Journal of MultiDisciplinary Evaluation
Volume 19, Issue 46, 2023

JMDE
Journal of MultiDisciplinary Evaluation

ISSN 1556-8180
http://www.jmde.com

Reflections on Required Competencies for Health Systems Monitoring, Evaluation, and Learning in India

Neethi Rao
Institute of Public Health Bengaluru

Devaki Nambiar
The George Institute for Global Health India

Background: The movement toward systems thinking and complexity-informed evaluation has been ongoing for some time. COVID-19 has increased the salience of contextually aware, adaptive forms of evaluation.

Purpose: Drawing very specifically upon experiences in the Global South, this article presents our reflections on competencies that have been integral to a collaboration involved with developmental health systems evaluation in India.

Setting: This work is based on our experience over 3 years of providing monitoring, evaluation, research and learning (MERL) to a large philanthropic organization's portfolio of health systems work in India.

Intervention: Not applicable.

Research Design: Not applicable.

Data Collection and Analysis: Not applicable.

Findings: We identify three types of evaluator competencies that derive from complexity-driven practice in a developing country.The first competency relates to deriving evidence in contexts where there are very few traditional forms of data. The second competency is related to skills required to play a convening role across varied stakeholders. The final competency relates to operational adaptiveness while evaluating an ecosystem with few constants. The pandemic created a natural experiment that required a reexamination of priorities in the unique context of reliable donor support in uncertain times. These three capacities are in fact interdependent and could only be developed iteratively, over time. Building competencies isn't simply about capacity building but rather requires a recognition of the diversity of skills and worldviews that need to be encompassed within our MERL functions for today's complex, discontinuous health systems.

Keywords: *complex adaptive systems, health system evaluation, Global South*

The movement toward systems thinking and complexity-informed monitoring and evaluation has been ongoing for some time (Gates, 2016; Hargreaves & Podems, 2012). From the 2000s onwards, scholars have recognized the importance of a developmental evaluation approach to support adaptive management of complex systems with dynamic and emergent properties (Patton, 2010; Patton, 2006). As Sridharan and Nakaima (2020) argue, such an approach is compatible with the principles and shifted lens introduced by a realist approach (Pawson & Tilley, 1997) to evaluation, where the focus shifts from "questions along the lines of 'does a program work?' to deeper interrogations into 'what is it about a program that makes it work [and for whom]?' " (Pawson & Tilley, 1997, p.2).

These shifted purposes and scopes of evaluation then introduce more complex and (yet) nimble designs and a larger range of methods, requiring, in turn, both wider and deeper sets of competencies on the part of evaluators (Gates, 2017; Sridharan & Nakaima, 2020). It is lamented, furthermore, that skills in project management, interpersonal communication, and presentation—which are required of evaluators—are not taught in training programs (Davies & MacKay, 2014; Dewey et al., 2008), even though the importance of these skills began to emerge in the 1990s (LaVelle, 2020). This is due in part to the multifarious nature of evaluation itself, the diverse contexts in which it has applied, as also perhaps the lack of a single unifying set of curricula or competencies required of evaluators (Brown, 2004; Gates, 2017; King & Ayoo, 2020). King and Ayoo's review of evaluator education, drawing mostly on literature from the Global North and focused on both formal and nonformal training, concluded that "a meaningful opportunity awaits researchers who care about understanding effective programs to prepare evaluators" (2020, p. 12). In service of this and drawing very specifically upon experiences in the Global South, we present our reflections on competencies that have been integral to a collaboration involved with developmental health systems evaluation in India.

The Context of Our Evaluation

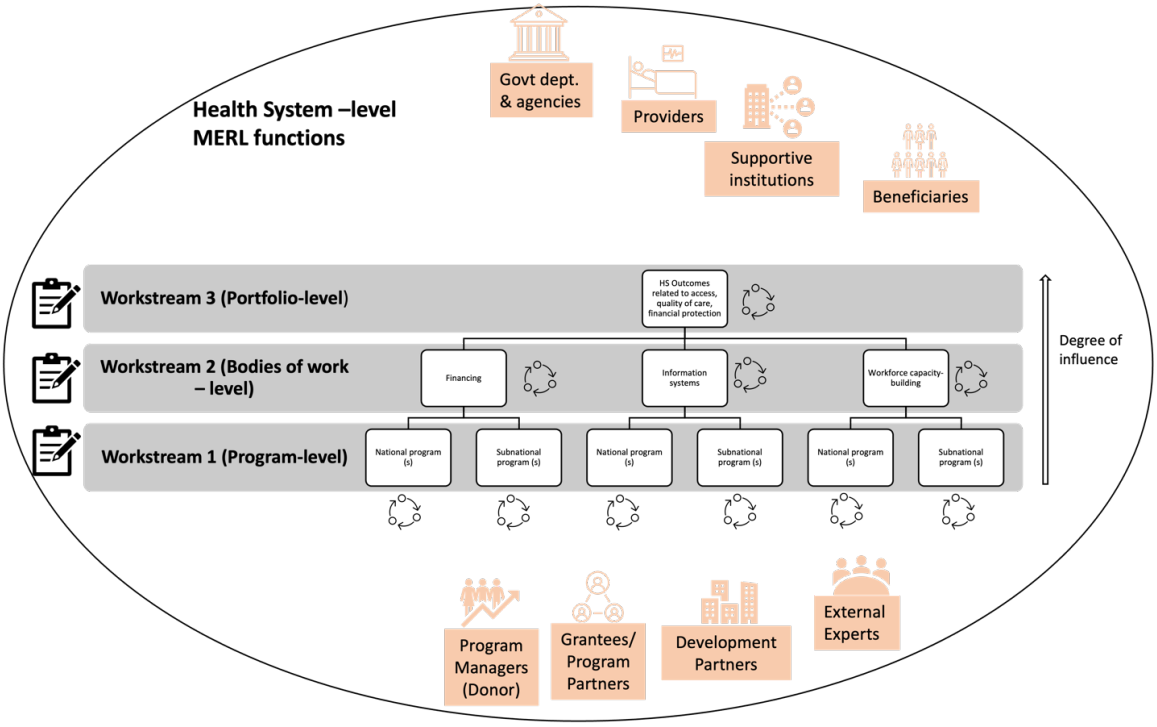
In 2016, a large private foundation with an office in India with a history of vertical funding in maternal and child health sought to integrate its portfolio by setting up a health systems design (HSD) team. In the first 2 years, the focus was on building the team and its leadership and vision as well as a strategic development process. By 2019, following a process of external consultation facilitated by a consulting firm, a draft strategy had been developed, alongside bodies of work (that is, clusters of individual projects) linked to health systems reform processes underway in India, such as health insurance, health service delivery, and formalization of health policy, regulation and incentives. Bodies of work ranged across health system building blocks and were primarily based on existing interests of program team members, who had been supporting partners and making investments, as well as on some strategic directions that experts advising the foundation saw as having potential.

At this point, a consortium of evaluators was convened by a U.K. agency and engaged by the foundation to provide monitoring, evaluation, research, and learning (MERL) functions to the team and its relatively small but growing set of investments. We, the authors of this paper, entered the fray as technical experts with experience in the Indian context and expertise in health systems research. Over the 2 subsequent years, the consortium carried out technical work to support an evaluation function, played a convening role to support a learning function for the foundation and its partners, and tried to fulfill a strategic and clearinghouse function in terms of supplying, on the one hand, contextual information for new players and stakeholders involved with the team, while also aiming to create spaces for frank discussion, reflection, and course correction in this area of the foundation's work on the other.

The design of this MERL project consisted of nested evaluations, serving varying purposes, at different levels: We distinguished these into three work streams (see Figure 1). Broadly, the design was of a series of nested evaluations—starting from the programs at national and sub-national levels, which represent individual grants made by the philanthropic organization (donor). First, at the most local level, there were individual grants, investments that the foundation had made opportunistically and wanted to evaluate the impact of. These grant programs were pooled into bodies of work based on a strategic development exercise; the bodies of work were roughly aligned with the health system building blocks (De Savigny & Adam, 2009). The second work stream was evaluation of these bodies of work with a view to building synergies across existing grants, combined with an exploratory evaluation approach for new domains of investment. Together, grants across all of the bodies of work made up the health systems design (HSD) portfolio of the donor. The final work stream was related to the creation of an evaluation and learning ecosystem for the HSD portfolio and

for health systems in the country more broadly. Here again, to some extent, there were existing investments but also emerging ambitions around creating learning opportunities across grants and bodies of work—both within the foundation and its network of funded partners and beyond. As we moved up from individual grants to bodies of work to the portfolio, the degree of influence that the donor and its partners had on overall health system outcomes diminished and correspondingly the theories of change as well as the associated signals and indicators used by the MERL team became more broad-based and less specific. A range of stakeholders were involved across work streams and are also depicted in Figure 1.

Figure 1. Representative Image Depicting the Overall Scope of Work for the MERL Consortium



Note. Specific bodies of work are only meant to be representative and in the interest of confidentiality do not refer to the specific bodies of work of the donor.

Around the time this work was being done, a flagship report on learning health systems (Sheikh & Abimbola, 2021) emerged, offering a great frame of reference to interpret and more deeply analyze our activities. A posteriori, we are able to see that across these work streams, we were aiming to create individual (i.e. within individual grant teams), team and group (i.e. across bodies of work), and organization and cross-organization levels (meaning between the foundation, its grantees and their partners). The Learning Health Systems model is based on systems learning and systems thinking, acknowledging that health systems are complex adaptive systems with emergent outcomes that are the result of multiple, interacting agents, often unpredictable in nature. This was is in line with developmental evaluation approaches as well as the acknowledged emphasis on the situational nature of evaluative practice. For instance, while it was already understood that the function of the evaluator in this instance was as a promoter of organizational learning (Preskill & Torres, 1999), the boundary around "organization" was loose and included a funder, grantees, and their partners, as well as other domain experts who could weigh in to multiple work streams. As Clinton and Hattie point out, the cognitive complexity of conducting such evaluations can be quite high, traversing "knowing that" and "knowing how" attributes (Clinton & Hattie, 2021).

Competencies

The delivery of this complex and adaptive evaluation required a series of competencies, of which we focus on three here. We chose these three for their key role in enabling such complexity-driven practice, and also to highlight the need for greater training and capacity building of evaluators in these areas. Each of these three competencies correspond to our operational reality in the Global South. While the need for these competencies—and, indeed, the challenges and deficiencies of health systems they pertain to—may not be confined to the Global South, our experience is limited to this context. Evaluative exercises in such settings are rarely championed by governments themselves and can often be supported by external funders and practitioners who enjoy greater resources, power, and privilege than the health system practitioners who are the subjects of the evaluation. Further, scholarship and evidence in relevant domains, such as social determinants of health care usage or epidemiological trends, may be underdeveloped or absent. We reflect here on competencies that are likely to be especially important in conducting systems evaluation with a large scope in a similar context to help orient training in evaluation. We note that much of the literature and thinking on evaluative capacity building has been from the Global North (Nielsen et al., 2023) and hope that documenting our experiences will help collectively inform evaluation capacity-building efforts, especially in the Global South. To be clear, we do not wish to suggest that our experience was unique in any way or that there were novel competencies identified through this exercise. Rather, we use our experience to demonstrate how these three types of competencies were applicable for a health systems evaluation and why they were relevant in the Indian context.

Deriving Evidence in Data-Poor Contexts

The first competency relates to **deriving evidence** in contexts where there are very few traditional forms of data. India's health data systems are notorious for being fragmented—with problems of data quality and lack of interoperability—and often publicly inaccessible (Khurana, 2021; Pandey et al., 2010; Rao et al., 2021). Even data related to flagship schemes such as the Pradhan Mantri Jan Arogya Yojana (PM-JAY; National Health Authority, 2019), which boasts an advanced digital architecture to support data-driven analysis, remains publicly inaccessible and siloed away from other existing data repositories, such as the health management information system. The MERL functions involved assessment of health systems outcomes of access, financial protection, and equity that required combining different kinds of data sets that were held with different government agencies, differed in quality, and were difficult to integrate. There were also large gaps in data, in part due to the inability of data systems to keep up with the rapid transformations in both policy priorities and implementation mechanisms.

A large part of the evidence gathered was thus qualitative, seeking to understand the perspectives of different actors within the system as individuals and as representatives of their organizations. Qualitative data collection took the form of interviews and focus group discussions. Findings were then compared with analysis of program documents and reports. Informal events, partner meetings, and workshops were additional ways to help synthesize experiential and tacit knowledge with more formal data across multiple components of the Indian health system. The goal was not to identify a single source of truth but rather to piece together a patchwork of validated information, wherein the relevance of different pieces of data was dependent on evolving outcomes of interest for the various stakeholders, as is common in evaluations of complex interventions (Gates, 2016; McGill et al., 2021).

Since the investments were responsive to fast-evolving government needs, the evaluators together with the project partners needed to first set out the underlying theory of change, sometimes working backwards from actions to uncover previously tacit assumptions about intended impacts—in other words, to bring to light the underlying theory or logic that explains how current investments—largely consisting of technical assistance to various government agencies—contribute to health system goals of financial protection, equity, and learning. Clearly this is not a linear path and is an open system involving a variety of stakeholders (including the government) and their own independent actions outside the remit of the investments themselves. Information from each of these disparate sources and perspectives needed to be gathered, joined up, and analyzed within the shared health system context. We also needed to obtain these perspectives over time, going back sometimes to the same set of stakeholders to ask similar questions within somewhat changed circumstances. Further, health system interventions by their very nature

involve interactions (often unpredictable ones) with other parts of the system, requiring longer time frames for outcomes and impacts to become apparent. Thus, the evaluators needed to consider possible trajectories, with milestones to determine degrees of success and impact, iteratively tweaking the assumptions underlying theories of change as well as the methodologies (Morell, 2018).

It bears mentioning that it was not always possible to assemble a coherent "story" or understanding of what was happening across these types of stakeholders, because information, networks, and data did not cover all the domains of interest or inquiry. For example, we were interested in understanding more about how national insurance schemes were faring in response to changed health needs and requirements in the COVID context. The evaluation team made attempts to actively incorporate researchers and practitioners and created processes to support ongoing dialogue between the two. This was not successful. We were unable to quantify or characterize this information, as key individuals were unreachable and specifically relevant data points inaccessible. Thus, even as the requirements of time and human resources to conduct effective evaluations in complex, adaptive systems were inevitably higher, the yield was not often commensurate with the effort put in, reflecting perhaps on the need to further develop these competencies among the evaluators, especially in times of disruption.

The MERL team consciously engaged different kinds of researchers with differential expertise, with the goal of building a much more pixelated picture. The MERL consortium included health economists, policy analysts, health system researchers, and policy practitioners. Thus, the team included a substantial number of experts who considered evaluations only a part of their overall practice and research. Formal data gathering was supplemented with extensive discussions, especially for validating the dynamic contextual factors to support more nuanced interpretation, recognizing that none among the team could claim complete knowledge or understanding. This meant going beyond simply methodological pluralism to purposively amplifying team members with different kinds of training and worldviews through deliberative practice and open debate. We were not always successful, of course. For example, identifying individuals familiar with evaluation methodology (to understand what was to be measured) as well as health system data characteristics (to determine where we would find information on it) in the Indian context was not possible. In some cases informal modes of deriving evidence were considered based on our networks, but acted upon only when necessary, based on steering and requests from the donor agency. This at times set some competencies (for example, the United Nations Evaluation Group's competency related to evaluation) in contest with contextual or communication/interpersonal skills (Vaessen et al., 2022).

Convening Across Stakeholders

The project sought to understand potential differences in perspectives of various stakeholders, regarding both the ends and the means. Stakeholders included the project implementing partners, governmental stakeholders at national and state (provincial) levels, and the program officers and health system portfolio managers at the donor organization. The perceptions of what constituted success varied for each of these stakeholders and, additionally, changed over time.

The implementers were focused rather understandably on the previously agreed upon project parameters. Program officers were concerned about implementation processes and potential prospects for future engagement in the area. Governmental stakeholders were fixated on the extent to which the external partners added value to their own areas of work. Donors had the broadest perspective and were most interested in the impact of the work, both real and perceived, on India's health system. This meant that the MERL consortium needed to engage differently with each of these stakeholders, presenting information in formats and contexts that reflected, revealed, and sometimes challenged these differing interests. Competencies in the interpersonal domain of existing competency frameworks—for example, those put forth by the American Evaluation Association (AEA) and the Canadian Evaluation Society (CES), among others—were thus key to the careful management of the stakeholder network over time through appropriately tailored communications.

The skills required to play **a convening role** for donors and implementing partners, support better understanding of the changing operating context, and help inform decision-making by program officers and partners were thus crucial. This involved building and sustaining relationships across different stakeholders at different stages of the project—from proposal development to review—while situating the evaluation within its programmatic and health system context (Garcia & Stevahn, 2020). The team organized several rounds of meetings and workshops that brought together

different mixes of relevant stakeholders. These meetings were carefully designed to support reflective discussions directly relating to the program of work as well as aimed at creating a sense of community beyond the immediate project parameters. However, there were inherent tensions, with actors having different degrees of association and involvement in prior interactions and engagements with the donor. This was particularly true for subnational stakeholders. Competencies relating to effective dialogue and developing an understanding of the core interests of international, national, and subnational partners as well as international donors and experts could only be developed iteratively and over time, but this was crucial in a distributed health decision-making ecosystem like India.

Once again, different members of the MERL team were called upon to anchor activities with different objectives with the overall aim of combining a range of insights. While some MERL consortium members were more focused on engaging governmental decision-makers, others had their ears closer to the ground with technical implementers. These aspects were then coordinated through the consortium lead organization, which provided both technical and logistical support while managing donor relationships. Given the diverse nature of the MERL consortium itself, the team also needed to invest time in collective reflection and consensus-building internally.

The positionality and composition of the consortium, then, became very critical. There are a couple of nuances of convening that are critical to mention here. The first is about power: While the remove of the consortium lead was seen to be essential in terms of the evaluative function, there was also consciousness about the lead being based outside of India and potential concerns of rehearsed power asymmetries that privileged Global North institutions. There could also have been the perception that this structure did not lend itself to understanding the sensitivities, peculiarities, customs, and contingencies in India and the geographies within it. While efforts were made to have distributed leadership and decision-making and to have "boots on the ground," this remained challenging and partners continued to perceive this as a factor bearing on the evaluation activities. Relatedly, it is unclear if the evaluation helped sufficiently strengthen the capacities of local practitioners in ways that supported ongoing developmental evaluative practice within India. The second is about representation: At various stages of the project the evaluators were specifically told NOT to engage with decision-makers or other key systems stakeholders in the projects and processes because of the possible strategic or reputational risks involved. The third is about scope: Breaking up the portfolio into domain areas (determined by the donor, somewhat artificially) meant that we were unable to ask larger questions related to governance (e.g. decentralization, conflicts of interest, positionality of partners) that are likely cross-cutting across all projects and a key factor in influencing the outcomes.

Operational Adaptiveness

The third and final competency relates to **operational adaptiveness** while evaluating an ecosystem with few constants. This can be the hardest competency to acquire because it is the farthest from the traditional notions embedded in the training of evaluation scientists and MERL practitioners. We found that the degree of buy-in and commitment to previously agreed upon frames of reference for evaluation can be shifted by changes in personnel or internal organizational structures. These shifts can lead to chain reactions of mismatched expectations that needed to be understood and managed in real time by MERL partners. Skills of interpretation, judgment, and diplomacy were called upon to reconcile different notions of what constitutes success or failure across different levels and stakeholders. For example, project implementers remained committed to previously identified areas of work, such as those related to improving insurance processes for tertiary care or standards of care for noncommunicable diseases, even as the priorities for other stakeholders (typically influential individuals) in the government and the donor organization had shifted to emphasizing, for example, primary health care, diagnostics, and digital data systems. Evaluators then had to support actors in reconciling the ongoing activities and expertise with shifting priorities in ways that built on the progress that had already been made.

The evaluation period was also interrupted by the emergence of COVID-19, which added to the uncertainty of the context while creating operational challenges. The pandemic created a natural experiment that on the one hand required a reexamination of program priorities and on the other depended on reliability of donor support. Implementing partners were called upon to rapidly redeploy existing resources toward short-term logistical needs related to the COVID response. Even as planned program activities were de-emphasized in some cases, close coordination with governmental stakeholders in supporting urgent

needs led to the building of trustful relationships. However, health system outcomes related to equitable access and social justice could not always be prioritized by project partners as part of the emergency response. Within our own consortium, there were very few people whose own personal lives were unaffected by COVID. Any conscious effort to redress inequities was more of a personal endeavor than a professional one as attention turned to individual, family, and institutional health concerns. The decentralization of the COVID response effort also revealed the relatively higher degree of centralization inherent in the HSD portfolio, raising questions about sustainability. The evaluator has a key role of observing these patterns and interpreting how they serve the existing goals partners may have vis-à-vis a project or body of work and what emergent properties exist. The evaluator also needs to ask how these patterns may challenge/subvert assumptions and even the trajectories that projects and portfolios are on.

The MERL consortium also needed to pivot to support donors and partners in adapting to changed operational realities by identifying minimum essential features that can still be leveraged to manage operational discontinuities and new potential opportunities to add value toward the larger health system goals. This aligns with the World Bank Group's Evaluation Principles (2019) (of responsive planning and strategic selectivity. For example, capacity-building programs could be quickly expanded based on fast-developing needs; data systems created previously could also now be utilized to identify neglected or vulnerable populations for the pandemic response. Skills related to systems thinking that allowed evaluators to focus on the big picture needed to be deployed in combination with a deep understanding of practical details on the ground. Once again, MERL functions relied critically on the combined talents of a team of diverse specialists to manage the health system–wide ambitions of the donor organization and its partners.

In addition to COVID-induced discontinuities, the project also underwent several changes in key personnel, including attrition among the program officers and senior government policy makers, which led to fairly significant changes in priorities and objectives for evaluation. The MERL team worked intensively with partners across projects on bodies of work to reexamine and adjust the strategies funded under those bodies of work to align with renewed priorities. In some cases, the MERL team's momentum dithered as it was unclear who our main client was and how to negotiate the changing dynamics and requirements of key actors across levels. Given these discontinuities, we placed emphasis on our own management as a consortium, trying to keep up with the changes and evolving needs and calibrating how we could add value and reallocate roles within our own team. We recognized the possible positive and negative impacts of carrying out evaluative activities at a time of flux and elected to instead facilitate knowledge sharing through webinars and informal meetings and adopt the role of a listener (Competencies 3.5 and 3.6 within CES's (2018) Situational Practice domain. Project managers of evaluation teams thus have a very key role in deploying the expertise within the team in an adaptive, responsive, and respectful manner while maintaining transparency across the whole team.

Discussion & Conclusions

In this paper we draw from our specific experience as evaluators as part of a consortium, supporting a philanthropic foundation's activities in the Global South. However, the three competencies are in fact applicable more broadly, going beyond the details of the project(s) under question, and indeed reflected in the evaluation literature we found (which, admittedly, has few specific references to our context per se). Complex interrogations of the sort discussed in this paper are required for complex fields of inquiry in evaluation, such as health systems strengthening and/or reform. The World Health Organization has broadly defined health system strengthening as "any array of initiatives that improves one or more of the functions of the health system and that leads to better health through improvements in access, coverage, quality, or efficiency" (Witter et al., 2019, p. 10). The elements of interest in a evaluation of health systems–strengthening efforts are commensurably multifarious, as our own experience demonstrated.

For domains like health systems evaluation, which deals with wicked problems, we have been largely building the ship as we are sailing it. More recent guidance on delineating both the components and the connections across various health system domains and functions—for example, the World Health Organization's *Health System Performance Assessment: A Framework for Policy Analysis* (2022)—will likely help in systematizing evaluations of health systems as well as providing greater alignment/coherence in terms of competencies. However, where a (funding) initiative seeks to operate across levels or functions of the health system, it will continue to present unique strategic and operational challenges. The

pathway toward a learning health system requires ongoing iterative improvement in different parts of the system, operating at multiple different levels through developmental evaluation (Shah et al., 2021).

The three competencies described in this article have indeed been outlined in existing frameworks of evaluation competencies developed by, for example, AEA (2018) and CES (2018). Our experience demonstrates how these competencies in fact operate interdependently in a real-world situation, emphasizing that competencies ought to be thought of (and taught) as fluid domains rather than specific elements. For example, our first competency, which relates to drawing evidence in data-poor settings, may be considered at first glance under AEA's Methodology domain or CES's Technical Practice domain. But in practice, it relied substantially on knowledge of the context (AEA's Context domain) and situational analysis, as well as drawing on skills of professional practice in identifying appropriate approaches and theories for the evaluation (AEA's Professional Practice domain). Similarly, our second competency (convening) is much more in the Interpersonal domain but required substantial skills in the Planning and Management domain to ensure coordination within the MERL consortium as well as accountability toward all stakeholders. The third competency (adaptiveness) keeps in mind professional practice (AEA's first domain) while focusing on interpersonal relationships (the subject of AEA's fifth domain) to facilitate shared redefinition of project priorities. So even as we endeavor to identify specific competencies, training has to be oriented toward evaluative thinking that supports an integrated conception of all evaluator competencies that can then be applied to different degrees as needed in different program evaluations (Clinton & Hattie, 2021).

As highlighted in virtually all existing evaluator competency frameworks, understanding of the context is an essential competency that forms the bedrock of all evaluative practice. Deep knowledge of context is necessary for understanding how the context (infrastructural, institutional, interpersonal, individual, and intersectional) might support or hinder the intervention(s; Shah et al., 2021). We found, for example, that technical assistance was being provided to improve the functioning of the health insurance program in India, mainly through supply-side interventions. However, there were legitimate reasons to hypothesize that the principal barriers to efficient and effective health care in these geographies were in fact demand- and supply-side constraints that were not covered through the interventions (Fadlallah et al., 2018; Saxena et al., 2019; Thakur, 2016). Even where data sets related to demand-side factors were unavailable, we sought to find other signals and proxy indicators to understand the impact of existing project activities.

Knowledge of the context can also help anticipate potential secondary or emergent impacts of the interventions that may or may not be desirable in the larger health system. For example, in our case, the same government officer was often in charge of increasing enrollment of private providers under the health financing program as well as regulating cost and quality of private health care, creating possible operational trade-offs in fulfilling their fiduciary duties. Interventions focused on improving enrollment without considering impacts on regulatory capacity can create greater problems for the overall health system than the gains obtained through increasing enrollment in health insurance programs. This recognition implied bringing together project partners focused on these separate pathways and supporting them to work together (convening/interpersonal practice).

Our work has potential implications for both evaluative practice and training. There needs to be much more emphasis on on-the-job training, including informal forums and conferences where early-career/novice evaluators can interact with more seasoned professionals (Garcia & Stevahn, 2020). Developmental systems evaluation is inevitably situated within research and learning functions, since it requires iterative and long-term work rather than episodic evaluation exercises. Recognition of this can also help reduce "evaluation anxiety" across all stakeholders (Donaldson et al., 2002). Evaluation teams need to include local practitioners and subject matter experts, with a combination of insiders and outsiders who are equally empowered. Finally, there needs to be a recognition that interventions are ultimately aimed to strengthen (health) systems as a whole and therefore evaluators must work closely with evaluation commissioners to ensure that goals of evaluation are aligned with goals of health systems as a whole—negotiating values as much as activities/outputs (House, 2015).

There are also challenges associated with these kinds of complexity-informed evaluations. They are time-consuming, need both methodological and substantive expertise, make it difficult to empirically attribute changes in outcomes directly to interventions, and can be challenging both for evaluators and for evaluation commissioners.

We are also cautious in recognizing that our experience may not be typical of program

evaluations. The scale and scope of this evaluation was rather large, constituting several programs across multiple domains as well as incorporating a systems lens (Figure 1). The resources at disposal for such a MERL exercise were significant, which is unusual for evaluations and limits the generalizability of our reflections. The MERL consortium was led by an organization based outside of India, in partnership with Indian organizations and individuals who were located in different parts of the country. The reflections presented here are from only two members of the consortium, neither of whom are professional evaluators, and likely represent a limited perspective on the complexities and learnings from the evaluation. The positionality of the authors is also constrained, since both authors experienced periods when they were unable to be involved in the MERL exercise (for different reasons). It is also not at all clear that there were uniformly positive impacts from the evaluation for the Indian health system, or indeed the evaluation commissioners. Given the changes induced by COVID and also changes within the donor organization and the government, the relevance and utility of the evaluation findings also remains to be seen.

There is an increasing amount of literature arguing for complexity-informed practice that places inordinate demands on the evaluator (Barnes et al., 2003; Sridharan & Nakaima, 2020; Vang et al., 2021). The range and skills required of evaluators to navigate the kind of complexity that is the norm today is ever expanding (Diaz et al., 2020). As we deliberate about what shape evaluation capacity building ought to take in such an environment, we argue in fact that evaluation capacities are about building task-appropriate teams. The evaluation function ought to be thought of more seriously as a team exercise with associated hierarchies and divisions of labor. Evaluator training thus ought to more consciously include the ability to engage closely with practitioners and other kinds of researchers in evaluative questions, as has been advocated before (Grack Nelson et al., 2019; Preskill & Boyle, 2008).

An additional area of reflection regards the expectations from the evaluation. Realist or complexity-informed evaluations by their very nature are unable to offer simplistic prescriptions of what needs to be done (or not done). This can create challenges for the program funders and implementers in understanding what can be done with the results of the evaluation. Evaluators may thus need additional competencies in anticipating possible implications of the evaluation for each stakeholder group and working closely with each of them to modify theories of change or action plans over the short term as well as over the long term.

The role of evaluation commissioners in facilitating the use of systems approaches has been recognized by some authors. For example, Gates (2017) suggests "new ways of constructing an evaluation contract that allow for flexibility and adaptability in evaluation design" (p. 167). But as we saw in our experience, complexity-informed evaluations in fact place considerable demands on the commissioners across all stages of the evaluation and, indeed, even after the results of the evaluation are available. More careful consideration of and clarity about the objectives of the evaluation in the near term for the project itself, in the medium term for the funders and implementers, and in the long term for the health system is necessary to obtain value out of such exercises. Even as funders and policy makers are increasingly demanding value-for-money and evidence-based practice, the complexity and network effects might necessitate a much softer, iterative developmental evaluation approach that reinforces existing learning pathways within health systems rather than building parallel infrastructures for evaluation.

References

American Evaluation Association. (2018). *The 2018 evaluator competencies*. https://www.eval.org/Portals/0/Docs/AEA%20Evaluator%20Competencies.pdf

Barnes, M., Matka, E., & Sullivan, H. (2003). Evidence, understanding and complexity: Evaluation in non-linear systems. *Evaluation*, *9*(3), 265–284. https://doi.org/10.1177/13563890030093003

Brown, E. (2004). Effective training of program evaluators: A mixture of art and science. *New Directions for Program Evaluation*, *1980*, 79–87. https://doi.org/10.1002/ev.1263

Canadian Evaluation Society. (2018). *Competencies for Canadian evaluation practice* (Revised version). https://evaluationcanada.ca/files/pdf/2_competencies_cdn_evaluation_practice_2018.pdf

Clinton, J. M., & Hattie, J. (2021). Cognitive complexity of evaluator competencies. *Evaluation and Program Planning*, *89*, 102006. https://doi.org/10.1016/j.evalprogplan.2021.102006

Davies, R., & MacKay, K. (2014). Evaluator training: Content and topic valuation in university evaluation courses. *American*

Journal of Evaluation, *35*(3), 419–429. https://doi.org/10.1177/1098214013520066

De Savigny, D., & Adam, T. (2009). *Systems thinking for health systems strengthening*. World Health Organization.

Dewey, J. D., Montrosse, B. E., Schröter, D. C., Sullins, C. D., & Mattox, J. R. (2008). Evaluator competencies: What's taught versus what's sought. *American Journal of Evaluation*, *29*(3), 268–287. https://doi.org/10.1177/1098214008321152

Diaz, J., Chaudhary, A. K., Jayaratne, K. S. U., & Assan, E. (2020). Expanding evaluator competency research: Exploring competencies for program evaluation using the context of non-formal education. *Evaluation and Program Planning*, *79*. https://doi.org/10.1016/j.evalprogplan.2020.101790

Donaldson, S. I., Gooler, L. E., & Scriven, M. (2002). Strategies for Managing Evaluation Anxiety: Toward a Psychology of Program Evaluation. *American Journal of Evaluation*, *23*(3), 261-273. https://doi.org/10.1177/109821400202300303

Fadlallah, R., El-Jardali, F., Hemadi, N., Morsi, R. Z., Abou Samra, C. A., Ahmad, A., Arif, K., Hishi, L., Honein-Abouhaidar, G., & Akl, E. A. (2018). Barriers and facilitators to implementation, uptake and sustainability of community-based health insurance schemes in low- and middle-income countries: A systematic review. *International Journal for Equity in Health*, *17*(1). https://doi.org/10.1186/S12939-018-0721-4

Garcia, G. L., & Stevahn, L. (2020). Situational awareness and interpersonal competence as evaluator competencies. *American Journal of Evaluation*, *41*(1), 107–124. https://doi.org/10.1177/1098214018814941

Gates, E. F. (2016). Making sense of the emerging conversation in evaluation about systems thinking and complexity science. *Evaluation and Program Planning*, *59*, 62–73. https://doi.org/10.1016/J.EVALPROGPLAN.2016.08.004

Gates, E. F. (2017). Learning from seasoned evaluators: Implications of systems approaches for evaluation practice. *Evaluation*, *23*(2), 152–171. https://doi.org/10.1177/1356389017697613

Grack Nelson, A., King, J. A., Lawrenz, F., Reich, C., Bequette, M., Pattison, S., Kunz Kollmann, E., Illes, M., Cohn, S., Iacovelli, S., Cardiel, C. L. B., Ostgaard, G., Goss, J., Beyer, M., Causey, L., Sinkey, A., & Francisco, M. (2019). Using a complex adaptive systems perspective to illuminate the concept of evaluation capacity building in a network. *American Journal of Evaluation*, *40*(2), 214–230. https://doi.org/10.1177/1098214018773877

Hargreaves, M. B., & Podems, D. (2012). Advancing systems thinking in evaluation: A review of four publications. *American Journal of Evaluation*, *33*(3), 462–470. https://doi.org/10.1177/1098214011435409

House, E. R. (2015). *Evaluating: Values, biases, and practical wisdom*. IAP.

Khurana, N. (2021). Issue analysis: A use-driven approach to data governance can promote the quality of routine health data in India. *Global Health: Science and Practice*, *9*(2), 238–245. https://doi.org/10.9745/GHSP-D-20-00347

King, J. A., & Ayoo, S. (2020). What do we know about evaluator education? A review of peer-reviewed publications (1978–2018). *Evaluation and Program Planning*, *79*, 101785. https://doi.org/10.1016/j.evalprogplan.2020.101785

LaVelle, J. M. (2020). Educating evaluators 1976–2017: An expanded analysis of university-based evaluation education programs. *American Journal of Evaluation*, *41*(4), 494–509. https://doi.org/10.1177/1098214019860914

McGill, E., Er, V., Penney, T., Egan, M., White, M., Meier, P., Whitehead, M., Lock, K., Anderson de Cuevas, R., Smith, R., Savona, N., Rutter, H., Marks, D., de Vocht, F., Cummins, S., Popay, J., & Petticrew, M. (2021). Evaluation of public health interventions from a complex systems perspective: A research methods review. *Social Science & Medicine*, *272*, 113697. https://doi.org/10.1016/j.socscimed.2021.113697

Morell, J. A. (2018). Systematic iteration between model and methodology: A proposed approach to evaluating unintended consequences. *Evaluation and Program Planning*, *68*, 243–252. https://doi.org/10.1016/j.evalprogplan.2017.09.006

National Health Authority [India]. (2019). *Best practices and innovations—PMJAY implementation*.

Nielsen, S. B., Lemire, S., Bourgeois, I., & Fierro, L. A. (2023). Mapping the evaluation capacity building landscape: A bibliometric analysis of scholarly communities and themes. *Evaluation and Program Planning*, *99*, 102318. https://doi.org/10.1016/j.evalprogplan.2023.102318

Pandey, A., Roy, N., Bhawsar, R., & Mishra, R. (2010). Health information system in India: Issues of data availability and quality 1. *Demography India*, *39*, 111–128.

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

Patton, M. Q. (2006). Evaluation for the way we work. *Nonprofit Quarterly*, *13*(1), 28–33.

Pawson, R., & Tilley, N. (1997). An introduction to scientific realist evaluation. In E. Chelimsky & W. R. Shadish (Eds.), *Evaluation for the 21st century: A handbook* (pp. 405–418). Sage Publications, Inc. https://doi.org/10.4135/9781483348896.n29

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

Preskill, H., & Torres, R. T. (1999). *Evaluative inquiry for learning in organizations*. Sage Publications, Inc. https://doi.org/10.4135/9781452231488

Rao, N. V., Prashanth, N. S., & Hebbar, P. B. (2021). Beyond numbers, coverage and cost: Adaptive governance for post-COVID-19 reforms in India. *BMJ Global Health*, *6*(2), e004392. https://doi.org/10.1136/bmjgh-2020-004392

Saxena, N., Singh, P., & Mishra, A. (2019). A qualitative comparative analysis of healthcare Supply-Demand side barriers under the publicly funded health insurance scheme in India. *Indian Journal of Public Health*, *63*(4), 318–323. https://doi.org/10.4103/IJPH.IJPH_409_18

Shah, N., Mathew, S., Pereira, A., Nakaima, A., & Sridharan, S. (2021). The role of evaluation in iterative learning and implementation of quality of care interventions. *Global Health Action*, *14*(1). https://doi.org/10.1080/16549716.2021.1882182

Sheikh, K., & Abimbola, S. (Eds.). (2021). *Learning health systems: Pathways to progress*. Alliance for Health Policy and Systems Research.

Sridharan, S., & Nakaima, A. (2020). Valuing and embracing complexity: How an understanding of complex interventions needs to shape our evaluation capacities building initiatives. *Evaluation and Program Planning*, *80*, 101440. https://doi.org/10.1016/J.EVALPROGPLAN.2017.05.001

Thakur, H. (2016). Study of awareness, enrollment, and utilization of Rashtriya Swasthya Bima Yojana (National Health Insurance Scheme) in Maharashtra, India. *Frontiers in Public Health*, *3*. https://doi.org/10.3389/FPUBH.2015.00282

Vaessen,J., Joppert,M., Stockman,R. (2022) Educating Evaluation Professionals: Core Knowledge and Competencies. https://www.globalevaluationinitiative.org/blog/educating-evaluation-professionals-core-knowledge-and-competencies

Vang, K. Y., Moore, M., & Nicklin, C. (2021). Applying a systems and complexity lens to building evaluation capacity: Learning from a multicountry donor's experience. *New Directions for Evaluation*, *170*, 101–111. https://doi.org/10.1002/EV.20463

Witter, S., Palmer, N., Balabanova, D., Mounier-Jack, S., Martineau, T., Klicpera, A., Jensen, C., Garcia, M. P., & Gilson, L. (2019). *Evidence review of what works for health systems strengthening, where and when?* FCDO. https://www.rebuildconsortium.com/wp-content/uploads/2019/08/HSS-revision-03-03-21.pdf

World Bank Group Evaluation Principles. (2019) https://ieg.worldbankgroup.org/sites/default/files/Data/reports/WorldBankEvaluationPrinciples.pdf

World Health Organization, European Observatory on Health Systems and Policies, Papanicolas, I., Rajan, D., Karanikolos, M., Soucat, A., & Figueras, J. (2022). *Health system performance assessment: A framework for policy analysis*. World Health Organization. https://apps.who.int/iris/handle/10665/352686