

https://dspace.org/

Gender responsiveness diagnostic of national monitoring and evaluation systems – methodological reflections

Caitlin, Blaser Mapitsa; Madri S., Jansen van Rensburg

2017-04-26
African Evaluation Journal
https://hdl.handle.net/10539/28253

Downloaded from a DSpace repository. https://dspace.org/

African Evaluation Journal
ISSN: (Online) 2306-5133, (Print) 2310-4988

Gender responsiveness diagnostic of national monitoring and evaluation systems – methodological reflections

Authors:
Madri S. Jansen van Rensburg[1,2]
Caitlin Blaser Mapitsa[3]

Affiliations:
[1]Africa Gender and Development Evaluators Network, South Africa

[2]Resilience Analysis Consulting, South Africa

[3]Centre for Learning on Evaluation and Results (CLEAR), University of the Witwatersrand, South Africa

Corresponding author:
Caitlin Mapitsa,
caitlin.blaser@wits.ac.za

Dates:
Received: 11 Nov. 2016
Accepted: 13 Feb. 2017
Published: 26 Apr. 2017

How to cite this article:
Jansen van Rensburg, M.S. & Blaser Mapitsa, C. 2017, 'Gender responsiveness diagnostic of national monitoring and evaluation systems – methodological reflections', *African Evaluation Journal* 5(1), a191. https://doi.org/10.4102/aej.v5i1.191

Copyright:
© 2017. The Authors. Licensee: AOSIS. This work is licensed under the Creative Commons Attribution License.

Read online:
Scan this QR code with your smart phone or mobile device to read online.

Background: This article reflects on the implementation of a diagnostic study carried out to understand the gender responsiveness of the national monitoring and evaluation (M&E) systems of Benin, South Africa and Uganda. Carrying out the study found that the potential for integrating the cross-cutting systems of gender and monitoring and evaluation (M&E) are strong. At the same time, it highlighted a range of challenges intersecting these two areas of work. This article explores these issues, which range from logistical to conceptual.

Objectives: This article aims to share reflections from the gender diagnostic study to enable more appropriate capacity building in the field of gender responsiveness in national M&E systems. Developing more sophisticated tools to measure gender responsiveness in complex contexts is critical. A better understanding of how gender and national M&E systems intersect is important to understanding firstly how we can more accurately measure the gender responsiveness of existing systems and secondly how better to engender capacity development initiatives.

Method: As part of the Twende Mbele programme, Centre for Learning on Evaluation and Results (CLEAR) commissioned Africa Gender and Development Evaluator's Network (AGDEN) to coordinate teams of researchers in Benin, Uganda, and South Africa to collaboratively develop the diagnostic tool, and then implement it by conducting a review of key documentation and to interview officials within the government wide monitoring and evaluation systems as well as the national gender machinery in each country.

Results: The study found that the gender responsiveness of M&E systems across all three systems was unequal, but more importantly, it is important to do more work on how M&E and gender are conceptualised, to ensure this can be studied in a more meaningful way. To strengthen national monitoring and evaluation systems, gender responsiveness and equity must serve as a foundation for growth. However, intersection M&E with gender is complex, and riddled with gaps in capacity, conceptual differences, and challenges bringing together disparate and complex systems.

Conclusion: A stronger understanding of the linkages between M&E and gender is an important starting place for bringing them together holistically.

Introduction

The Africa Gender and Development Evaluator's Network (AGDEN) was commissioned by the Centre for Learning on Evaluation and Results in Anglophone Africa (CLEAR AA) to conduct a diagnostic study to assess the Gender Responsiveness of the National Monitoring and Evaluation Systems (NMES) and Policies (NMEP) of South Africa, Uganda and Benin. This exercise is part of the multi-country peer-learning programme Twende Mbele, which aims to strengthen the NMES of the three countries. This assessment explored how the three countries can strengthen their efforts in making gender equity and equality integral to the NMES and NMEP.

While the assessment found certain strengths and weaknesses in the gender responsiveness of the NMES, the aim of this article is not to look at the findings of the study, which have been discussed elsewhere (Houinsa & Etta 2016; Jansen van Rensburg 2016a, 2016b; Wokadala & Sibanda 2016). These findings have important implications for planning capacity building interventions, informing tools to use within the national M&E systems and building systematic linkages to the national gender machinery in each country. However, this article focuses on the lessons learned through the process of carrying out the study. By grappling with some of the conceptual issues in tools development and engaging with stakeholders who are working either in crosscutting

monitoring and evaluation or gender initiatives, the study uncovered important lessons for how both are conceptualised, and what the real constraints are to bringing together two fields filled with intersections. Understanding the way gender and monitoring and evaluation intersect is an important first step in understanding how gender responsive monitoring and evaluation capacity can be strengthened. Without this foundation, it is possible that evaluation capacity development initiatives will fail to correctly identify areas of gender blindness or focus capacity development initiatives on areas that fail to intersect with gender.

Problem statement

Gender responsiveness and gender transformative evaluations are currently receiving more attention in light of the imperative to measure the Sustainable Development Goals (SDGs) through a gendered lens. Various documents are being developed and task groups of different organisations are actively raising awareness and promoting including gender in all aspects of evaluations, from commissioning through better structured terms of reference (TOR) to ensuring recommendations and implementation plans are gender inclusive and specific (Bamberger, Segone & Tateossian 2016; UN Women 2016a, 2016b, 2016c).

At the same time, there is a growing interest in national monitoring and evaluation systems in Africa. While monitoring and evaluation functions have long existed within the public sector, the experiences of Benin, South Africa, Uganda and others have demonstrated both the value and the need to systematise these functions and processes in order to integrate often disjunct systems of planning, budgeting, implementing and reporting. Globally, national systems of long-term planning are being aligned with medium-term planning and reporting processes.

In spite of these simultaneous, cross-sectoral trends, at the time of this study there was no widely accepted tool to specifically measure the gender responsiveness of monitoring and evaluation systems, let alone of national monitoring and evaluation systems in the African context. The gender responsiveness diagnostic study entailed developing a tool and method that could capture information that could also act as a baseline. Many valuable lessons were learned through this process and this article reflects on the key messages that need to be shared from the experience.

Evaluators play an important role as knowledge brokers and in enabling evidence to inform practice and policies. While practitioners themselves may be gender sensitive, when they are working within gender-blind systems, it can be difficult for a specific technical skill, like gender responsive evaluation, to transform otherwise gender-blind monitoring and evaluation systems. Similarly, evaluators should take more responsibility to reflect on and discuss the methodological aspects of their own studies (Rogers & Hummelbrunner 2011; Snowden 2002; Van Hemelrijck 2015, 2016; White 2009). This article aims to reflect on the tool, process and practical difficulty in investigating and measuring the intersection of gender responsiveness and monitoring and evaluation.

Development of gender responsiveness assessment tool

Foundational to the understanding of the capacity needs of developing gender responsive systems is a description of the process and content of the tool used in the study. AGDEN was engaged in conducting gender responsive assessments of national monitoring and evaluation policies and systems in three African countries, namely Benin, Uganda and South Africa. The central contribution of this exercise and a critical innovation by AGDEN was the development of a gender diagnostic tool for national policy environments and systems.

The tool is based on an ecological systems conceptual framework with different levels. The levels include macro level (national and African regional level), meso level (including government department and organisation) and micro level (consisting of project and evaluation levels). An ecological systems framework recognises the relationships between different systems and the influence of different systems and subsystems. The model of Bronfenbrenner (1979) further includes an individual level and an exo level that were not deemed of interest for the scope of this study. However, the systems approach enables one to examine and understand the influences in a structured manner from a wider cultural level to the project level. It was useful in framing items in the tool across systems and specifically in the multi-country context.

The aim was to develop a generic quantitative tool that can be used as a dashboard to capture the current gender responsiveness of each country's national evaluation system and compare the strengths and gaps across the three countries. The comparison was a means to share information between the three countries in the Twende Mbele[1] programme and enable peer learning by the participating countries. The ultimate goal was to inform collaborative capacity development that will be planned in light of the strengths and weaknesses in each country.

The alternative to this would have been to use a more exploratory approach. This would have had the advantage of accommodating the available time and stakeholders. However, it also would have meant that many of the challenges of comparability and variances in administrative and institutional context would have been amplified. Even with a very structured approach, it was difficult reaching consensus on what exactly the scope of measurement was and how different aspects of gender would be captured.

1.Twende Mbele is a peer learning initiative to strengthen monitoring and evaluation systems in African governments by stimulating demand, strengthening learning, increasing sharing, developing tools collaboratively and building communities of practice. The programme builds on a long-standing collaborative relationship between Benin, Uganda and South Africa.

Dimensions and criteria of the tool

Three dimensions were included to cover the most important elements of national M&E systems. The dimensions were: national M&E policies, systems (which included the M&E framework and institutional arrangements) and advocacy for implementation and practical application of M&E practices. There was a deliberate decision taken to include national M&E systems; it was important to include monitoring functions, since it is this monitoring that informs evaluation capacity and practice. An assessment of the national evaluations conducted in each country was not included in the study, since it required a different method and tool. Additionally, a decision was taken not to measure practice and capacity. While this would be an important next step, it would be a significant piece of research outside the scope of the diagnostic study.

In order to investigate if gender was included in the various dimensions, specific criteria were identified as minimum requirements. The development of the criteria started with a distinction between gender-negative or gender-blind, gender-sensitive or gender-responsive and gender transformative concepts (UNDP 2014). The first was deemed too restrictive and the last an outcome or function of the system rather than a characteristic under review in the current study. Initially, a more comprehensive list of criteria was compiled from existing tools from AGDEN and other lead organisations (ACDI/VOCA 2012; AGDEN 2011; UNAIDS 2014; UNEG 2011; UN Women 2015). This list included equity, equality, access to and control of resources and benefits, intentionality, non-discrimination, inclusion and participation of men and women, fair power relationships and empowerment, transparency, accountability, utility – use of information on results to benefit women and men (how does gender information of evaluations get incorporated in policies?) – sustainability (how sustainable is the involvement, inclusion, participation and short and longer term benefits to women and men?), dissemination (will or are the resourcing, planning, progress and results shared with women and men in ways that ensure knowledge sharing and generation of appropriate decision-making?) and alignment to SDGs (including Goal 5 and mainstreaming gender across all 17 goals).

An iterative process was followed to determine the final criteria that were included (i.e. a discussion meeting between the project consultants, input from other experts and review of more literature). Some of the criteria were merged, such as gender equality, equity and non-discrimination. This included combining criteria that are either related or sequential to a central characteristic of the system. The final criteria included the following:

- gender equality
- participation
- decision-making
- gender budgeting
- evaluability, review and revision
- sustainability

Each of the criteria was investigated through a set of questions that had to be answered from the documents under review and by the stakeholders who were interviewed. The following section provides a short description of each of the criteria to gain an understanding of what items were included in the tool.

Gender equality refers to the equal rights, responsibilities and opportunities of women and men, girls and boys. It implies that the interests, needs and priorities of both women and men are taken into consideration, recognising the diversity of different groups of women and men and ensuring the same enjoyment of opportunity, privilege and reward between men and women.

Gender budgeting means incorporating a gender perspective at all levels of the budgetary process and restructuring revenues and expenditures in order to promote gender equality.

Participation refers to different mechanisms for both men and women to express opinions and exert influence regarding issues that affect them. Participation includes two aspects. The first is a gender balance between men and women involved in different parts of the system. The second is that there is a gender expert, or a person knowledgeable about gender issues, as a part of all teams.

Decision-making examines who is empowered to make decisions in general and decisions related to gender and implementation of the national M&E policies. It examines the different power relations at play in the process.

Evaluability, review and revision: public policies are often made for specific time periods and it is considered good practice to review, assess, evaluate and revise them from time to time. This criterion examines if and how the national evaluation policy is reviewed and assessed from a gender equality perspective.

Sustainability is concerned with measuring whether the benefits of an activity are likely to continue after donor or project funding has been withdrawn. Sustainability measures the ability of the system to sustain change and to ensure that gender responses will continue to be developed and maintained. Adequate gender budgets are important for implementation of gender mainstreaming efforts and form an important aspect of financial sustainability. Financial sustainability however depends on funding beyond gender budgets.

Gender diagnostic matrix

From the dimensions and criteria, a gender diagnostic matrix (GDM) was developed. The GDM is an Excel-based instrument of 57 items which interrogate two levels (representing the policy and system levels, with the advocacy dimension embedded in both levels). The 57 items, which target seven equity and gender-responsive criteria, are accompanied by

performance scales for each item. The GDM is being revised for future use.

Conducting the gender-responsiveness diagnostic

After development of the tool was completed, three local teams conducted the gender responsiveness assessments in the three countries. Evaluation practitioners from Benin and Uganda were recruited by Twende Mbele and AGDEN conducted the assessment in South Africa. AGDEN provided oversight of the data collection, guided the analysis and report writing in each country and coordinated the results into a summary report (Houinsa & Etta 2016; Jansen van Rensburg 2016a, 2016b; Wokadala & Sibanda 2016).

The main sources of information for the assessment of the policy dimension were the M&E policies, as well as other policies and guidelines related to the M&E function across government. To fully understand the organisational structure and the development of the M&E functions and policies, a review of documents was performed. The document reviews aimed not only to assess the inclusion of the gender criteria in national policies, but also to understand the complex systems and their evolution. For example, in South Africa more than 80 documents were reviewed to gain an understanding of the development of the system.

Stakeholder interviews were then conducted to further contextualise the documentation. Individuals approached included representatives from government departments responsible for various elements of the national (government-wide) M&E functions. Departments responsible for gender mainstreaming (e.g. Department of Women), national gender machinery, constitutional bodies, civil society and other implementers were also included.

Positioning the diagnostic study

One central aim of the gender diagnostic was to maintain some kind of comparability across country contexts and, where possible, across existing instruments and measures. Twende Mbele is a collaborative initiative that looks at the joint development of tools and approaches among member countries as a particularly innovative approach to capacity building. However, for this collaboration to be effective there needs to be sufficient understanding of both the context in all countries and the common intervention. The gender diagnostic tool therefore had to be able to capture areas of similarity and difference across three different settings.

The countries have several key differences, which will be explored in more detail later in the article, and which are presented in detail in papers about the content of the diagnostic, cited above. Broadly, however, their differences included logistical and institutional levels. Logistical differences included differences in language, which posed a considerable challenge for coordinating a multi-country team with tight timelines. More significantly, though, it also posed a challenge in defining concepts around both gender and M&E, and trying to promote some sort of consistency in the application of different concepts. Institutional differences are built into the monitoring and evaluation systems and policies themselves. While each country had a designated body that was responsible for leading evaluations, its institutional linkages to long-term strategic planning mechanisms and performance monitoring functions varied widely. With each country having its own defined outcomes around gender, as well as institutions to promote gender equality, a key component of capacity building and building gender responsiveness at an institutional level lies in understanding these institutional mechanisms and their linkages to the M&E system.

The activities, coordination and influence of the national gender machinery varied substantially between the countries. South Africa has progressive legislation and policies on gender (e.g. policies linked to employment equity, gender-based violence, etc.). Unfortunately, the mandate of the national gender machinery (NGM) in South Africa does not seem clear to all the stakeholders. There is a fully functional and mandated NGM in Uganda that advocates for gender responsiveness in M&E. In Benin the NGM is coordinated by two different bodies.

These contextual differences between the countries had a major influence on the use of the tool and the interpretation of results. For example, the potential to create institutional linkages between the organisation responsible for coordinating the evaluation function and the NGM depends significantly on any interlinkages between mandate and institutional positioning. For example, there are shared responsibilities across the departments of Women, Social Development and Planning, Monitoring and Evaluating for measuring certain outcomes that include a gender dimension. A change is currently underway in the region to shift the focus of capacity development in the regional space from exclusively individual skills building and training to institutional strengthening. As this happens, having stronger tools for understanding these institutions and their roles is an important first step for capacity building effectiveness.

Methodological reflections from a regional perspective are important to ensure that the lessons learned are not only applicable to the three countries in this study, but that further studies and the application of the tool and method can be contextualised in other countries on the continent.

Methodological reflection

This reflection exercise took into account feedback from the evaluators working in the three countries and comments by the governmental agencies involved in the study. Information and viewpoints expressed by participants at the Africa Evidence Network and the European Evaluation Society conferences, where the initial results were presented, also informed this article.

This article reflects on three different aspects of the gender diagnostic study. The first are the methodological and

practical considerations, such as scoping and preparation of the project, tools and data collection process. This will uncover some of the technical elements of gender responsiveness that will need to be included in evaluation capacity development initiatives. The second looks at the contextual and institutional differences between the countries. This will look at the way capacity building interventions will need to address institution building to strengthen gender responsiveness of M&E systems. The third and main reflection looks at the points of integration and fragmentation of the gender and M&E systems themselves. This suggests both reconceptualising how gender responsiveness is conceived in M&E, as well as uncovering specific lessons about how cross-cutting fields can fail to intersect. This will help to identify the areas in which capacity building initiatives will gain traction and synergy, versus areas where it may fall flat.

Reflection on practical aspects

Decisions on the scope and focus of the diagnostic approach

Sufficient time and resources are needed to develop an appropriate tool to be used in different contexts and to investigate the complex interplay between two systems (gender and M&E). It was clear from this study that time needs to be allocated to ensure piloting of the tool in different contexts and training of those involved in its application. In multi-country studies the use of mixed teams of evaluators is necessary, with in-country members and members of a central team (who are central to developing the tool and method, as well as ensuring coherence in analysis). However, this is a costly approach in the African situation with resources needed for travelling to a central training venue, communication and translation.

Another challenge was the decision-making process on the scope of the project. What could easily have been a large, multi-year research initiative was limited by both a practical demand to use the results and logistical considerations of contracting timelines and budgets. Decisions had to be made on how to focus the study. One criterion for limiting the scope was to remove aspects that could form part of follow-up studies. For instance, the current study had to restrict the assessment of the gender responsiveness to the system (NMES) and the policies (NMEP) and exclude aspects that are part of the system such as the engendering of evaluations themselves. The assessment of gender responsiveness of the national evaluations completed in each country will form part of another study and was easier to limit. Other limitations and restrictions involved the decision not to look at the actual implementation of interventions and beneficiaries, which would have given valuable insight about current capacity, gaps, and requirements. Ideally, additional elements of process and implementation, and stakeholder results, would have added a valuable dimension to the diagnostic study, but was not possible due to time and resource constraints.

Another decision was to limit the scope of the study to exclude elements of performance management and measurement. Certain elements of the national M&E systems, particularly the monitoring functions, were closely linked to performance management. However, the overlap between the monitoring aspects of the system and the performance measurement aspects raised some issues during the analysis. For example, in South Africa the performance management aspects form an integral part of the system and aspects that were not included in the tool seemed to indicate a lack of gender responsiveness in monitoring, such as economic empowerment aspects and affirmative action. The boundaries between these elements of the M&E system and the measurement of the gender responsiveness demonstrate already that there is a range of conceptual approaches to M&E and that there is not necessarily a consensus among key role players about what is included or what is prioritised.

Additionally, conceptualising the focus and scope of the diagnostic tool was further clouded in that many respondents across the three countries mistook a gender responsive M&E system to mean evaluating the effectiveness of gender projects and interventions (such as measuring the impact of M&E, or even the evaluation of specific interventions on affirmative action on women, or evaluating gender projects such as gender-based violence) for gender responsiveness of the M&E system. In other words, in applying the tool, the research team had to clearly indicate that it was the gender responsiveness of the system itself being measured and not a tool to look at how gender activities and projects should be monitored and evaluated by the national system. This confusion demonstrates the complexity of the intersection of the two concepts and how gaps in capacity range and amplify from individual skills to institutional integration.

Certain assumptions about the nature of the two systems led to further decisions about the scope and focus of the project. Deciding on a geographic limitation of the diagnostic to the national components of the M&E system was a challenge. This was relevant to both gender and M&E structures. For example, in South Africa both the systems include a decentralised strategy for mainstreaming both functions. This means provincial and local government M&E departments and gender focal persons are considered part of the national systems. These components are important to understand mainstreaming efforts. At the same time, this is something that varied considerably by national context and the public administration system in each country. Local M&E departments, grass roots organisations and civil society play an important part in the gender responsiveness of the system at a district or municipal level in all three countries. The diagnostic began with an assumption that gender mainstreaming efforts across the national M&E system would be largely top down, since this has been true of many national policy initiatives on the African continent. However, following the process of implementing the diagnostic, this assumption did not necessarily hold. Capacity development strategies taking a bottom-up approach also have merit.

Without additional focus on the implementation aspects of the national M&E system, it is hard to draw conclusions

around this, but if the tool is to be revised and reapplied, it is an important area to revisit. It also implies that in capacity building interventions, it is not always a straightforward process to determine the capacity within the system, which can be limited to specific components, versus the capacity of the system, which is very much informed by the environment in which the system operates and the ecosystem of stakeholders.

Conceptual challenges in developing the diagnostic tool and its application

Apart from the real limitations of budget and time, which influenced the scope of the diagnostic tool, additional challenges around the conceptualisation and content of the study shaped the development of the tool. These are considerations that need special attention in similar studies regarding both diagnostic tool design and conceptualisations of gender responsiveness or national M&E systems.

Defining both the concepts of gender and M&E across all three countries was difficult. While there were legal definitions in place in the countries, it wasn't evident that these were accurate reflections of the way the concepts were used. At the same time, without a more exploratory, qualitative research design, and with a limitation on the ability to look at actual implementation issues, there was limited scope to explore these variations of definitions and their application.

One challenge in the initial stages of the study was the varied use of both gender and M&E terminology. Similar concepts were often reflected by different terms in different countries, and even within different components of the M&E systems in the same country. Alternatively, the same terminology, particularly around gender, was often used to represent quite different concepts. In some cases, 'gender equality' was a broad, inclusive concept that included gender diversity, sexual orientation and identity, while, more often, 'gender equality' referred specifically to the inclusion of women in certain spaces or processes.

Political differences in the way gender is conceptualised, both socially and politically, across all three countries, complicated the inclusion of measuring the responsiveness of M&E systems to gender. For the purpose of this particular study, the team adopted a narrower approach to defining gender. This was both due to a deliberate assumption that the development of gender responsiveness in a system would most probably firstly be visible with women before being responsive to other gender groups. However, it still remains the case that the contribution of the study to understand the level of gender responsiveness (taking a more narrow definition of gender) could apply to other groups, and would aid in development of a system that is more broadly gender responsive in all aspects and could be more inclusive of other excluded groups. This issue particularly highlights the necessity of building gender into evaluation capacity development interventions. Given the significant blind spots to gender in these systems, it is likely other aspects of equality are being equally overlooked. This is critical to understand, for planning systems to run well.

A further challenge to implementing the study was around existing stereotypes about both gender and M&E. The resistance to gender is reasonably well understood and documented in feminist evaluation theories (Diamond et al. 1990; Hall & Bucholtz 2012; McNay 2013). As an alternative explanation of the way power is articulated, this is not a surprise. However, M&E is also subject to a range of stereotypes. While some of the attitudes and perceptions about M&E have been studied, scholarship on this comes from a nearly exclusively northern context. While certain perceptions may be the same (such as intimidation about a highly technical subject or resistance to compliance mechanisms), there is a need for more research about attitudes towards M&E in an African context, especially as evaluation moves from a donor imposition to an imperative of governance. The close relationship between the field of M&E and the donor community is central to this, and under-studied. While some research on this will be forthcoming through initiatives by CLEAR, anecdotal avoidance of M&E as a function across government departments is evident.

Due in part to a lack of consensus about the purpose and role of M&E, different departments have articulated M&E functions differently (Cloete 2009; Estrella & Gaventa 1998; Mayne & Zapico-Goñi 2007; Sanderson 2001; Seasons 2003). While all three countries have a coordinating department within the office of the presidency or the prime minister, this department's relationship to the M&E function within departments varies. While this is of general interest for understanding the system, pragmatically, it meant that entry into departments was inconsistent and often challenging. Had there been a longer process of obtaining buy-in for the study and its goals and objectives, this may have been less of a limitation, but the fact that this hurdle was encountered across all three country contexts demonstrates a strong discomfort with the unclear mandate for both gender and M&E functions within departments. This was an important methodological consideration for the diagnostic, but is equally important in considering the target and structure of capacity building initiatives which follow. These are specific to each context. Understanding where leadership and ownership for evaluation capacity is within departments and coordinating mechanisms are important in determining how best to target interventions.

Comparisons in context

As described above, part of the analysis and interpretation of the results was a comparison between the M&E systems of the three countries in the form of a dashboard highlighting the different elements of gender responsiveness. This comparability was important to allow all three countries to plan certain common, follow-up capacity building interventions. The intention of this dashboard approach was

to briefly highlight the relative strengths and weaknesses within each country, to enable cross-learning and highlight some areas of good practice. Unfortunately, these results are often seen as punitive, which was not intended in the research design. Diagnostic studies are more geared towards learning than accountability, but even when situated firmly in the learning space, there are sensitivities around weaknesses. When the tool is revisited for future application, this may be taken into account, to look for an approach that highlights strengths and weaknesses, but also highlights similarities and differences in a non-hierarchical way.

An even more significant challenge to comparability across countries is that it has been difficult to contextualise each country's national M&E system. To understand the gender responsiveness of each system, it was important to understand how all three developed. However, assessment across different levels of development of systems in the countries was a difficult task. As a follow-up step, Twende Mbele is developing a diagnostic tool of the national evaluation systems themselves, to allow a more structured analysis of various aspects of the systems. When the gender diagnostic itself was carried out, it was difficult to compare elements of gender responsiveness of three systems that were located in different public administration contexts and had different approaches and different mandates. This was equally true of the national gender machinery. Without some sort of mapping process of the systems and their respective purposes, components, capacities and functions, it was very difficult to develop a tool that could have any sort of comparable role. The tool to be used in the future needs to be flexible enough to allow contextual differences to be captured to enable comparison.

Finally, there was a lack of contextual information across all three countries that would have allowed for stronger comparability. As mentioned previously, there is little consensus even across different components within one country's system about the concepts, mandates and purpose of the M&E function. To add to this, the M&E systems in all three countries are situated differently politically. These are all before considering the different understandings of gender-related concepts by different stakeholders within these systems. While uncovering this varied landscape was part of the richness of the study, a much more nuanced comparison will only be possible when this diversity of context is better researched.

Intersecting gender responsiveness with monitoring and evaluation

The mainstreaming of gender and M&E in national systems pose or face similar challenges. They both include a wider range of technical concepts that are not understood in the same way across departments and other stakeholders. The mandates for strengthening government action on both are complex and not always clear. Both areas are recognised as priorities in the current global political climate, which has prompted a lot of activity, but still without common definitions or widespread capacity. Furthermore, it is not clear how to define, let alone measure, capacity in these complex institutional arrangements.

In the absence of common conceptualisations of either gender responsive government, or national M&E systems, it is a very challenging task to develop a tool that will gauge either one, let alone in a comparative context. In spite of the challenges, the very lack of a common understanding and definition made the process of carrying out the diagnostic study even more important. It is precisely through the development of a tool that sought to measure the gender responsiveness of each country's NMES that the research team was able to identify areas of contestation, and it is through repeated efforts at defining and redefining the scope and focus of these systems that concepts will begin to be used consistently and consensus will be built about various definitions. This is one case where the research process has played a very active role in defining the problem. Similarly, further research has the potential to be an agent in a solution. The collaborative research itself will be one step in capacity building.

Some of the issues mentioned in the previous reflections are also relevant for this description, including the tool development and the application of the tool in contexts that vary in development and in defining central concepts such as gender. The differences between the three countries regarding certain central aspects of the study (such as defining gender and the level of development and structure of the M&E system) played a significant role in the issue of intersecting gender and M&E.

There is also an intersection in roles and responsibilities to make the M&E system more gender responsive. However, it is importance to delineate the differences (and the mandate) to mainstream gender in the government (generally the role of a department or ministry such as the Department of Women) and that of making sure the system that is monitoring and evaluating all government programmes and activities looks at the application of gender equity and equality and ensures recommendations and implementation plans include women.

The complexity this diagnostic study faced included all three countries having different political, social and administrative contexts. Within these differing contexts, both gender and M&E were facing contestations around conceptualisation, institutionalisation, bureaucratic location and the development of systems. The diagnostic was a first step in making some of these differences explicit, and understandable, in a way that will allow us to start the process of looking at the current capacity across different contexts.

Lessons learned on integrating gender into M&E capacity building

The lessons learned from the study and process of reflecting on its implementation highlight a few considerations for

designing and implementing capacity building to integrate gender in M&E systems on a national level.

Firstly, there are aspects to consider for building the capacity of the national M&E systems. Any capacity building exercise will need thorough assessment of the extent of intersection between the gender and M&E functions to ensure targeting of all relevant persons. Basic sessions on clarification of key concepts should form the basis of all events. The format of the capacity development activities can include online platforms (including self-paced electronic learning and webinars) which would be very useful for developing insights into the key concepts. Face-to-face engagements with system personnel are critical for ensuring integration of gender in practical aspects and implementation. Clear guidelines (including checklists) and reference materials need to be developed on topics such as how to include gender in terms of reference and requests for studies. It is further important to include stakeholders of all levels in the capacity building strategy. Identification of a sponsor is imperative. The capacity building should not only focus on gender focal persons or key M&E officers, but all departments and institutions involved in the M&E system. The complexity of the intersection of gender and M&E systems, as described in this document, implies that the capacity building exercises need to be extensive and flexible to accommodate not only different levels of staff, but also different perceptions and interest. Using the resistance to discussing and engaging about gender can be useful as a training strategy.

Secondly, there were valuable lessons learned in regard to building capacity for the assessment of gender responsiveness. This includes correct scoping of the assessment as illustrated in this reflection. The allocation of sufficient human and other resources is critical, especially when conducting multi-country assessments. The development of a context-appropriate tool will need the input of gender experts as well as persons who understand the nuances of the local context. It has to be an iterative process that takes place on a local level. Stakeholder buy-in needs to be done early in the assessment to ensure a comprehensive understanding of what is clearly a complex intersectional system.

Conclusion

The actual findings of the diagnostic study on the gender responsiveness of the national evaluation systems of Benin, Uganda and South Africa had a number of limitations, due to the range of factors discussed above. Some of them were practical, ranging from issues of language and coordination. However, the most important factors were conceptual. Carrying out the study exposed the extent to which both M&E and gender concepts and mandates are not shared among stakeholders who play active roles in working on them. This unexpected finding made the implementation of the study all the more important, since the process of developing a diagnostic tool itself served to identify some of the key areas of contestation and confusion. Continuing with this research agenda will make an important contribution to the creation of common definitions and applications of different concepts among stakeholders within the NMES and national gender machinery. It will also allow all three governments to begin to understand the existing capacity for gender responsive national evaluation systems and help to identify some of the most effective interventions for strengthening this capacity.

In conclusion, the exercise has been important in the SDG climate, which prioritises both gender and M&E, but is still grappling with some of the complexities of measuring transformative change regarding gender responsiveness. It has provided valuable insight into the development, adjustment and application of a gender responsiveness measurement tool. The development of the gender diagnostic tool is an iterative, ongoing process that gained valuable insight from assessment in the three countries.

Acknowledgements

The support in the gender diagnostic project by James Wokadala, David Houinsa, Florence Etta and Adeline Sibande is appreciated. The authors acknowledge the financial support by Twende Mbele and DFID.

Competing interests

The authors declare that they have no financial or personal relationship that may have inappropriately influenced them in writing this article.

Authors' contributions

M.J.v.R. was the project leader and led the development and application of the tool during the diagnostic study. She was therefore also able to lead the analysis on how the tool was developed and implemented. C.B.M. was the co-author and, as the programme manager of Twende Mbele, was able to speak to how the issues of context and comparability are important for programme design. Both authors contributed to the documentation of the manuscript.

References

ACDI/VOCA, 2012, *Gender analyses, assessments, and audits brief*, ACDI/VOCA, Washington, DC.

AGDEN, 2011, *Human rights and gender responsive monitoring and evaluation: AGDEN toolkit for practitioners*, AGDEN, Nairobi.

Bamberger, M., Segone, M. & Tateossian, F., 2016, (Draft version), *Evaluating the Sustainable Development Goals within a 'No one left behind' lens through equity-focused and gender-responsive evaluations*, UN Women, New York.

Bronfenbrenner, U., 1979, *The ecology of human development: Experiments by nature and design*, Harvard University Press, Cambridge, MA.

Cloete, F., 2009, 'Evidence-based policy analysis in South Africa: Critical assessment of the emerging government-wide monitoring and evaluation system', *Journal of Public Administration* 44(2), 293–311.

Diamond, I., Quinby, L., Benhabib, S. & Cornell, D., 1990, 'Feminism and Foucault: Reflections on resistance', *Hypatia* 5(3), 118–124.

Estrella, M. & Gaventa, J., 1998, *Who counts reality? Participatory monitoring and evaluation: A literature review*, Institute of Development Studies, Brighton.

Hall, K. & Bucholtz, M. (eds.), 2012, *Gender articulated: Language and the socially constructed self*, Routledge, United Kingdom.

Houinsa, D.G. & Etta, F., 2016, *Rapport De L'analyse Diagnostique De La Prise En Compte De Genre Dans Les Systemes Nationaux De Suivi Et D'evaluation: Cas Du Benin*, unpublished report.

Jansen van Rensburg, M.S., 2016a, *Gender responsiveness diagnostic report of national monitoring and evaluation systems in South Africa, Uganda and Benin.*

Jansen van Rensburg, M.S., 2016b, *Gender responsiveness diagnostic report of South Africa government-wide monitoring and evaluation system*, unpublished report.

Joint United Nations Programme on HIV/AIDS (UNAIDS), 2014, *UNAIDS Gender assessment: Towards a gender-transformative HIV response*, UNAIDS, Geneva, Switzerland.

Mayne, J. & Zapico-Goñi, E. (eds.), 2007, *Monitoring performance in the public sector: Future directions from international experience*, Transaction Publishers, New Jersey.

McNay, L., 2013, *Gender and agency: Reconfiguring the subject in feminist and social theory*, John Wiley & Sons, New Jersey.

Rogers, P. & Hummelbrunner, R., 2011, 'Methodological challenges in using programme theory to evaluate pro-poor and equity-focused programmes', in I. Sanderson, 2001, (ed.), 'Performance management, evaluation and learning in 'modern' local government', *Public Administration* 79(2), 297–313.

Sanderson, I., 2001, Performance management, evaluation and learning in 'modern' local government, *Public Administration* 79(2), 297–313.

Seasons, M., 2003, 'Monitoring and evaluation in municipal planning: Considering the realities', *Journal of the American Planning Association* 69(4), 430–440. https://doi.org/10.1080/01944360308976329

Snowden, D., 2002, 'Complex acts of knowing: Paradox and descriptive self-awareness', *Journal of Knowledge Management* 6(2), 100–111. https://doi.org/10.1108/13673270210424639

UNDP, 2014, *On course: Mainstreaming gender into national HIV strategies and plans – A roadmap*, viewed 11 August 2016, from http://livelifeslowly.net/genderinghiv/

UNEG, 2011, *Evaluation norms and standards + integrating human rights and gender equality in evaluations*, viewed 8 August 2016, from http://www.unevaluation.org

UN Women (Independent Evaluation Office), 2015, *How to manage gender-responsive evaluation: Evaluation handbook*, UN Women, Nairobi.

UN Women, 2016a, *UN women training centre: Compendium of good practices in training for gender equality*, UN Women, Nairobi.

UN Women, 2016b, *Building gender responsive and transformative national evaluation capacity: Training manual*, UN Women, Nairobi.

UN Women, 2016c, *Compendium of good practices in training for gender equality: Concept note*, UN Women, Nairobi.

Van Hemelrijck, A., 2015, *Improved learning initiative for the design of a Participatory Impact Assessment & Learning Approach (PIALA): Methodological reflections following the second PIALA pilot in Ghana*, IFAD.

Van Hemelrijck, A., 2016, *Methodological strengths and limitations of the SenseMaker® pilot in AMD*, IFAD, Rome, Italy.

White, H., 2009, *Some reflections on current debates in impact evaluation* (Working paper No. 2009-1). International Initiative for Impact Evaluation (3ie), New Delhi.

Wokadala, J. & Sibanda, A., 2016, *Diagnostic of the gender responsiveness of national monitoring and evaluation system in Uganda*, unpublished report.