Identifying the Need for Evaluation Capacity Assessment Tools and Guidance

Ana Claudia Franca-Koh, MEASURE Evaluation
Lwendo Moonzwe, MEASURE Evaluation

January 2017

Identifying the Need for Evaluation Capacity Assessment Tools and Guidance

Ana Claudia Franca-Koh, MEASURE Evaluation
Lwendo Moonzwe, MEASURE Evaluation

January 2017

ACKNOWLEDGMENTS

ABBREVIATIONS

CDC   U.S. Centers for Disease Control and Prevention

ECB   evaluation capacity building

ECD   evaluation capacity development

M&E   monitoring and evaluation

MECAT   Monitoring and Evaluation Capacity Assessment Toolkit

PEPFAR   President's Emergency Plan for AIDS Relief

UNAIDS   Joint United Nations Programme on HIV/AIDS

UNEG   United Nations Evaluation Group

USAID   U.S. Agency for International Development

WHO   World Health Organization

INTRODUCTION

In 2011, the U.S. Agency for International Development (USAID) published its *Evaluation Policy* (USAID, 2011). The policy emphasizes the need to conduct more evaluations of its programs to ensure greater accountability and learning, and it outlines best practices and requirements for conducting evaluations. Since releasing the policy, USAID has commissioned an increasing number of evaluations of its programs (USAID, 2016).

The importance of evaluations for international public health programs has been long recognized (CDC, 1999, 2011; RAND, 2005; WHO, 2001), with demand for such evaluations coming from both internal and external sources. Donors or those external to program implementation seek evidence of accomplishments and accountability for resources spent, whereas those involved in program implementation seek evidence to inform and improve program design (WHO, 2001). Within USAID, the need for more evaluations was driven by the understanding that evaluations provide information and analysis that prevent mistakes from being repeated and increase the likelihood of greater yield from future investments (USAID, 2011). Finally, there is overall recognition that evaluations should be of high quality and driven by demand, and that results should be communicated to relevant stakeholders (PEPFAR, 2014).

Despite the increased demand for evaluations, there is limited evaluation capacity in many countries where international development programs are implemented (RAND, 2005). Before strategies to strengthen evaluation capacity can be implemented, it is important first to assess existing evaluation capacity and develop action plans accordingly.

We conducted a review of existing assessment tools and guidance documents related to assessing organizations' capacity to carry out evaluations of international public health programs in order to determine the adequacy of those materials. Here, we summarize the key findings of our review of the literature and provide recommendations for the development of future tools and guidance documents.

METHODS

The purpose of our literature search was twofold. First, we wanted to broaden our understanding of evaluation capacity and how it was defined, and second, we wanted to identify existing tools for measuring evaluation capacity. We were specifically interested in identifying tools that measured capacity to carry out evaluations of international public health programs.

We began our review of the literature by conducting a general Internet search using the Google search engine and the key words "evaluation capacity assessment," "evaluation capacity assessment tools," "evaluation capacity assessment guidance," "evaluation standards," "evaluation dimensions," and "evaluation competencies."

This was not a systematic review of every paper or report related to the above search criteria but rather a purposeful selection process that was iterative in the sense that decision making was based on what we were learning during review of the literature. As we found documents that were relevant, we discussed them as a group and established whether they should be included in our database. Oftentimes, review and discussion of the documents promoted further Internet searches based on new, emergent themes or topics.

In addition, we reviewed two capacity assessment tools used by the USAID MEASURE Evaluation Project: (1) the Monitoring and Evaluation Capacity Assessment Toolkit (MECAT), and (2) the Monitoring and Evaluation Reference Group 12 Components Monitoring and Evaluation System Strengthening tool. In reviewing these two tools and associated guidance, we identified additional relevant references to review. After this initial process, we found our results to still be limited and thus broadened our search to include any tools or publications on evaluation capacity on organizational websites, including USAID, CDC, PEPFAR, UNAIDS, and World Bank. During this broader search, we also reviewed various evaluation policy documents, including USAID's Evaluation Policy (USAID, 2011), for any content on evaluation capacity specifically.

RESULTS

As we searched for evaluation capacity assessment tools and guidance documents, it became apparent that the definition and use of "evaluation capacity" varied greatly in the literature. We therefore begin by describing these definitions as context for our findings related to the assessment tools and guidance documents.

Definitions of Evaluation Capacity

The term "evaluation capacity" was defined and used in several different ways in the literature, including the eight definitions we found during our review (Table 1). One definition comes from extensive literature on the evaluation capacity of organizations, which refers to the capacity to conduct evaluations and use their results. However, this type of evaluation capacity has nothing to do with program evaluation; instead, it refers to internal evaluations intended to assess the performance of organizations for their own learning and improvement (Preskill & Torres, 2000; Volkov & King, 2007). Some other, broader definitions are unclear as to whether they refer to evaluations of programs or evaluations geared toward organizational learning (for example, Preskill & Boyle, 2008). We also found that evaluation capacity is often used as part of the term "evaluation capacity building" (ECB), or with "development," as in "evaluation capacity development" (ECD). The ECB literature typically refers to organizations' internal capacity to evaluate themselves (Cousins, et al., 2011; Labin, et al., 2012), whereas ECD refers to the capacity to carry out evaluations of programs (OECD, 2006). Other definitions are more organization-specific and include details about the type of organization—for example, HIV prevention organizations (Gilliam, et al., 2003), or health departments and community-based organizations (CBOs) (Gibbs, et al., 2002)—or the type of program that is the focus of the evaluation (for example, HIV prevention programs [Gilliam, et al., 2003]). Finally, some definitions focus on the capacity to conduct evaluations at just one level (for example, organizational level only, as in Gilliam, et al., 2003) while others apply to multiple levels (for example, individual, organizational, and system levels [OECD, 2006]).

Table 1. Definitions of evaluation capacity

Source	Definition of Evaluation Capacity
1. Cousins, et al., 2011	**Evaluation capacity building**: the capacity both to *conduct* evaluation studies, as well as the capacity within organizations to *use* evaluation results and processes.
2. Gilliam, et al., 2003	**Evaluation capacity**: the ability of HIV prevention organizations (for example, health departments and CBOs) to conduct evaluations of their own HIV prevention programs.
3. Gibbs, et al., 2002	**Evaluation capacity**: the extent to which a CBO has the necessary resources and motivation to conduct, analyze, and use evaluations.
4. James Bell Associates, 2013	**Evaluation capacity**: an organization's ability to conduct, use, and continuously learn from evaluation processes.
5. Labin, et al., 2012	**Evaluation capacity building**: an intentional process to increase individual motivation, knowledge, and skills, and to enhance the ability of a group or organization to conduct or use evaluations.
6. OECD, 2006	**Evaluation capacity development**: the process whereby people, organizations, and society as a whole strengthen, create, adapt, and maintain evaluation capacities over time.
7. Preskill & Boyle, 2008	**Evaluation capacity building**: a process by which strategies are designed and implemented to assist individuals, groups, and organizations with the process of conducting effective, useful, and professional evaluation practice.
8. Stockdill, et al., 2002	**Evaluation capacity building**: intentional work to continuously create and sustain overall organizational processes that routinize quality evaluations and their uses.

Evaluation Capacity Assessment Tools

Table 2 summarizes the three assessment tools identified in the literature review that included components related to the capacity to evaluate public health programs. All of the tools focused on monitoring and evaluation (M&E) more broadly and so inherently had limited content specific to evaluation capacity. We also found some tools that were specific to assessing organizational evaluation capacity (the capacity of organizations to evaluate themselves for organizational learning; for example, Informing Change, n.d. and Capacity for Health, n.d.); these tools were not included in Table 2 because they were not designed to assess capacity to evaluate programs.

Table 2. Summary of tools identified with components related to evaluation capacity

	Source	Description	Components Related to Evaluation Capacity	Gaps and Limitations
1	**MEASURE Evaluation, 2014** *Monitoring and Evaluation Capacity Assessment Toolkit* (MECAT)	The MECAT is a group of tools for assessing an organization's capacity to carry out M&E functions. It allows organizations to identify key M&E stakeholders, understand current capacity in M&E, and determine the most appropriate interventions for strengthening the M&E system. The MECAT examines M&E capacity across four capacity dimensions (status, quality, technical autonomy, and financial autonomy) and is composed of four tools (participatory group assessment, individual self-assessment, key informant interviews, and desk review).	The self-assessment tool has seven general questions about evaluation competencies, ranging from design and methods to evaluation management. It also includes questions on data analysis, dissemination, and use competencies relevant to evaluation. The participatory group assessment tool is based on the 12 Components Tool (UNAIDS, 2010, see below) and includes several questions relevant to evaluation capacity, including human capacity, availability of training, and dissemination of findings.	Because the overall focus is on M&E, it does not address evaluation capacity in any depth.

		Source	Description	Components Related to Evaluation Capacity	Gaps and Limitations
2		**UNAIDS, 2007** Checklist for Assessing and Monitoring Capacity of the HIV/AIDS M&E System. Appendix F in *Building National HIV/AIDS Monitoring and Evaluation Capacity: A Practical Guide for Planning, Implementing, and Assessing Capacity Building of HIV/AIDS Monitoring and Evaluation Systems*.	The checklist provides a framework for assessing the capacity of the national HIV/AIDS M&E system in terms of five performance objectives and three capacity domains. The five performance objectives are: 1. Develop a national HIV/AIDS M&E plan 2. Implement the national HIV/AIDS M&E pan 3. Assess, upgrade, and standardize M&E capacity of stakeholders and national and subnational HIV/AIDS M&E organizations and personnel 4. Produce high quality data on a regular and timely basis 5. Communicate information on HIV and AIDS to relevant stakeholders and facilitate the use of information among stakeholders The three capacity performance objectives for a national HIV/AIDS M&E system are: -Resources -Organizational structure and operations -Organizational culture	Under the "Resources" domain, the tool asks whether there is a sufficient number of staff devoted to M&E related to HIV and AIDS, and then whether general M&E skills are available among the human resources.	The tool focuses on assessing M&E capacity more broadly and not evaluation capacity specifically.

	Source	Description	Components Related to Evaluation Capacity	Gaps and Limitations
3	**UNAIDS, 2010** *12 Components Monitoring and Evaluation System Strengthening Tool*	The *Organizing Framework for a Functional National HIV Monitoring and Evaluation System* (UNAIDS, 2008) describes 12 components of a functional HIV M&E system. This tool includes questions to assess the status of each of the following 12 components: 1. Organizational structures with M&E functions 2. Human capacity for M&E 3. M&E partnerships 4. M&E plan 5. Costed M&E workplan 6. M&E advocacy, communications, and culture 7. Routine program monitoring 8. Survey and surveillance 9. M&E databases 10. Supervision and data auditing 11. Evaluation and research 12. Data dissemination and use	Two of the twelve components include elements related to evaluation capacity: Component 2 (human capacity) and Component 11 (evaluation and research). The human capacity performance goal is intended to ensure that an adequate number of skilled individuals are available at all levels of the M&E system in order for all tasks defined in the annual costed national HIV M&E workplan to be completed. The evaluation and research performance goal is to identify key evaluation and research questions, coordinate studies to meet the identified needs, and enhance the use of evaluation and research findings.	The tool focuses on M&E broadly, with limited elements of evaluation capacity included in only two of the 12 components.

Evaluation Capacity Guidance Documents

We also identified four relevant guidance documents that included information about evaluation capacity (Table 3). Of the four documents, the *PEPFAR Evaluation Standards of Practice* (PEPFAR, 2014) focused the most on evaluation and emphasized the importance of strengthening evaluation capacity; however, it did not include a tool for assessing evaluation capacity as defined by the evaluation standards proposed. The other three documents focused more on capacity building within the M&E system in general and included limited information on evaluation capacity, specifically. We also identified a guidance document related to capacity building more generally (PEPFAR, 2012), but it did not refer to evaluation capacity in particular and so is not included in Table 3.

Table 3. Summary of guidance documents and frameworks that include evaluation capacity-related components

	Source	Description	Components Related to Evaluation Capacity	Gaps or Limitations
1	**PEPFAR, 2014** *PEPFAR Evaluation Standards of Practice*	Purpose of the document is to articulate evaluation standards that should improve the quality of evaluations and their contribution to decision making.	The document acknowledges the importance of strengthening evaluation capacity and describes 11 evaluation standards of practice: 1. Engage stakeholders 2. Clearly state evaluation questions, purpose, and objectives 3. Use appropriate evaluation design, methods, and analytical techniques 4. Address ethical considerations and assurances 5. Identify resources and articulate budget 6. Construct data collection and management plans 7. Ensure appropriate evaluator qualifications and evaluator independence 8. Monitor the planning and implementation of an evaluation 9. Produce quality evaluation reports 10. Disseminate results 11. Use findings for program improvement	Though the document provides evaluation standards, it does not include a tool or criteria to measure or assess the standards.
2	**UNAIDS, 2007** *Building National HIV/AIDS Monitoring and Evaluation Capacity:* *A Practical Guide for Planning, Implementing, and Assessing Capacity Building of HIV/AIDS Monitoring and Evaluation Systems*	The purpose of the guide is to assist national HIV/AIDS programs to plan and implement HIV/AIDS M&E system capacity building. The guide provides a step-by-step approach to assessing M&E system capacity, developing a plan for addressing capacity gaps, and monitoring and evaluating capacity-building efforts.	The guide lists examples of M&E capacities to assess, including: • Motivation • Decision-making approaches • Workforce and professional development • Resources and supports • Technical knowledge and tools	Examples of M&E capacities to assess are provided, but they are broad and not specific to evaluation capacity.

	Source	Description	Components Related to Evaluation Capacity	Gaps or Limitations
3	**UNAIDS, 2010** *Guidance for Capacity Building for HIV Monitoring and Evaluation*	The purpose of this guidance is to provide practical advice for national AIDS programs that are planning and implementing capacity-building activities as part of their effort to develop a unified and effective national HIV M&E system.	The document provides a list of key knowledge, skills, and competency areas related to M&E in general and includes "evaluation methods" as one of the skills.	The guidance document focuses more on providing examples of capacity-building strategies that can be implemented instead of assessing capacity. However, it includes a useful list of competencies needed for M&E capacity, while being limited in terms of evaluation capacity specifically.
4	**United Nations Evaluation Group (UNEG), 2012** *National Evaluation Capacity Development: Practical Tips on How to Strengthen National Evaluation Systems*	The document presents a framework for understanding national evaluation capacity development, including clarification of terms and definitions. It also discusses four essential building blocks for an effective M&E system: 1. Vision of leadership 2. Enabling environment 3. Technical capacity to supply M&E 4. Capacity to demand and use M&E information	These tips include a list of "key considerations for success" related to "technical capacity," including: • Is there capacity (data systems and infrastructure) to collect reliable data and report credible information? • Is there adequate analytical capacity (that is, skilled personnel)? • Are there existing institutions that could serve as credible partners (for example, a national statistical agency or research institutes)?	The document is very broad, focusing on higher-level development of a national monitoring and evaluation system while not offering much guidance on evaluation capacity specifically.

CONCLUSION AND RECOMMENDATIONS

Our review of the literature was intended to identify specific tools and guidance to assess capacity to evaluate international public health programs. It revealed that the term evaluation capacity is defined and used in several ways. More specifically, it is often used to refer to the ability of organizations to evaluate internal capacity for the benefit of organizational learning, and not to their ability to evaluate public health programs—regarding which we found no existing tools or guidance documents. Instead, the tools and guidance documents we identified primarily concerned M&E in general and were all limited in the extent to which evaluation capacity specifically was included.

Based on these key findings, we recommend the following:

(1) When writing about and defining evaluation capacity, the type of evaluation capacity being referenced should be specified.
(2) A specific tool and guidance document to assess capacity for evaluating international public health programs needs to be developed to fill the gap in this area.
(3) Since international public health programs are implemented in different contexts with varying levels of in-country evaluation capacity, we recommend development of a tool specific to the assessment of evaluation capacity that is also generic enough to be adaptable to different contexts and programs.
(4) The tool should assess the capacity to implement different types of evaluations of public health programs, given that they each require different designs and methods (for example, rigorous impact evaluations, quasi-experimental evaluations, process evaluations, and formative evaluations). To do so, the different evaluation competencies or standards for the various types of evaluations will also need to be developed and assessed through the tool.
(5) As noted by Tarsilla (2014), international development capacity-building efforts have often been dedicated to improving individual-level capacity only and have had a short-term focus—for example, by providing a small group of individuals with the knowledge and skills needed to implement a specific evaluation. He argued that evaluation capacity development should be seen as a long-term effort aimed at both sustainability and empowerment at the individual, organizational, and system levels, which will foster greater country ownership and sustainability. These same three levels were also emphasized in some of the guidance documents we reviewed (including PEPFAR, 2012; UNAIDS, 2007; UNAIDS, 2010), and PEPFAR's capacity-building framework (PEPFAR, 2012) stressed the interrelationship of the three levels and the importance of strengthening capacity at all levels over time to increase sustainability. Given this, we recommend that the development of an evaluation capacity tool and guidance should also focus on multiple levels rather than only at the individual-level capacity of evaluators.
(6) Last, as the demand for evaluations of programs continues to increase, so too will the demand for professional local evaluators. We therefore recommend that the specific evaluation capacity assessment tool also include a component to quantify the existing in-country evaluation workforce (for example, within an organization or a larger system); this quantification component will help determine whether the existing number of evaluators is adequate for the number of evaluations planned. This, too, will encourage longer-term capacity-building strategies to be developed and implemented, thus contributing to the sustainability of evaluation capacity-strengthening efforts.

REFERENCES

Capacity for Health. (2011). *Monitoring and evaluation (M&E) capacity assessment tool.* Retrieved from

http://apiahf-c4h-wp-uploads.s3.amazonaws.com/wp-content/uploads/2015/05/Evaluation-Capacity-Assessment-Tool3.pdf

CDC. (1999). Framework for program evaluation in public health. MMWR 1999;48 (No. RR-11). Retrieved from http://www.cdc.gov/eval/framework/

CDC. (2011). *Introduction to program evaluation for public health programs. A self-study guide.* Retrieved from https://www.cdc.gov/eval/guide/

Cousins, J. B., Amo, C., Bourgeois, I., Chouinard, J., Goh, S.C., & Lahey, R. (2011). Organizational capacity to do and use evaluation: Results of a pan-Canadian survey of evaluators. *The Canadian Journal of Program Evaluation*, 23 (3), 1-35.

Gibbs, D., Napp, D., Jolly, D., Westover, B. & Uhl, G. (2002). Increasing evaluation capacity within community-based HIV prevention programs. *Evaluation and Program Planning*, 25, 261-269.

Gilliam, A., Barrington, T. Davis, D., Lacson, R., Uhl, G., & Phoenix, U. (2003). Building evaluation capacity for HIV prevention programs. *Evaluation and Program Planning*, 26, 133-142.

Informing Change. n.d. *Evaluation capacity diagnostic tool.* Retrieved from http://informingchange.com/cat-resources/evaluation-capacity-diagnostic-tool

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.

MEASURE Evaluation. (2014). *Monitoring and evaluation capacity assessment toolkit (MECAT).* Unpublished manuscript.

OECD. (2006). *The challenge of capacity development: Working towards good practice.* Paris, France: OECD. Retrieved from http://www.fao.org/fileadmin/templates/capacitybuilding/pdf/DAC_paper_final.pdf

PEPFAR. (2012). *Capacity building and strengthening framework.* Retrieved from http://www.pepfar.gov/documents/organization/197182.pdf

PEPFAR. (2014). *PEPFAR evaluation standards of practice.* Retrieved from http://www.pepfar.gov/reports/guidance/221111.htm

Preskill, H., & Boyle, S. (2008). A conceptual model of evaluation capacity building: A multidisciplinary perspective. *American Journal of Evaluation,* 29, 443-459.

Preskill, H., & Torres, R. (2000). The readiness for organizational learning and evaluation instrument (ROLE). Retrieved from http://www.fsg.org/tools-and-resources/readiness-organizational-learning-and-evaluation-instrument-role

RAND. (2005). Challenges in program evaluation of health interventions in developing countries. Retrieved from http://www.rand.org/content/dam/rand/pubs/monographs/2005/RAND_MG402.pdf

Stockdill, S. H., Baizerman, M., & Compton, D. W. (2002). Toward a definition of the ECB process: A conversation with the ECB literature. *New Directions for Evaluation,* 93. San-Francisco: Jossey-Bass.

Tarsilla, M., (2014). Evaluation capacity development in Africa: Current landscape of international partners' initiatives, lessons learned and the way forward. *African Evaluation Journal* 2 (1).

UNAIDS Monitoring and Evaluation Reference Group (MERG). (2008). *Organizing Framework for a functional national HIV monitoring and evaluation system.* Geneva: MERG. Retrieved from http://www.unaids.org/sites/default/files/sub_landing/files/20080430_JC1769_Organizing_Framework_Functional_v2_en.pdf

UNAIDS Monitoring and Evaluation Reference Group (MERG). (2010). *Guidance on capacity building for HIV monitoring and evaluation.* Geneva: MERG. Retrieved from http://www.unaids.org/sites/default/files/sub_landing/files/5_4_MERG_Guidance_HIV_ME_Capacity_Buidling.pdf

UNAIDS. (2007). *Building national HIV/AIDS monitoring and evaluation capacity: A practical guide for planning, implementing, and assessing capacity building of HIV/AIDS monitoring and evaluation systems.* Manuscript in preparation. Retrieved from http://gametlibrary.worldbank.org/files/1170_USG%20M&E%20system%20capacity%20building%20guide1.pdf

UNAIDS. (2007). *Checklist for assessing and monitoring capacity of the HIV/AIDS M&E system.* Retrieved from http://gametlibrary.worldbank.org/FILES/91_USG%20M&E%20System%20Capacity%20Assessment%20Guide%20ANNEXES.pdf

UNAIDS. (2010). *12 components monitoring and evaluation system strengthening tool.* Retrieved from http://www.unaids.org/sites/default/files/sub_landing/files/2_MERG_Strengthening_Tool_12_Components_ME_System.pdf

United Nations Evaluation Group (UNEG). (2012). *National evaluation capacity development: Practical tips on how to strengthen national evaluation systems.* Retrieved from http://www.uneval.org/document/detail/1205

USAID. (2011). *USAID evaluation policy.* Washington, DC: USAID. Retrieved from https://www.usaid.gov/evaluation/policy

USAID (2016). *Strengthening evidence-based development: Five years of better evaluation practice at USAID.* Washington, DC: USAID. Retrieved from https://www.usaid.gov/documents/1870/strengthening-evidence-based-development-five-years-better-evaluation-practice-usaid

Volkov, B., & King, J. (2007). *A checklist for building organizational evaluation capacity.* Retrieved from https://www.wmich.edu/sites/default/files/attachments/u350/2014/organiziationevalcapacity.pdf

WHO. (2001). *Evaluation in health promotion: Principles and perspectives.* Retrieved from http://www.euro.who.int/__data/assets/pdf_file/0007/108934/E73455.pdf

WORKING PAPER