals, and hiring external resource persons to help staff build their knowledge (King & Volkov, 2005; Preskill & Boyle, 2008).

      In addition to ECB structures and processes at the organization, evaluation capacity is influenced by organizational culture (Bourgeois & Cousins, 2013; King & Volkov, 2005; Preskill & Boyle, 2008). Organizations that have an environment of trust and a culture of learning are more open to viewing evaluation as a performance improvement tool rather than a judgment on their work and are more willing to develop evaluation capacity (Preskill & Boyle, 2008; Taut, 2007). In the context of organizations implementing international development interventions, donors' interests, vision, and support for evaluation determine the level of evaluation capacity and their interest in the evaluation (Bamberger, 2000; Sen, 1987). Empirical research shows that organizational evaluation capacity and organizational learning share a close link with each other (Bourgeois & Cousins, 2013). Organizations with a high learning capacity tend to internalize evaluation in the culture and decision making of the organization. In other words, organizations with a high learning capacity also have a good organizational evaluation capacity. Evaluation, in these organizations, is a part of the learning process.

Conclusion

Countries from the Global South with their histories, cultures, economic conditions, and institutional structures are markedly different from the countries in the Global North. The South, therefore, requires evaluation frameworks that reflect values and its unique context. The evaluation field in the development sector of the Global South needs to address information and feedback needs of different stakeholders - donors, government, implementing organizations, and communities. This paper argues that current dominant frameworks that focus on accountability and narrowly defined impact measurements have been detrimental to the development sector in the Global South. The evaluation approaches discussed in this paper are just a few examples of recent thinking in evaluation that acknowledge the complex nature of development work while providing relevant feedback to practitioners and stakeholders. Prospects of alternatives, such as Developmental Evaluation, Human-Centered Designs, and Realist Evaluation grounded in different worldviews, is one step towards more effective, inclusive, and context-sensitive evaluation in international development.

References

Bamberger, M. (1989). The monitoring and evaluation of public sector programs in Asia: Why are development programs monitored but not evaluated? *Evaluation Review*, *13*(3), 223–242. https://doi.org/10.1177/0193841X8901300303

Bamberger, M. (1991). The politics of evaluation in developing countries. *Evaluation and Program Planning*, *14*(4), 325–339. https://doi.org/10.1016/0149-7189(91)90015-9

Bamberger, M. (2000). The evaluation of international development programs: A view from the front. *American Journal of Evaluation*, *21*(1), 95–102. https://doi.org/10.1177/109821400002100108

Bason, C. (2017). *Leading public design: Discovering human-centred governance.* Policy Press.

Bollen, K., Paxton, P., & Morishima, R. (2005). Assessing international evaluations: An example from USAID's democracy and governance program. *American Journal of Evaluation*, *26*(2), 189–203. https://doi.org/10.1177/1098214005275640

Bourgeois, I., & Cousins, J. B. (2013). Understanding Dimensions of Organizational Evaluation Capacity. *American Journal of Evaluation*, *34*(3), 299–319. https://doi.org/10.1177/1098214013477235

Chen, H. T. (2015). *Amazon.com: Practical program evaluation: Theory-driven evaluation and the integrated evaluation perspective eBook: Huey T. (Tsyh) Chen: Kindle Store* (Second). Retrieved from https://www.amazon.com/Practical-Program-Evaluation-Theory-Driven-Perspective-ebook/dp/B00O81XTSM

Cousins, J. B., & Whitmore, E. (1998). Framing participatory evaluation. *New Directions for Evaluation*, *1998*(80), 5–23. https://doi.org/10.1002/ev.1114

Faulkner, L., Ayers, J., & Huq, S. (2015). Meaningful measurement for community-based adaptation. *New Directions for Evaluation*, *2015*(147), 89–104. https://doi.org/10.1002/ev.20133

Funnell, S. C., & Rogers, P. J. (2011). *Purposeful program theory: Effective Use of theories of change and logic models* (1 edition). San Francisco, CA: Jossey-Bass.

Gates, E., & Dyson, L. (2017). Implications of the changing conversation about causality for evaluators. *American Journal of Evaluation*, *38*(1), 29–46. https://doi.org/10.1177/1098214016644068

Gran, G. (1983). *Development by people; citizen construction of a just world*. Retrieved from http://agris.fao.org/agris-search/search.do?recordID=XF2015027383

Gray, K., & Gills, B. K. (2016). South–south cooperation and the rise of the Global South. *Third World Quarterly*, *37*(4), 557–574. https://doi.org/10.1080/01436597.2015.1128817

Gugerty, M. K., & Karlan, D. (2018). *Goldilocks challenge: Right-fit evidence for the social sector - Oxford Scholarship*. Retrieved from https://www.oxfordscholarship.com/view/10.1093/oso/9780199366088.001.0001/oso-9780199366088

Hall, P. D. (2004). A historical perspective on evaluation in foundations. In M. T. Braverman, N. A. Constantine, & J. K. Slater (Eds.), *Foundations and Evaluation: Contexts and Practices for Effective Philanthropy* (1st ed.). San Francisco, CA: John Wiley & Sons.

Hay, K. (2010). Evaluation field building in South Asia: Reflections, anecdotes, and questions. *American Journal of Evaluation*, *31*(2), 222–231. https://doi.org/10.1177/1098214010366175

Kazi, M. A. F. (2013). *Realist Evaluation in Practice*. Retrieved from https://uk.sagepub.com/en-gb/eur/realist-evaluation-in-practice/book210924

Khagram, S., & Thomas, C. W. (2010). Toward a platinum standard for evidence-based assessment by 2020. In *Source: Public Administration Review* (Vol. 70).

King, J. A., & Volkov, B. (2005). A framework for building evaluation capacity based on the experiences of three organizations. *CURA Reporter*, 10–16.

Kogen, L. (2018). What have we learned here? Questioning accountability in aid policy and practice. *Evaluation*, *24*(1), 98–112. https://doi.org/10.1177/1356389017750195

Labin, S. N., Duffy, J. L., Meyers, D. C., Wandersman, A., & Lesesne, C. A. (2012). A research synthesis of the evaluation capacity building literature. *American Journal of Evaluation*, *33*(3), 307–338. https://doi.org/10.1177/1098214011434608

Mackay, K. (2002). The World Bank's ECB experience. *New Directions for Evaluation*, *2002*(93), 81–100. https://doi.org/10.1002/ev.43

Mehrotra, S. (2013). Monitoring, evaluation and performance management in South Asia: The challenge of building capacity. *Evaluation*, *19*(1), 74–84. https://doi.org/10.1177/1356389012471257

National Evaluation Policy of Sri Lanka. (2018). Retrieved from: http://www.pmm.gov.lk/resources/Draft%20National%20Evaluation%20Policy_English.pdf

National Evaluation Framework of The Philippines. (2015). Retrieved from: http://www.neda.gov.ph/wp-content/uploads/2015/07/NEDA-DBM%20Joint%20Memorandum%20Circular%20No.%202015-01%20-%20National%20Evaluation%20Policy%20Framework%20of%20the%20Philippines.pdf

Ofir, Z. (2013). Strengthening evaluation for development. *American Journal of Evaluation*, *34*(4), 582–586. https://doi.org/10.1177/1098214013497531

Ofir, Z., & Shiva Kumar, A. K. (2013). Evaluation in developing countries: What makes it different? In *Emerging Practices in International Development Evaluation* (p. 250). Charlotte, NC: Information Age Publishing.

Patton, M. Q. (2008). *Utilization Focused Evaluation* (4th ed.). Sage Publications.

Patton, M. Q. (2010). *Developmental evaluation: Applying complexity concepts to enhance innovation and use*. New York: The Guilford Press.

Patton, M. Q., McKegg, K., & Wehipeihana, N. (Eds.). (2016). *Developmental evaluation exemplars: Principles in practice*. New York: Guilford Press.

Patton, M. Q. (2018). Building a community of practice through adaptation and integration of multiple methods and diverse approaches to evaluative inquiry. *INTERNATIONAL JOURNAL OF MULTIPLE RESEARCH APPROACHES*, *10*(1), 183–198. https://doi.org/10.29034/ijmra.v10n1a12

Pawson, R., & Tilley, N. (1997). *Realistic Evaluation | SAGE Publications Inc* (1st ed.). Retrieved from https://us.sagepub.com/en-us/nam/realistic-evaluation/book205276

Picciotto, R. (2012). Experimentalism and development evaluation: Will the bubble burst? *Evaluation*, *18*(2), 213–229. https://doi.org/10.1177/1356389012440915

Preskill, H., & Boyle, S. (2008). A multidisciplinary model of evaluation capacity building. *American Journal of Evaluation*, *29*(4), 443–459. https://doi.org/10.1177/1098214008324182

Raimondo, E. (2018). The power and dysfunctions of evaluation systems in international organizations. *Evaluation*, *24*(1), 26–41. https://doi.org/10.1177/1356389017749068

Rossi, P. H., Lipsey, M. W., & Freeman, H. E. (2003). *Evaluation: A systematic approach, 7th Edition* (7th edition). Thousand Oaks, Calif.: SAGE Publications, Inc.

Sen, B. (1987). NGO self-evaluation: Issues of concern. *World Development*, *15*(Supplement 1), 161–167. https://doi.org/10.1016/0305-750X(87)90154-9

Shiva Kumar, A. K. (2010). A comment on "evaluation field building in South Asia: Reflections, anecdotes, and questions." *American Journal of Evaluation*, *31*(2), 238–240. https://doi.org/10.1177/1098214010366176

Sridharan, S., & De Silva, S. (2010). On ambition, learning, and co-Evolution: Building evaluation as a field. *American Journal of Evaluation*, *31*(2), 246–251. https://doi.org/10.1177/1098214010366297

Taut, S. (2007). Studying Self-Evaluation Capacity Building in a Large International Development Organization. *American Journal of Evaluation*, *28*(1), 45–59. https://doi.org/10.1177/1098214006296430

The World Bank. (2019). *World Bank Group evaluation principles* (p. 18). Retrieved from https://ieg.worldbankgroup.org/sites/default/files/Data/reports/WorldBankEvaluationPrinciples.pdf

UNDP. (2019). *Annual Report on Evaluation* (p. 72) [Annual report]. Retrieved from United Nations Development Program website: http://web.undp.org/evaluation/documents/annual-report/2019/ARE2018.pdf

UNEG. (2016). *Norms and standards for evaluation* (p. 32). Retrieved from United Nations Evaluation Group website: http://www.uneval.org/document/detail/1914

Walton, M. (2016). Expert views on applying complexity theory in evaluation: Opportunities and barriers. *Evaluation*, 22(4), 410–423.