# EVALUATION LANDSCAPE *in Africa*

## *Context, Methods and Capacity*

**Caitlin Blaser Mapitsa**
**Precious Tirivanhu**
**Nedson Pophiwa**

*Editors*

SUN PRESS

*Evaluation Landscape in Africa – Context, Methods and Capacity*

Published by AFRICAN SUN MeDIA under the SUN PReSS imprint



This publication was subjected to an independent double-blind peer evaluation by the publisher.



First edition 2019

ISBN 978-1-928480-18-1
ISBN 978-1-928480-19-8 (e-book)
https://doi.org/10.18820/9781928480198

Set in Avenir Light 10/15
Cover design, typesetting and production by AFRICAN SUN MeDIA

SUN PReSS is a licensed imprint of AFRICAN SUN MeDIA. Scholarly, professional and reference works are published under this imprint in print and electronic formats.

This publication can be ordered directly from:
www.sun-e-shop.co.za
www.africansunmedia.store.it.si (e-books)
www.africansunmedia.co.za

# CONTENTS


*The editors would like to thank CLEAR-AA for their support in compiling this volume and CREST for their work on the database. We appreciate the efforts of all contributing authors, as well as the many colleagues at Wits University and beyond, who supported the research.*

# LIST OF ACRONYMS

| | |
|---|---|
| 3ie | International Initiative for Impact Evaluation |
| ACBF | African Capacity Building Foundation |
| AEA | American Evaluation Association |
| AEGs | African Evaluation Guidelines |
| AfDB | African Development Bank |
| AfrEA | African Evaluation Association |
| AfrED | African Evaluation Database |
| ALNAP | Active Learning Network for Accountability and Performance in Humanitarian Action |
| ANZEA | Aotearoa New Zealand Evaluation Association |
| CHESS | Checklist for Evaluation-Specific Standards |
| CIMES | County Integrated Monitoring and Evaluation System |
| CLEAR-AA | Centre for Learning and Results Anglophone Africa |
| CoMEC | County Monitoring and Evaluation Committee |
| CPD | continuous professional development |
| CREST | Centre for Research on Evaluation, Science and Technology |
| CSW | United Nations Commission on the Status of Women |
| DBSA | Development Bank of Southern Africa |
| DAC | Development Assistance Committee |
| DANIDA | Danish International Development Agency |
| DFID | Department for International Developmen |
| DHET | Department of Higher Education and Training |
| DPME | Department of Planning, Monitoring and Evaluation |
| ECB | evaluation capacity building |
| ECD | evaluation capacity development |
| EU | European Union |
| FGD | focus group discussion |
| G&E | gender and equity |
| GIZ | *Deutsche Gesellschaft für Internationale Zusammenarbeit* |
| GoK | Government of Kenya |
| HEI | higher education institution |
| HIV/AIDS | Human Immunodeficiency Virus/Acquired Immunodeficiency Syndrome |
| IDEAS | International Development Evaluation Association |
| IDI | in-depth interview |
| IMF | International Monetary Fund |

| | |
|---|---|
| IMR | infant mortality rate |
| KII | key informant interview |
| M&E | monitoring and evaluation |
| MAE | Made in Africa Evaluation |
| MBNP | Ministry of Budget and National Planning |
| MDPMED | Ministry of Devolution and Planning, Monitoring and Evaluation Department |
| MMR | maternal mortality rate |
| NES | National Evaluation System |
| NGO | non-governmental organisation |
| NMR | neonatal mortality rate |
| NORAD | Norwegian Agency for Development |
| OECD | Organisation for Economic Cooperation and Development |
| P2P | peer to peer |
| PEPFAR | The US President's Emergency Plan for AIDS Relief |
| PIK | Peace Initiative Kenya |
| PRA | participatory rapid appraisal |
| QF | Quality Framework |
| READ II | Reading for Ethiopia's Achievement Developed Institutional Improvement |
| RCTs | randomised control trials |
| SAMEA | South African Monitoring and Evaluation Association |
| SDGs | Sustainable Development Goals |
| Sida | Swedish International Development Cooperation Agency |
| TOC | theory of change |
| TQF | total quality framework |
| UNAIDS | Joint United Nations Programme on HIV/AIDS |
| UNDG | United Nations Development Group |
| UNDP | United Nations Development Programme |
| UNEG | United Nations Evaluation Group |
| UNFPA | United Nations Population Fund |
| UNICEF | United Nations International Children's Emergency Fund |
| UNWomen | United Nations Entity for Gender Equality and the Empowerment of Women |
| USAID | United States Agency for International Development |
| VOPE | voluntary organisations for professional evaluation |
| WB | World Bank |
| WFP | World Food Programme |
| WoS | Web of Science |

# INTRODUCTION

This book emerged from a need to better understand the trends and dynamics of evaluations on the continent. While the evaluation field has been growing in Africa over the last two decades, there is a dearth of empirical research available to allow scholars to better understand the relationship between evaluation practice and the political economy that underpins the evaluation agenda on the continent. This book is an effort to tease out some evidence on evaluation trends from a database on evaluations over a decade (2005–2015) across 12 countries within Anglophone Africa. This may help substantiate some themes that are beginning to characterise the discipline on the continent.

Evaluation systems are increasingly playing a key role in both global governance and the national government architecture. The Sustainable Development Goals (SDGs) call for country-led evaluation systems and are revealing the limited capacity of most African governments to track progress of the SDGs through their existing monitoring and evaluation (M&E) systems. This pressure is occurring alongside a growing call by African governments to strengthen key areas within this burgeoning discipline and practice. Central to this is strengthening of national systems of evaluation practice, enacting appropriate legislation around the practice of monitoring and evaluation, and building institutions to carry out monitoring and evaluation functions across the public sector. While this momentum is commendable, it also demands that evaluation scholars and practitioners support this growth by a provision of appropriate empirical data and analysis around the way these systems can be built to best meet the needs of the region.

## Poverty and development as a broad contextual issue for evaluation in Africa

Poverty reduction remains an elusive agenda in Africa within the framework of the SDGs that aim at eradicating global poverty by 2030. Although statistics by the World Bank suggest that the proportion of the African population in extreme poverty declined from 57% in 1990 to 43% in 2012, Africa's population has continuously expanded, increasing the number of people living in extreme poverty by more than 100 million (Beegle, Christiansen, Dabalen & Gaddis, 2016). In addition, climate change could account for close to

10,1 million more poor people by the middle of the century. Although average economic growth in Africa has been estimated at 4,5% per annum, assessing the impacts of such growth on well-being remains a key challenge. There is growing evidence that, unless distributive mechanisms are put in place to arrest the iniquitous dimensions to the current economic growth, this potential is unlikely to reduce poverty.

Understanding and addressing poverty is complicated by the fact that poverty statistics in the region are often limited and sometimes of poor quality. Such a scenario is partly due to the poor quality of data generated from evaluations (Devarajan, 2013). Lack of funding and low capacity are often cited as main drivers for the data gaps in Africa. Poverty estimates are based on data from a patchwork of household surveys that are conducted at irregular intervals and are sometimes incomparable and of questionable quality. In terms of capacity, the production of high-quality consumption surveys and statistics is technically complex, involving the mobilisation of financial and human resources on a large scale and requiring the establishment of robust quality-control mechanisms. Without accurate statistics and careful analysis, the extent of the problem cannot be gauged and policy directions on how or where to accelerate investments or measure the success of efforts to end extreme poverty, will be ill informed (UNICEF & World Bank, 2016).

Although the research on poverty has been ongoing for decades, in recent years focus has been on improvements to the quantity and quality of data, as well as advances in conceptualising and measuring poverty (see, for example, Cauthen & Fass, 2008; Pradhan & Ravallion, 2000). A significant shift is the transition from an exclusive focus on unidimensional consumption-poverty domain towards incorporation of non-income dimensions of well-being, such as literacy achievements, nutritional and health status and individual empowerment. In recent years, attention has been given to the measurement of poverty, with considerable attention towards blending both monetary and non-monetary dimensions of well-being (Christiaensen & Shorrocks, 2012:137). This has led to lengthy debates in the way in which various dimensions of well-being are best combined into a single multi-dimensional measure. A daunting question for the evaluation discourse in Africa is: to what extent has the design of development-related evaluations adjusted to these progressions in research? For the sector to be relevant in Africa, there is a need to maintain a

constant awareness of both the way evaluations measure development and the way that they contribute to development.

This book covers a wide array of insights from an examination of 2 635 evaluation reports across 12 countries over a ten-year period. While these may have a seemingly eclectic compilation of chapters that offer diverse insights into evaluation practice on the African continent, there are a few cross-cutting themes to the book.

## *Appropriate capacity*

Monitoring and evaluation as a discipline took root in Africa in conjunction with the rise of New Public Management (NPM), linking it inexorably to the role of donors in both public finance and social development. This enmeshed the monitoring and evaluation function very strongly with discussions on accountability for donor and tax expenditure, accountability and auditing. Systems of monitoring and evaluation often grew out of performance management of budgeting processes, emanating from offices of the auditor general and international donors (Mouton, Rabie, Cloete & De Coning, 2014:34).

This history has shaped the epistemological framework for what counts as knowledge in the evaluation sector. Central to this is a positivist paradigm that gives huge import to quantitative measurement as the best means to assess impact, with randomised control trials being the gold standard. This quantitative approach embraces a Cartesian ontological approach where statistical validity provides certainty in what is being measured, even though it may well miss what statistics fail to count.

These power relations have been further entrenched by a growing trend of "collaboration" in European and American practice of evaluation that brings Western evaluators to the African continent to do evaluations. In these institutional arrangements, local evaluators are part of these evaluation teams but are reduced to field collectors and excluded from the design of evaluations, let alone the analysis. The chapters in this book provide ample evidence of this trend in "collaboration". The advocacy behind the Made in Africa monitoring and evaluation (MAE) approach would call for knowledgeable local actors to help define the design for an evaluation or a monitoring framework to ensure they are culturally relevant.

Indigenous societies in Canada, Australia and New Zealand have for some time now been calling for a "decolonisation" of the curriculum and this has been picked up in the concept of M&E as an approach that is people centred and values and culturally relevant and indigenised evaluation processes and methodologies which are informed by African world views and paradigms (Wilson, 2008; Kovan, 2009; Chilisa, 2016). Central to this is a participatory approach where not only local evaluators are part of the design but where beneficiaries also take part to ensure their perspectives begin to influence the ontological framing of the evaluation (Bellagio Centre, 2012; Cloete, 2016). Despite an emergent trend in the region of understanding evaluation practice as part of a wider system, there are limits to trying to transform methods and approaches to evaluations when public institutions themselves are embedded in systems of planning and implementation that remain untransformed in identifying and valuing Eurocentric models of knowledge. The challenge is to foster greater appreciation amidst bureaucracies on the continent of African epistemologies, which could in turn create greater openness towards valuing indigenous methods.

### *Bureaucratic capacity for states to learn*

As national evaluation systems in the region grow, governments are becoming increasingly important role players in setting the evaluation agenda. This is an important step away from the nearly exclusively donor led landscape and opens a door for local ownership, coupled with more contextually appropriate systems of knowledge creation and participation. However, it also comes with cautions. Interrogating the way evaluation practice and evaluation systems currently work within African governments will help us better understand the nature of the problem. So far, approaches to evaluations within the public sector have largely been technocratic. This implies that the use of evaluations is at best to provide performance assessments for commissioners of evaluations (Chouinard, 2013:242) and, at worst, to perform strict accountability and compliance functions. Technocratic approaches to evaluation focus on creating monitoring systems, measuring outputs and providing M&E training, which are far from sufficient for successful M&E systems (MacKay, 2007). The demand-side issues around M&E are neglected, such as engaging with the incentives to conduct and to use M&E information as well as training that raises awareness of the uses and value of M&E which in the end can help strengthen cultures of learning (MacKay, 2007).

The dilemma of a growing technocratic creep that responds to the accountability drivers of evaluation at the expense of promoting the learning agenda through evidence emerging from evaluations is a risk to the sector. With reference to the civil service, in the efforts to institutionalise the practice of evaluations within the bureaucracy, the opportunities for innovation and risk taking in applying lessons learnt to other domains of programmatic work will be limited. This is because the check list to ensure evaluations get done rather than how they get done takes precedence. An example of this was seen in the Evaluation Technical Working Group (ETWG) for the National Evaluation System (NES) in South Africa. The ETWG had been driven by the Department of Planning, Monitoring and Evaluation (DPME) in partnership with heads of M&E Units in various line departments as well as heads of M&E in offices of the premier. The ETWG had historically helped to establish the NES in South Africa and played a significant role in mainstreaming the evaluation function across line-ministries and provinces, a process that had historically been voluntary as the DPME had originally chosen not to go the statutory route in establishing the NES. But after five years of practice, the findings from the first evaluation of the National Evaluation System emerged and Departmental officials observed the time-cycle it took to see the programmatic impact from the use of evaluation evidence. They then made a decision to legislate evaluation practice. This decision was driven by the motive to institutionalise the evaluation function. While noble in intent, the unintended consequences was to diminish the space for innovation and driving a learning agenda as it got subsumed by the compliance of how many evaluations were in the Annual Performance Plans, what their budgets were, what the consequences were in not meeting these targets, and so forth. This is just an example of the delicate balance required to try to ensure an evaluation function has the financial and human resources it needs to operate effectively within a bureaucracy while not being so prescriptive that the space for ingenuity and individual leadership is obliterated.

Accountability-driven technocratic approaches are in sharp contrast with approaches that are more sensitive and responsive to community needs, such as participatory or collaborative approaches to evaluation. As argued by Chouinard (2013:242), a technocratic approach to evaluation "may well fall short of capturing the range of local views, contextualised meanings and culturally relevant perspectives that are increasingly relevant today". Participatory approaches emphasise the need for participation of different stakeholders in

the M&E process, during planning, implementation, results, as well as further decision-making (Weber, 2008:18). While not discarding the accountability function of evaluations, Latib (2014:469) contends that:

> In appreciating that M&E is fundamentally about building accountability and inclusivity, efforts can be focused on strengthening the ability of those within the political space to engage with information. Contrary to the technocratic approach of training politicians and stakeholders on engaging with 'complex reports', inclusivity would require re-shaping reports so that they are understandable to those within the larger accountability space. This requires an active move towards simplifying reports generated.

Ensuring that inclusivity becomes a critical aspect of evaluations on the continent will lead to a shift from technocratic to more participatory approaches of evaluation. At present, "technocratic reports which, by virtue of their source, tend to close off the spaces for effective politics" (Latib, 2014:463).

### *Varied quality of evaluations*

As expected of an emergent field of study and practice, the issue of quality, standards and norms as well as levels of professionalisation will at best be varied. There are pockets of excellence in evaluation designs and methods that have been conducted on the continent in as much as there are other examples where room for improvement is needed. Considering that evaluations in Africa are still largely commissioned by international donor or development agencies that run or fund development programmes on the continent, issues pertaining to quality will always surface. Quality is usually assured within the framework of the donors, which potentially could disregard locally relevant approaches as sub-standard. In highlighting the sensitivities on who sets criteria for measuring quality, Mouton, Rabbie and De Coning (2014:58) remark:

> This is still a sensitive issue for many African evaluators, because perceptions have emerged in circles both in Africa and outside the continent that African evaluators have to improve their international competitiveness compared to their northern hemisphere counterparts, because their profession is relatively new and there is much room for improvement.

The authors, however, explain that this situation is changing fast and African evaluators are becoming increasingly competitive on an international level (Mouton et al., 2014:58).

There have been efforts to establish standards, competencies and guidelines for evaluations in Africa. One such example is the evaluation guidelines developed by AfrEA's in 2006 which comprise 34 rules that are divided into four major principles, namely utility, feasibility, precision and quality, respect and ethics of evaluations (Cloete, 2014:316). While this is commendable, the reality is that Western countries still dominate the standards, competencies and guidelines for evaluations. At regional level the South African Monitoring and Evaluation Association (SAMEA) is still working in collaboration with stakeholders, such as DPME, to set up evaluation standards and competencies which are applicable to South Africa (Cloete, 2014:316). It can be noted, nevertheless that some evaluation criteria such as the OECD/DAC, which is currently being reviewed resonate well with contexts of developing countries (Cloete, 2014:343).

The lack of quality in evaluations or the empirical data utilised in evaluations has in some cases been linked to the relationship between donors and host governments. An evaluation of the Kenya education sector support showed that as long as relations with the government were good "donors were willing to look the other way concerning the data quality of indicators" (Grane, 2013:111). The same study revealed that as long as there was "no resistance or bad press" donors were happy to continue supporting the programmes (Grane, 2013:111). These are some of the issues which undermine the quality of evidence that are derived from evaluations on the continent.

These quality issues have certainly had implications on poor uptake. A compounding issue linked to uneven quality of evaluations is commissioners selecting methods without enough attention to local context. Inevitably, the epistemological frameworks of commissioners have been the key influencers in shaping the selection of methods. Furthermore, the African Evaluation Database (AfrED) provides empirical evidence of a preponderance of summative evaluations that are outcome and impact stories done by external non-African evaluation actors. This provides the lens for interrogating the epistemological framework of evaluators as the cause for why evaluations may have had limited use.

## *Barriers to capacity building*

Some of the challenges presented, related to the quality of the evaluations reviewed and the analysis of trends in this book, speak to significant obstacles in both evaluation practice and evaluation capacity development on the continent. Short course offerings have mushroomed, but many are delivered

by fly-by-night consultancies and do not meet the requirements of evaluation practitioners and commissioners. Furthermore, the firms that are doing evaluations themselves and therefore have contextually relevant, practice-based experience do not have capacity development as a mandate because they are more focused on technical assistance.

As such there has been a growth in demand for higher education institutions to play a larger role in institutionalising deeper offerings for this growing discipline. Nevertheless, there is a lack of coordination amongst universities to standardise curriculum and an absence of institutional mechanisms to regulate evaluation education. Furthermore, there is a paucity of skilled evaluation educators in the region, which is related to poor instructional design. There is no standardised pathway on how to acquire the right set of skills for this. The lack of resources in many African universities constrains quality, which impedes processes for approving new academic programmes.

With these weaknesses on the supply side, the gaps in evaluation capacity development are: an absence of standards and regulatory mechanisms for quality assurance to ensure robust post graduate M&E degrees offered by HEIs as well as a lack of consensus amongst them in how to standardise competencies; limited attention to systems level capacity building in emerging post-graduate degree courses; limited attention in training to ensuring methods are fit for purpose and use of evaluation. This has led to a reliance on building technical skills as a panacea for strengthening evidence use.

In relation to practitioners who make up voluntary organisations for professional evaluation (VOPEs) in this emergent field, there are limited pathways to development in the absence of standardised competencies. The continued donor reliance on Western consultancies and consultants begins with an absence of a comprehensive accurate African database of professional evaluators and M&E practitioners. As such, there is a growing creep of government advisory services from the big five taking on the role of M&E technical services, e.g. KPMG playing a growing role in the technical assistance space related to evaluation services. More senior evaluators continue to trot out a traditional reliance on Western methods, often leading to inappropriate analysis in practice. Local M&E practitioners are constrained in their involvement of donor-led evaluations as they cannot match the years of experience or technical requirements in the specifications of the bids that get put out. The few senior continental evaluators that have been active in this space are also part of the gatekeeping

that prevents a wider entry of newer and younger evaluators. This leads to a limited pool of professional evaluators and an overuse of the usual suspects, many of which have not questioned the ontological perspectives of the terms of refence drafted by commissioners. VOPEs have to date been ineffective in matching the supply and demand at the country level as the supply has not been adequately confident to articulate indigenous approaches and methods nor has it been good-governance driven (questioning the power dynamics around who commissions and degrees of inclusivity involved in the process). Nevertheless, through AfrEA, the tide is beginning to turn in establishing an agenda for an African-centred approach to evaluation that acknowledges culture, history and beliefs.

## The AfrED database

The African Evaluation Database (AfrED), the first database of its kind, provides us with a ground-breaking opportunity to use empirical data to discover patterns of evaluation practice on the African continent over a ten-year period. Anecdotal statements on the quality of evaluations conducted on the continent, the absence or presence of competent evaluators and the use of evaluation for evidence-based decision-making can now be tested using the vast repository of evaluations and evaluation-related articles that have been curated into the AfrED. It presents fertile ground for the establishment of research-led communities of practice (COP), particularly amongst academics and practitioners, towards building African scholarship in evaluation. This book is the first of a larger research strategy being driven by CLEAR-AA in the hope of widening and deepening African scholarship on evaluations to understand what interventions are needed to influence the learning agenda associated with this burgeoning practice.

At the start of this journey, some of the key research questions driving the research agenda that could draw from empirical findings in the database were:

- What implications do the findings of trends in evaluation have for training and education providers in these countries?
- How do we understand the changes in trends across sectors, actors and types of evaluations across the years? For example:
  - Who is commissioning evaluation?
  - In which fields?
  - What are the kinds of evaluation approaches and methods being requested?

- What can we learn from the different evaluation types in three different countries (comparative analysis of evaluation types and approaches) and what are the gaps in knowledge?
- To what extent are there differences in the intended uses of evaluation between governments and donor organisations?
- To what extent are cross-cutting issues like gender being incorporated into different evaluations?

This book is a first effort to either answer or add granularity to some of these questions through empirical findings from 2635 evaluation reports across twelve countries over a decade. As the following chapter will reveal, mining information from this database is not easy and a series of collective discussions with prominent practitioners, universities and development partners is key in terms of making it more accessible and useful. The remainder of this introduction provides a brief overview of the content of the book and some of the debates that chapter findings feed into or try to instigate.

The first chapter, 'Building the AfrED Database' by Blaser Mapitsa, simply provides an account of the approach taken to building the database. It includes a review of the basis on which articles and evaluation reports were included, as well as some reflections on the scope of the database. This chapter provides some descriptive information around the commissioning agencies and sectors reflected in the database. It concludes with reflections on some limitations of the database and both the opportunities and challenges faced by the partnership.

The second chapter of the book, 'Learning or Complying? Finding Consensus on the Purpose of Evaluations' by Blaser Mapitsa and Tsotsotso have drawn on a sample of 163 evaluations to determine trends in use by virtue of whether evaluation findings were being used for management purposes (measuring outputs) or governance (strategic reflection). They did this by developing a systematic review protocol and process by which evaluations got allocated according to whether M&E as a discipline is better located in governance or management. They then determined whether evaluations were measuring progress based on an agreed upon target, which would indicate evaluations are fulfilling more of a monitoring function or whether evaluations were being used for strategic decision-making and helping to foster a learning agenda. Of the 163 evaluations, more than half measure progress towards predetermined results, while only 12% have an exclusive focus on the strategic positioning of

programmes. The conclusion is that there has been a strong trend of focusing on progress towards specific results rather than evaluations looking more broadly at the programme strategy in context. The database analysis shows that evaluations are overwhelmingly managerial in nature and serve more of an output monitoring function than a platform for strategic decision-making.

Reasons underlying how evaluation evidence is disproportionately used for managerial purposes is related to the limited standardisation around what is termed an evaluation. While this is a common feature of emergent professions, it is also a symptom of the gap between the supply and demand of the evaluation profession. There is often a lack of shared understanding or vision between what commissioners want and commission versus what the evaluators interpret and are capable of. This speaks to the challenges presented later by Morkel and Mangwiro in Chapter 8, that the evaluation capacity development field needs to understand in order to try to build the bridge across this schism.

The third chapter of the book, 'Gender and Equity-Focused Evaluation in the SDG Era: A Call to Action' by Akibu, presents a content analysis, drawing on 36 academic articles linked to the database with a focus on gender and equity perspectives in evaluations. The analysis is guided by a set of quality criteria developed by the United Nations Evaluation Group. The findings under the "general report category", evaluation of projects not focused to address gender or equity, found 20% of the evaluation reports reviewed as satisfactory, 33% unsatisfactory and 45% highly unsatisfactory in their gender and equity focus. Similarly, under the "thematic report category", evaluation of reports focused on addressing gender and equity, the review found 25% of the evaluation reports highly satisfactory, 50% satisfactory and 25% unsatisfactory based on the gender and equity focus.

The findings from the qualitative appreciative enquiry, conducted on the reasons for the non-inclusion and consideration of the use of equity and gender perspectives in evaluations in selected West African countries, revealed a unanimous agreement from all the key stakeholders around the primary reasons. These are:

1. Gender and equity issues are not a requirement in the national evaluation policy for countries who have one.
2. The capacity to conduct gender and equity evaluations is weak.
3. Funds to conduct gender and equity evaluation is not readily available.

The study concludes that, for African countries to be on track in achieving the SDGs by 2030, their evaluation framework/policies need to include and consistently track the gender and equity dimensions in all commissioned evaluations. This approach could enable a more equitable delivery of projects and programmes in compliance with the SDG policy thrust, i.e. of leaving no one behind and reaching the furthest behind first.

The fourth chapter, 'Scandinavian donors in Africa: A Reflection on Evaluation Reporting Standards' by Ngwabi and Wildschut, provides a scan of the practice of Nordic commissioners (DANIDA, NORAD and Sida) of evaluations in Africa through a systematic review of 142 evaluations from the AfrED database. This sample makes up 20% of the evaluations in the database. The chapter reflects on the degree of structure and guidance provided to evaluators to write reports when drawing on the OCED/DAC criteria. Several key areas of findings emerge. First, the chapter provides the empirical evidence that the bulk of Scandinavian donors source consulting firms located in Nordic countries with the majority located in Sweden. Only 33% of these evaluations based in Africa were carried out by African firms, academic institutions or consultants. These findings speak to the first chapter regarding the need to address the barriers to entry in the evaluation market on the continent.

The second key important dimension of this chapter is what it raises around global standards set through the DAC criteria that have historically set the trend for defining evaluation practice. While Scandinavian donors look to the OECD as the governing body on standards, there is no standardisation in the guiding documentation given by Scandinavian commissioners, resulting in poor quality evaluation reports with limited credibility. The issue at hand here was limited quality assurance in place by Scandinavian commissioners. The chapter reveals that most of the 142 evaluations were based on non-experimental designs with limited mention of the evaluation designs. Of this sample, methodological sections of the evaluations were poor in that they did not even mention sampling techniques, partially due to the significant drawing on secondary sources of data for the evaluations. Finally, there was little attention to theories or approaches used. Out of the sample of 142 evaluations, 102 did not use an intervention logic or theory of change (TOC). The limitations of the findings reveal the evaluation reports commissioned by Scandinavian donors did not follow the detailed guides from the OECD. This reveals a lack of quality assurance of the evaluations published and the credibility of the results produced in these reports.

Clearly, these findings concur with Morkel and Mangwiro's hypothesis in Chapter 8 that a lack of professionalisation has led to a lack of uniformity in guiding the evaluation field. This is an interesting tension because it raises concerns about poor quality assurance by commissioners who prescribed the obligatory requirement of the DAC criteria covering evaluations commissioned between 2005–2015. These retrospective findings emerged in 2018 and 2019 amidst growing debates globally as to whether it is not time to revisit the DAC criteria itself.

The DAC criteria set the standard for what should be considered and measured and remain the most referenced guideline in international development. Specifically, these are five criteria – relevance, effectiveness, efficiency, impact and sustainability – that underpin most evaluation systems in international development. The empirical findings from scanning Nordic practice in commissioning evaluations reveal a prevailing vagueness in the use of these criteria. In reviewing quality assurance frameworks as employed by commissioners, the sixth chapter illustrates how the DAC criteria have a limited focus on the needs of the beneficiaries, as well as limiting the scope of an evaluation to the quality of the implementation process rather than the exportability of programme design.

These misgivings with the DAC criteria as they stand, resonate with growing public debate in the international community on how to move beyond the DAC criteria. The DAC criteria were originally used to support comparable thinking on assessing performance, as a need arose to standardise certain measures of quality in evaluation reports. In her January 2017 blog, the Secretary General of the Independent Evaluation Group, Caroline Heider, questioned whether it was not time for a "Copernican shift" in the evaluation field by revisiting the DAC criteria, considering how faithful adherence to its framework left many questions unanswered. She suggested greater attention to key issues, such as values, complexity, technology and the role these play in understanding local context and the underlying systems of the initiatives being assessed. Crawley takes this provocation a step further and provides a framework for a deeper and more purposeful interpretation of the DAC criteria through the Six Sphere framework (2017). This introduces additional analytical lenses (ideology, political, relational dimensions, contextual, technical and logistical) that can help shed a deeper light into impact and sustainability questions. For instance, the framework emphasises the importance of the relationship between donors,

recipients and beneficiaries for programme effectiveness. It also focuses on the need for a shift in beliefs and political buy-in for sustainable impact.

The fifth chapter, 'An Assessment of Methods used in Evaluations from the African Evaluation Database' by Ndhlovu, Khumalo and Krakra, provides a longitudinal trend analysis of the methods used from a sample size of 338 evaluations from the AfrED database.[1] It confirms the empirical findings of Chapter 4, which draws on a much smaller sample focusing on Scandinavian evaluations, that a disproportionate number of evaluations carried out on the African continent over a decade were done by Northern firms and consultants. Of the sample size of 338 evaluations, only 24% of evaluations were carried out by local academics and non-academics on the continent. As the vast majority of commissioners of these evaluations were Northern donors (e.g. Sida, USAID, DFID, OECD and UNICEF), 86.6% of evaluation commissioned were summative with a focus on outcome, impact and synthesis evaluations and the predominant methods used were qualitative and mixed methods. Of the 29 evaluations that were quantitative, making up 13,4% of the overall AfrED sample of evaluations, the most prominent methods were RCTs (34%) of which 59% were done by 3ie, survey designs (24%) and systematic reviews (24%). The large donors commissioning randomised control trials (RCTs) were 3ie, USAID, UNDP, World Bank and Sida. Most of the 12% that were formative fell into the implementation evaluation category and were predominantly done through mixed methods.

Ndhlovu, Khumalo and Krakra provide a qualitative framework with standardised questions to determine the quality of the designs used. This draws on using different data collection techniques by focusing on the logic of conducting the research, how data collection was carried out, how they went about collecting the data and so forth. The authors found that the logic of the research design for the qualitative evaluations was not adhered to. They concluded that there was no clear trajectory on how the research methods were conducted and that the overall reporting on methods was poor.

Picking up on this dilemma in Chapter 6, 'Appraising Qualitative Evaluations', bring Chirau, Tirivanhu and Ramasobana to narrow their scope to assess the calibre of quality assurance as practised by commissioners drawing on the

[1] This sample was drawn from evaluations in 2005, 2010 and 2015 where methods were discernible.

sample of 95 qualitative evaluations in the database. The chapter reviews the history of quality assessment frameworks and how they have come to be employed in the evaluation field by drawing out their key principles. The authors assess whether these principles were actually carried out by commissioners. In echoing some of the findings from Ndhlovu, Khumalo and Krakra (Chapter 5), the chapter finds that while data collection methods are outlined in the majority of evaluation reports, there is inadequate discussion as to why these methods were chosen, nor is there any outline of who collected the data or how the fieldwork could have influenced the data collected. Such attention to the "who" and "how" in qualitative methods helps strengthen the trustworthiness of the findings. The findings show a lack of consistency in how quality is maintained from a donor perspective. Finally, as most of the evaluations carried out in the sample were done by consulting firms or evaluators from the North, the epistemological framework underlying the evaluations may not have been sufficiently contextually grounded.

Methods take its starting point from the values and beliefs of evaluators and how this shapes their identity and their choice of knowledge frameworks for carrying out their evaluation practice. The deliberate focus on use by programme managers is what sets evaluations apart from research, with use as the focal point in carrying out an evaluation.

Chapter 7 examines the epistemological frameworks that underlie the evaluation practice that have largely remained implicit and therefore have been absent in the evaluation discourse. In 'Indigenising Evaluation Knowledge: Exploring the Epistemic Identity of African Evaluators', Tirivanhu and Blaser Mapitsa point to the importance of the epistemological grounding of evaluation in influencing the practice in this emergent field and how this view of knowledge must be integrated into evaluation capacity development. In the effort to look at the relationship between epistemological identity and how these shape choices of evaluation methods and approaches, the authors try to link evaluation to indigenous knowledge systems. The authors make this case by outlining how the deliberate focus on use by programme managers sets evaluations apart from research. The distinguishing point between evaluation and research is that evaluation entails a value judgement by determination of its alignment to a particular value system, a portion of which is inevitably linked to the geographical, historical, socio-economic and cultural context the evaluator comes from.

Chapter 8, 'Implications of Evaluation Trends for Capacity Development' by Morkel and Mangwiro, speaks to one of the prominent debates on the continent regarding evaluation as a profession as opposed to learning from it as a field of practice. Fleshing out the issues within this debate requires examining how to increase the number of local evaluators while simultaneously addressing the overall quality of evaluation practice by local evaluators on the continent. In order to address the latter, the chapter provides a snapshot of trends in evaluation supply, demand and practice in order to search for the kind of skills that may be required of evaluators in the future. What the scan of the landscape of ECD in Africa found is a noted emphasis on short-term training with little sense of what impact this has had over the years. This is due to a limited understanding of the multi-faceted dimensions of ECD that pay just as much attention to systems work and thinking as to individual capacity building.

The concluding chapter of the book reflects on the role of the database in generating knowledge for the evaluation sector and looks at how this particular foundation of empirical data has contributed to specific components of the learning and transformative agenda for evaluations. It especially considers the role data has to play in development for Africa, as well as some of the limitations it faces.

***Laila Smith, Nedson Pophiwa & Precious Tirivanhu***

## References


Australian Government. 2017. *Aid Program Performance Report 2016–2017*. Sub-Saharan Africa. Department of Foreign Affairs and Trade.

Beegle, K.; Christiansen, L.; Dabalen, A. & Gaddis, I. 2016. *Poverty in a rising Africa*. Washington, DC: International Bank for Reconstruction and Development, The World Bank. https://doi.org/10.1596/978-1-4648-0723-7

Carletto, C.; Jolliffe, D. & Banerjee, R. 2015. From Tragedy to Renaissance: Improving Agriculture Data for Better Policies. *Journal of Development Studies*, 51(2):133-148. https://doi.org/10.1080/00220388.2014.968140

Cauthen, N.K. & Fass, S. 2008. *Measuring Poverty in the United States*. New York: National Center for Children in Poverty.

Chilisa, B. 2016. A Synthesis Paper on the Made in Africa Evaluation Concept. Unpublished paper commissioned by the African Evaluation Association.

Chilisa, B. & Malunga, C. 2012. Made in Africa Evaluation: Uncovering African roots in evaluation theory and practice. Paper presented at the African Thought Leaders Forum in Evaluation and Development, Bellagio, Italy.

Chouinard, J. 2013. The Case for Participatory Evaluation in an Era of Accountability. *American Journal of Evaluation*, 34(2):237-253. https://doi.org/10.1177/1098214013478142

Christiaensen, L. & Shorrocks, A. 2012. Measuring poverty over time. *Journal of Economics and Inequality*, 10:137-143. https://doi.org/10.1007/s10888-012-9225-4

Cloete, F. 2016. Developing and Africa-rooted program evaluation approach, *African Journal of Public Affairs*, 9(4):55-70.

Devarajan, S. 2013. Africa's Statistical Tragedy. *Review of Income and Wealth*, 59(S1):S9-S15. https://doi.org/10.1111/roiw.12013

DFID (Department for International Development). 2017. *Statistics on International Development. Final 2016 UK ODA spend statistics*. London: Department for International Development.

Grane, M.D. 2013. "Indicatorism": The context, politics and effects of monitoring and evaluation in the Kenya education sector support program. PhD thesis, University of Iowa.

Heider, C. 2008. Rethinking Evaluation. [Online]. http://ieg.worldbankgroup.org/sites/default/files/Data/RethinkingEvaluation.pdf

International Monetary Fund. 2004. *Sub-Saharan Africa regional economic outlook*. Washington, DC: International Monetary Fund.

Kovaan, M. 2010. *Indigenous Methodology: Characteristics, Conversations and Contexts*. Toronto: University of Toronto Press.

Latib, S. Bringing politics and contestation back into monitoring and evaluation. *Journal of Public Administration*, 49(2):460-473.

Mackay, K. 2007. *How to Build M&E Systems to Support Better Government*. Washington, DC: The World Bank. https://doi.org/10.1596/978-0-8213-7191-6

Magombeyi, M.T. & Odhiambo, N.M. 2017. Causal relationship between FDI and poverty reduction in South Africa. *Cogent Economics & Finance*, 5(1):1357901. https://doi.org/10.1080/23322039.2017.1357901

Mouton, C.; Rabie, B.; Cloete, F. & De Coning, C. 2014. Historical Development and Practice of Evaluation. In: F. Cloete, B. Rabie & C. de Coning (eds), *Evaluation Management in South African and Africa*. Stellenbosch: African Sun Media. https://doi.org/10.18820/9781920689513

OECD (Organisation for Economic Cooperation and Development). 2017. Development aid at a glance, statistics by region: Africa. [Online]. https://www.oecd.org/dac/stats/documentupload/Africa-Development-Aid-at-a-Glance.pdf [Accessed 14 August 2018].

Podems, D. & Cloete, F. 2014. Evaluation Professionalisation and Capacity Building. In: F. Cloete, B. Rabie, & C. de Coning (eds), *Evaluation Management in South Africa and Africa*. Stellenbosch: African Sun Media. 312-336. https://doi.org/10.18820/9781920689513

Pradhan, M. & Ravallion, M. 2000. Measuring poverty using qualitative perceptions of consumption adequacy. *The Review of Economics and Statistics*, 82(3):462-471. https://doi.org/10.1162/003465300558821

UNICEF & World Bank. 2016. *Ending Extreme Poverty: A Focus on Children*. Briefing Note, October 2016.

Weber, K. 2008. *Accountability, Monitoring & Evaluation and Tailor-Made Monitoring*. The Hague: Netherlands Ministry of Foreign Affairs, Civil Society Division.

Wilson, S. 2008. *Research is Ceremony: Indigenous Research Methods*. Fernwood Publishing, Black Point, NS, Cananda.

1

# BUILDING THE AfrED DATABASE

Caitlin Blaser Mapitsa

## Background to the project

In 2015, a range of stakeholders in the African monitoring and evaluation sector, including the Centre for Learning on Evaluation and Results for Anglophone Africa (CLEAR-AA), the Centre for Research on Evaluation, Science and Technology (CREST), the South African Monitoring and Evaluation Association (SAMEA) and Twende Mbele, noted a gap in availability of empirical evidence in the African evaluation landscape. It further noted that this empirical data was fragmented. Stakeholders had differing perceptions about the scale of the field and the location of information. To address this, CLEAR-AA commissioned the Evaluation Research Agency, affiliated with CREST, to build a database of both peer-reviewed journal articles in the field of monitoring and evaluation, as well as evaluation reports.

Due to CLEAR-AA's mandate to support evaluation capacity development in the region, the database itself is a public good and can be accessed online by anyone interested.[1]

The development of the AfrED database was a first step to capture and present information on evaluation-related research and evaluations themselves in Africa. By providing this empirical foundation about the landscape for evaluation, it is hoped that scholars and practitioners will be able to more accurately map the gap between supply and demand, propel forward specific debates that are heavily based on anecdotal evidence, and develop effective and appropriate evaluation capacity building interventions.

## Scope of the database

The boundaries on the scope of the database were agreed on for pragmatic reasons. The database aims to capture academic journal articles, conference presentations, evaluation reports, terms of reference, and other clearly related evaluation documents between the years 2005 and 2015.[2] Only open access articles and reports were considered, due to the need for the resultant database

[1] https://crest2.sun.ac.za/african_evaluation_db/default/african_eval_db_01

[2] Acknowledgement: This is an adaptation of the Evaluation: Database and Survey Report, written by the Evaluation Research Agency, which was commissioned to build the database. Any mistakes remain those of the author.

to be open access. The geographical scope currently covers 12 sub-Saharan Anglophone countries:

1. Botswana
2. Ethiopia
3. Ghana
4. Kenya
5. Namibia
6. Nigeria
7. Rwanda
8. South Africa
9. Tanzania
10. Uganda
11. Zambia
12. Zimbabwe

## Methodology

### *Academic articles*

CREST extracted from the Thomson Reuters (TR) Web of Science database all articles from the 12 selected countries for the period 2005 to 2015. In this first extraction a number of keywords were entered in order to identify possible academic papers in the field of monitoring and evaluation (M&E). These keywords were: evaluation, impact assessment, monitoring and performance. The output files for the respective countries were subsequently imported into our Access database (CREST Evaluation Database). A total of 4503 unique papers were imported. For the Web of Science papers, these keywords were only applied to the 'article title' field as a broader search across other fields yielded too many results.

The second step of the process then involved a visual inspection of the article titles by the project lead. This manual process was time consuming but focused on eliminating "obviously irrelevant" papers in the field of the natural resources (such as the impact of climate change on crop production) or papers that reported on clinical trials. The end result of this process produced a list of unique evaluation papers by country. As a result of this process, a total of 543 unique evaluation papers remained.

Up to this point, the manual inspection of the database utilised the article title field. However, there were concerns that the title would not necessarily adequately indicate whether the study was in fact an evaluation study. The third step was to relink the 543 papers and extract them again from the Web of Science (WoS) database, but now with the Abstract of the paper included. Another round of manual inspection of the list of papers followed, taking into account the additional information provided by the paper Abstract. The end

result of this process was that 435 unique evaluation papers remained. The final step involved relinking the papers – through their unique identifier IDs – with the TR Web of Science in order to re-insert additional fields.

CREST extracted from the Elsevier Scopus online version of the database all articles from the 12 selected countries for the period 2005 to 2015. In this first extraction, the same keywords (as above) were entered to identify possible academic papers in the field of M&E. In this case, as we were working with the online version, these keywords were applied to the article title, abstract and keywords fields. The output files for the respective countries were subsequently imported into the Access database (CREST Evaluation Database). A total of 41 687 unique "evaluation" papers were imported. A subsequent filtering process applied these same keywords only to the article title field. As a result of the second filtering process, a total of 12 083 unique papers remained. After visual inspection 285 unique and appropriate evaluation articles remained. These records were merged with the 435 unique articles extracted from the Web of Science. At this point a final check revealed that there were some duplications between these two data files and these were removed. This resulted in a final datafile of 658 unique evaluation articles in the database.

This process reflected the situation of the evaluation sector: search terms were not necessarily effective, due to the use of monitoring and evaluation related terminology outside the sector, as well as varied use of semantics within the sector. This required an iterative process to define the scope of the database and the most effective way to fit within this scope. It also required actual engagement with the research methods section of each evaluation study, to ensure its relevance.

In the iterative definition process, it became clear that evaluation should be limited to refer to project or programme evaluation in the field of social services. Specifically, the database aims to target evaluations of interventions that are relevant to human development. This excludes the large bodies of literature that are evaluating natural events or physical phenomena. Even with this focus, there is still a large grey area of evaluations that are not technically of projects or programmes, but still relevant to the database. Table 1.1 below gives some examples of studies that were excluded, since they did not address programmes. On the other hand, those that may be looking at systems, policies, etc. are also included as they are deemed sufficiently relevant to the audience of the database.

**Table 1.1** Illustrative studies for exclusion or inclusion in the database

| Excluded study | Included study |
|---|---|
| The accuracy of mobile teleradiology in the evaluation of chest X-rays | Evaluation of the Agricultural Information Service (AIS) in Lesotho |
| Evaluation of spatial and temporal characteristics of rainfall in Malawi: A case of data scarce region | Impact of digital revolution on the structure of Nigerian banks |
| Mixed methods evaluation of targeted selective anthelmintic treatment by resource-poor smallholder goat farmers in Botswana | The impact of financial integration in Botswana |

*Source:* Mouton, Burg & Wildschut, c.2017. Close-out Report on the AfrED Database

## Evaluation reports

The process of identifying evaluation reports was more challenging. While academic journal articles are already indexed by a relatively small number of recognised role players, the evaluation space is not as coordinated. As a result, this process was less exhaustive, but created many important methodological lessons.

Concurrent with the downloading of scientific papers from international citation databases, commissioned evaluation reports/studies/working papers were downloaded from major international bodies that fund and commission such studies. The major web sites and online databases identified for this task are listed below and cover:

- Government and parastatal websites in the selected countries;
- Major NGOs operating in these countries; and
- The websites of the main international donors that work in Africa.

The process involved the following steps:

1. Identification of appropriate websites with databases of evaluation studies or reports (often through snowballing).
2. Reports that were produced between 2005 and 2015 for the countries included in this study were downloaded. These reports are currently archived in folders on a server at Stellenbosch University.
3. The core metadata that were available for these downloaded reports was entered into an Access database.

The results of these currently form part of the first version of the African Evaluation Database (AfrED). In addition, the documents have been downloaded and are available through a web-based Open Access repository. The African Evaluation Database contains the following standard bibliographic metadata (fields in the database):

1. Author(s)
2. Author affiliation (organisation)
3. Author affiliation (country)
4. E-mail address of corresponding author
5. Article title
6. Source
7. Document type (article, review, book review, report, policy document, etc.)
8. Year of publication
9. Keywords
10. Sector
11. URL Link to document

Table 1.2 provides information on the number of reports included in the database by source organisation.

**Table 1.2** Number of documents by source organisation

| Source | No of reports |
|---|---|
| 3ie | 713 |
| Government Website | 383 |
| Sida | 352 |
| ALNAP | 252 |
| OECD | 192 |
| UNICEF | 115 |
| NORAD | 111 |
| USAID | 88 |
| MEASURE | 78 |
| UNEG | 73 |
| DANIDA | 61 |
| DFID | 60 |
| IFAD | 60 |
| World Bank | 22 |
| UNAIDS | 6 |
| Other sources | 69 |
| **Total** | **2635** |

The process of sourcing evaluation reports was not exhaustive, which is one of the major limitations of the ability to use the AfrED database to generalise about the sector in the region as a whole. The scope of the databases contributing to

this study was ultimately limited by the availability of resources. The developers of the database do not make claims that it is exhaustive, and it is inevitable that certain important sources of evaluations have not been included. These include evaluation reports by the Department of Planning, Monitoring and Evaluation (DPME) in South Africa. CLEAR-AA is already in the process of identifying important gaps, and in the second phase of the database development an effort will be made to be more systematic and exhaustive in identifying and including evaluation reports.

## Description of the database content

This section presents some results from a preliminary bibliometric analyses of the datafile on evaluation articles (n=658).

Figure 1.1 presents the output by year for the period 2005 to 2015. The decline in numbers of papers in 2015 is most likely due to the lag in indexing such papers, which is common to both the WoS and Scopus databases. The overall trend shows that the current production of evaluation-specific articles (by African authors) averages around 85 per year.

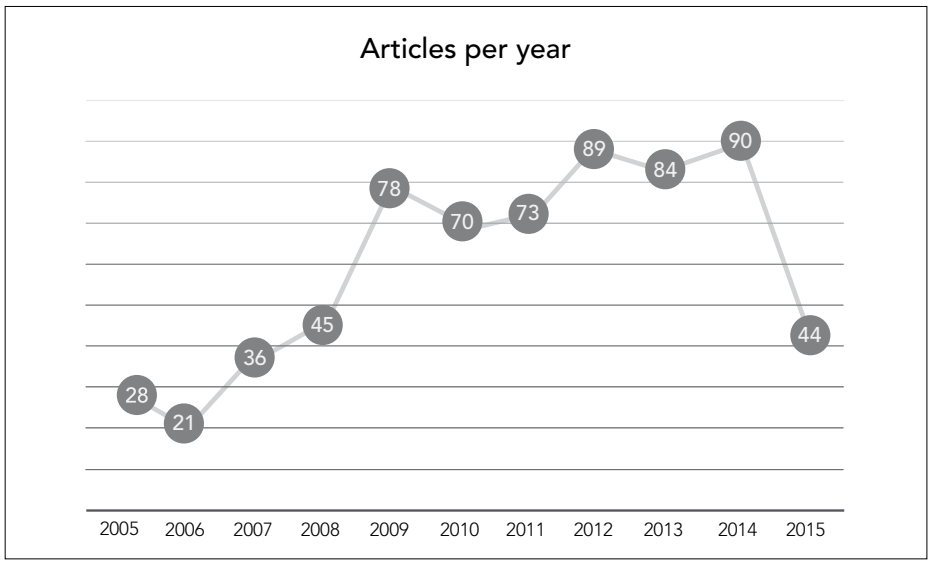



**Figure 1.1** Output of the year for the period 2005–2015

Only a very small number of articles (n=19) have been published in evaluation specific journals.

**Table 1.3** Articles published in dedicated evaluation specific journals

| Source title | No of papers |
|---|---|
| Journal of Testing and Evaluation | 3 |
| Evaluation and Programme Planning | 3 |
| New Directions in Social Impact Assessment: Conceptual and Methodological Advances | 3 |
| American Journal of Evaluation | 2 |
| Environmental Impact Assessment Review | 2 |
| Impact Assessment and Project Appraisal | 1 |
| Evaluation and Research in Education | 1 |
| Evaluation Review | 1 |
| Assessment & Evaluation in Higher Education | 1 |
| Research Evaluation | 1 |
| Journal of Evaluation in Clinical Practice | 1 |
| Total | 19 |

## *Survey methodology*

Recognising that the people behind the evaluation studies in the database represented an extensive collection of practitioners and experts in the field of evaluation in Africa, CREST combed the database for the e-mail addresses of contributing authors and evaluators. While the evaluation studies represent a valuable source of information about evaluation practice (the meeting place between supply and demand), the analysis of the database would be more robust with additional information about the background of those who participated in AfrED's evaluation studies. A survey instrument was designed to better understand the demographic profile, origin, educational background and professional identity of the individuals in the AfrED database.

Between 10 and 29 November 2016, e-mails were sent to 3032 persons identified as individuals involved with evaluations in Africa. Eligible candidates for the survey were obtained from three sources.

- The documents listed in the African evaluation database were examined and all e-mail details of authors recorded.
- E-mails were obtained from the South African, Ethiopian, Ghanaian and African Evaluation Associations.
- E-mails from CREST's own internal database (previous students of evaluation and a list of various individuals making enquiries).

Based on these sources, 3032 individuals were sent an invitation to participate in the survey. Of these, 426 e-mails returned undelivered and nine individuals indicated that they did not want to participate in the survey. We, therefore, assume that around 2597 individuals received the invitation. Each e-mail contained a description of the survey, some information regarding CLEAR-AA and CREST, and a request to participate in the survey. In total, 549 individuals indicated their willingness to complete the survey. These 549 individuals were then sent a subsequent e-mail with a link to participate in the survey. A reminder was sent after one week to complete the survey. Of the 549 persons who agreed to complete the survey, 421 completed the survey. On 11 November, the remaining 2051, who had not indicated their willingness to participate, were e-mailed a link to the survey. Of these 2051 individuals, 139 completed the survey. There were also 4 individuals who completed the survey through a general link provided. By 4 January (the date on which the survey was closed), a total of 564 individuals had completed the survey. This amounts to a 22% response rate.

### *Limitations*

An initial challenge, faced by both the database designers and the authors in this book, is that the AfrED database is a first of its kind in the region. As a result, the research questions from the database were not yet well established. For an initiative like this database, an iterative process can be expected of analysing the available data, focusing the research questions, and revisiting the data collection strategy. Having now reviewed the database and mapped out trends, gaps and key areas for research, important sources for evaluation reports and important fields of metadata are missing. However, due to a need to ensure an ongoing contribution to rapidly evolving debates, it is important to work with what we have while building what we need.

One of the most significant limitations of the database is around evaluations not included, due to a combination of limited resources and also a larger than expected pool of academic journal articles in the evaluation space. Accessible databases of evaluations were included, while some of the more difficult to access, importantly including public sector evaluations, particularly those with decentralised evaluation systems, did not form part of the database. Many evaluations from the public sectors of countries in the region that are leading in evaluation systems development are not included in the database. This means that the conclusions we draw about evaluations should not be generalised

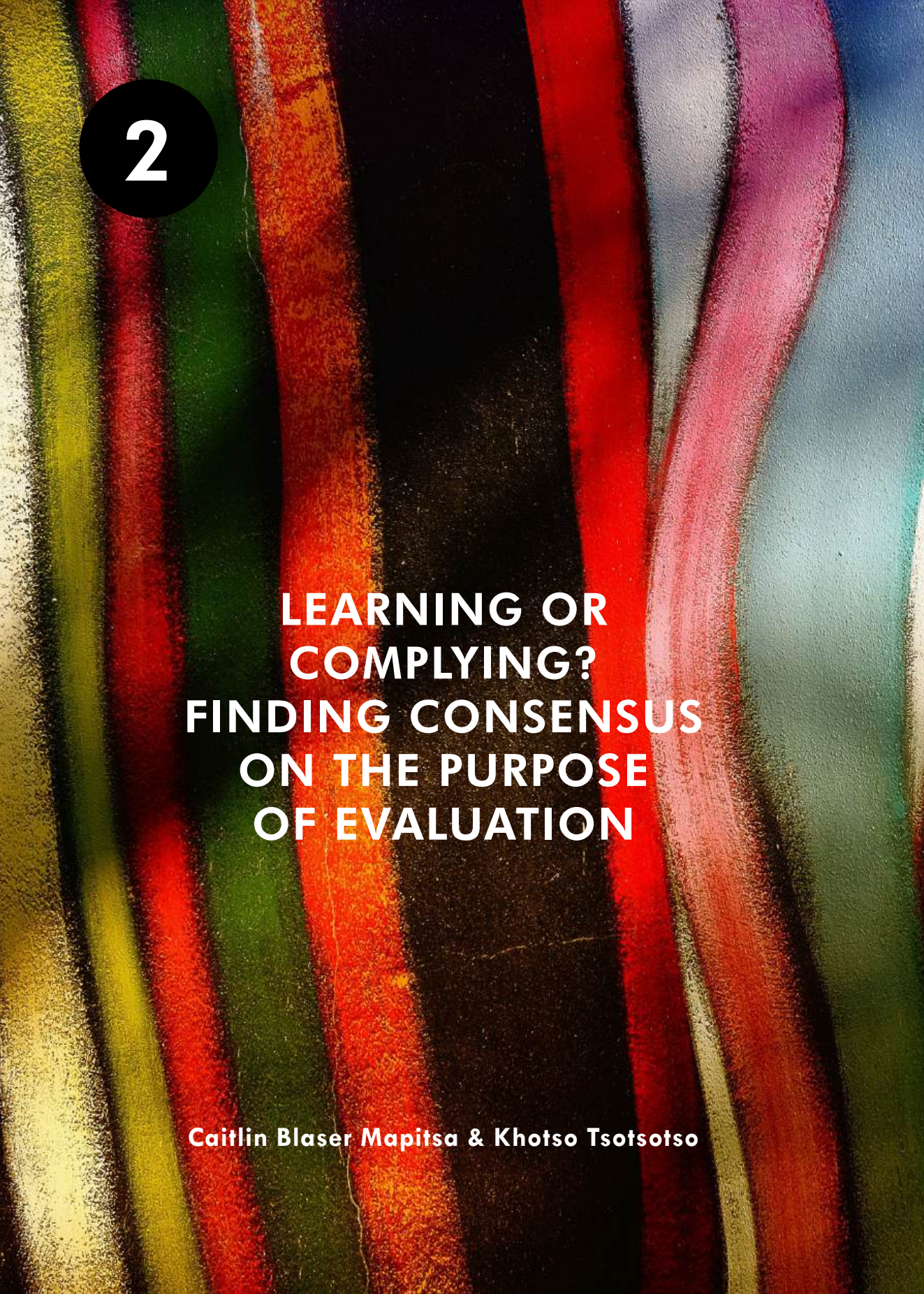
across the sector, but rather seen as indicative of certain trends. The conclusions also raise pertinent questions.

Finally, the database has limited metadata and as websites are updated, links expire quickly. In a future iteration of the database, first the research team has a much better understanding of the kind of metadata we need for the database to be a valuable contribution to scholarship in the region. Having engaged with many evaluations in detail, it is also more clear what the challenges are to accessing data in certain categories.

### *Way forward*

The first iteration of the AfrED database has been an important first step both in generating data on the evaluation landscape that has contributed to this book, but also in understanding key sources of information and how to navigate them. The process of writing this book, and engaging with the material in the database, has been critical in determining future directions for growth.

The concluding chapter of this book will discuss priorities for future research and the second iteration of the database, expected in 2019, which will continue to provide an empirical foundation for understanding the evaluation landscape in Africa.



2

# LEARNING OR COMPLYING? FINDING CONSENSUS ON THE PURPOSE OF EVALUATION

Caitlin Blaser Mapitsa & Khotso Tsotsotso

## Introduction

Monitoring and evaluation systems are growing rapidly across the public sector in Africa, sparking enthusiastic debate about what makes these systems most appropriate and effective in design (Goldman, Byamugisha, Gounoug, Smith, Ntakumba, Lubanga, Sossou and Rot-Munstermann, 2018; Porter & Goldman, 2013). One of the factors contributing to this challenge is the disproportionate focus governments in the region place on compliance and monitoring (Tarsilla, 2014), while evaluation aspires to play a strategic learning function. This chapter explores this tension by looking at the extent to which evaluations in the region are answering strategic questions, versus monitoring processes and outputs. This is important because there is a widespread assumption that a spread of evaluation systems both reflects and supports a concomitant expansion of a culture of learning and evidence-based decision-making. It is possible, however, that if there is not a shared view of the organisational role evaluations play, the evaluation function will simply become another tool of compliance.

Practitioners and commissioners often differ on the purpose for the evaluation function and, as a result, they place it differently on the spectrum from management to governance. Since evaluations also have a disciplinary focus based on the programme being evaluated, such as education, health and agriculture, the more subtle difference of organisational location often goes unnoticed. Different stakeholders in the evaluation process often make tacit assumptions about which level on the results chain an evaluation is targeting for decision-making, as well as appropriate evaluation questions and methodologies.

These differing assumptions often manifest as tension around the most appropriate process, method and product for an evaluation.

This is not a challenge that is new to evaluation. One of the founding fathers of the field described evaluation as "an overdue child of allegedly incompatible parents whose families opposed the marriage. And its difficult birth began a confused childhood exhibiting multiple personality disorder" (Scriven, 1994). One of the results of this disorder is that the diverse stakeholders of an evaluation process, who often hail from diverse disciplinary backgrounds themselves, often fail to establish shared aims of an evaluation process.

Furthermore, they may not fully understand and consider differences in values and perspectives, the appropriateness and utility of an evaluation (Reeve & Paperboy, 2007). Other authors frame this issue as a "lack of unified approach" (Khan, 1998).

One of the key fault lines of this debate is whether evaluation sits primarily in the field of management, or governance. This debate is as old as the practice of evaluation itself. Linimon and Borish (1981) describe a continuum in this role, "from the evaluator as a technician, to the evaluator as a change agent. At one pole of the continuum, the evaluator is portrayed as an illuminator, fulfilling the technical functions of information gathering and reporting for the party commissioning the evaluation. As the evaluator role shifts from this pole, the evaluator fulfils the additional functions of interpreting information, describing the social-political context of the evaluation, and identifying and advocating alternative courses of action." Often in evaluation reports, it is possible to tease out differences of opinion held by the commissioning organisation and the evaluator; evaluators often aim to place the evaluation function in a more strategic governing role, while commissioning agencies are more likely to locate it as a managerial requirement, with a tighter focus on outputs and compliance.

While the authors do not hold a normative view around a single, optimal way to locate the evaluation function, an important foundation to understanding the evaluation landscape in Africa is understanding the nature of current practice. This chapter explores the purpose of evaluations in the database, to try and understand whether evaluations are primarily sitting in the domain of management, or evaluation. Evaluations performing a management function are fulfilling monitoring objectives, of determining the extent to which predetermined goals and objectives have been achieved. Evaluation in the governance domain has a more strategic purpose, namely to understand the way in which interventions have worked in context, consider alternatives and explore whether the objectives set have been the right ones.

It is particularly important to understand this distinction, because one of the critiques of the monitoring and evaluation field in Africa is that it is donor-led and practice is still driven by non-Africans (Khan, 1998; Porter & Goldman, 2013). First, while the problem is widely acknowledged, it is usually mentioned as an assertion made without empirical backing. The AfrED database provides an opportunity to begin bringing evidence to this debate. The suggestion is that

donors may come with a different approach to the evaluation function, different methods of measurement, as well as different definitions for success. As a result, evaluations may not be contextually relevant (Kabore, 2013; Khan, 1998). However, data on this is limited, and this chapter will make some contribution towards this gap, interrogating whether various commissioning agencies place evaluation as a management or a governance function.

One more point of contestation in discussions of the African evaluation landscape is on the balance between monitoring and evaluation (Cloete, Rabie & De Coning, 2014). It is often assumed that, as monitoring and evaluation systems expand, monitoring is privileged over evaluation. Sufficient investment in systems of evaluation is rare. While monitoring is critical to build effective systems of management and understand processes of implementation, it is insufficient to make decisions of programmatic strategy and design. While evaluation has been the centre of attention over the last decade, it has long been recognised that even systems using the word evaluation may not necessarily fulfil an evaluative function. As one scholar pointed out nearly thirty years ago, "In fact, there was a wide gulf between evaluation precepts and practice. In practice, most evaluative reports of many donor agencies were monitoring types of efforts, concentrating on implementation issues rather than on assessing development results" (Binnendijk, 1989). Unfortunately, the debate does not seem to have shifted significantly in the intervening years, perhaps in part due to the dearth of empirical data to propel the debate.

While public sector M&E systems in the region put more emphasis on monitoring than on evaluation (Cloete et al., 2014; Abrahams, 2015), one of the findings of this study is that there are still several layers to peel away before we can reach any conclusions about the balance of monitoring and evaluation, and their implications for both evaluation practice and evaluation demand and use.

Finally, questions remain about the role of evaluations within the current public sector culture (Goldman, Mathe, Jacob, Hercules, Amisi, Buthelezi, Narsee, Ntakumba & Sadan, 2015; Cronin & Sadan, 2015). By looking at the purpose of evaluations as they currently stand, it will become more apparent whether or not evaluation can be a tool to either trigger changes in public sector culture and norms, or, alternatively, better understand the limitations of evaluation to catalyse change (Mayne, 2017).

## Background to management and governance in Africa

The lines between management and governance are not necessarily clear cut; with management being the implementation system of governance, the two functions can be described colloquially as "rowing" versus "steering". Governance explains how goals are set, while management determines the processes through which they are achieved.

Monitoring and evaluation in Africa have historically been linked to the donor space. One of the implications of this is that monitoring and evaluation related reforms have often been introduced to the public service by institutions of Treasury, linking the function to audit and compliance (Abrahams, 2015). The compliance-oriented nature of state practice has had a range of implications for the development of the monitoring and evaluation sector. One is that there is a limited choice of methods and analysis available to evaluators. Variable data systems have made evaluation practice costly and make the quality of M&E systems and practice variable.

From a governance perspective, this association with the donor space has created questions about how decisions are made and, in evaluation design, who decides on the most appropriate approach. There are questions of whether donor-led evaluations are contextually relevant, and whether there are appropriate power relationships between commissioners and users of evaluation. Finally, there is a risk that the learning loop is broken, given the different stakeholder influences.

Leaders within the evaluation sector often share an expectation that evaluation does not exist "merely to assess merit or worth in any given programme or project, but also to establish and institutionalise roles, relationships and practices among a wide breadth of development partners" (Rutkowski & Sparks, 2014) and "a component of political strategy" (Taylor, 2005). These expectations may not be shared by evaluation commissioners. This chapter explores the extent to which the evaluations actually carried out primarily address issues of management, or of governance. It begins a conversation about how different stakeholders are negotiating the role of evaluation, its implications for evaluation methods and ethical practice.

## Research methods

The objective of the study on which this chapter is based is to use the evaluations in the AfrED database to develop an empirical foundation to better understand the current purpose of evaluations being undertaken in Africa. This may be reflected in patterns identified in the purpose section of evaluations and methods employed, or trends linked to sectoral areas of work or commissioning agencies. The hope is that by mapping some empirical results, we will uncover additional questions around the current purpose of evaluations. In a context where national evaluation systems are growing within states that often focus heavily on compliance, it is important to understand if evaluation systems are 'pulling' states into a learning space, or, alternatively, if evaluations are adapting to a monitoring function.

In order to do this, the first step was to map out an approach that would allow us to categorise the evaluations, and then to identify emergent patterns.

## Inspired by systematic reviews

In order to locate the evaluations in the AfrED database in the fields of either governance or management, this study drew methodological inspiration from a systematic review approach. The study did not carry out a full systematic review, nor would it have been appropriate, given the database already providing a limited population of evaluations. However, following the steps of developing, testing, refining and implementing a protocol to categorise evaluations as playing a management role, a governance role, or both, drove the research approach. The study aimed to develop a tool that would allow for a replicable process for evaluations outside of the database population. With this in mind, the first five of eight steps outlined in the Cochrane Handbook (Higgins & Green, 2005) were adapted for this specific study as follows:

1. *Defining the review protocol, and developing criteria for including evaluations*

A first hurdle in this study was creating categories that broadly described the purpose of the evaluation exercise and its implications for both the location of the field from a professional standpoint, and the level of the results chain from a commissioning and use standpoint. To respond to these multiple goals, we found that the broad categories of management and governance best

encapsulated the concepts we were trying to apply. They both matched to disciplines in higher education and broad purposes of evaluation – from the commissioning and use side.

While there are some areas of overlap between management and governance, they are concepts that are widely understood as fulfilling different functions within the organisation. With a bit of iterative research, we could find mutually exclusive definitions that would serve the purpose of the study. Building on certain consensus definitions (Fukuyama, 2013; Dayton, 2001; UNESCO, 1998; De Janvery & Dethier, 2012), the following components were identified to allow us to describe the categories of management and governance more comprehensively.

**Table 2.1** Identifying the core constructs of management and governance in evaluations

| Management | Governance |
|---|---|
| Efficiency | Setting strategy |
| Working to a goal | Stakeholders |
| Implementing | Situational analysis/ context |
| Meeting targets | Relevance |
| Adhering to norms, standards and policies | Setting a goal |
| | Setting norms, standards and policies |

These constructs were then developed into a protocol which would be used to categorise evaluations as sitting primarily in either the discipline of management or in the discipline of governance, or broadly across both disciplines.

## 2. *Selecting evaluations*

While a true systematic review would have cast a wide net to identify evaluations, this study simply sampled within the database to search for evaluations. In our initial plan to sample the database, we anticipated that some subset of evaluations would explicitly discuss management or governance. We hoped to construct a purposive sample through searching for the terms directly. An initial sample of the database using the search terms of management and governance very quickly demonstrated that this was not the case. An unanticipated challenge of terminology was that a large number of the evaluations in AfrED database are in the health sector, and of these many deal with the management of disease.

While rigorous definitions of concepts of management and governance are critical to carrying out the review, this challenge demonstrated very early on that no shortcut to conceptualising these two areas would be possible.

Realising that the initial approach would not yield the expected results, a random sampling strategy was chosen. This would allow the authors to analyse results across the broadest range of sectors and countries, and also help uncover trends across the database as a whole. Furthermore, by constructing a representative sample, the authors hoped to refine the protocol in such a way that the conceptual constructs would be theoretically sound across differences in commissioning agency terminology or thematic area of work.

This study used Excel as a random sampling statistical package. Excel allocated a random number adjacent to each evaluation name in a separate column in a spreadsheet. Once each item was assigned a random number, the spreadsheet was ordered from the smallest random number to the largest. In order to arrive at an appropriate sampling size, a power analysis was conducted. The sample size is based on a 95% level of statistical confidence. From a population of 2635 evaluation reports in the AfrED database, a sample size of 163 was achieved.

### 3. *Testing and refining the protocol*

Once the sample was constructed, the protocol itself was tested and refined. To do this, an additional ten evaluations were brought in as a trial, and three reviewers applied the protocol to the same ten evaluations to see if it was valid. It was revised and retested twice with additional sets of ten evaluations before the researchers had a construct for each discipline that was sufficiently valid. Through the testing process, two additional categories were created. One was both management and governance, for evaluations which clearly included at least one component from both categories. A fourth category was created for evaluations that were primarily focused on methods, or the evaluation process itself, since these evaluations did not fit into any of the categories created. They were nevertheless included in the AfrED database.

As part of the process of trailing the protocol, the research team discovered that few evaluations were explicit about both the purpose of the evaluation, and the way in which their methods achieved this purpose This reflected previously discussed differences in perspective between evaluation commissioners and

practitioners. Rather than being a quick process of matching stated objectives to the categories developed, categorising the evaluations was a much more analytical process of teasing out and clarifying the stated objectives, interpreting them (based on what was actually done in the evaluation process), and then looking at the level of results and conclusion.

### 4. *Applying the protocol*

Due to the challenges above in constructing a valid protocol, the researchers chose a rigorous process of applying the protocol. Two senior researchers categorised each of the 163 evaluations into one of the four categories (management, governance, both, or methods) indicating at least two components of the category from Table 2.1. The researchers compared lists and identified all differences. Each difference was individually discussed and resolved, either as an issue of human error or substantive disagreement. In the cases of human error, the mistakes were rectified, and the evaluation was considered categorised. In the cases of substantive disagreement, each researcher presented their logic; in most cases, this process identified an area of either substantive oversight or logical fallacy. This was then rectified and the evaluation categorised. In two of the 163 cases the researchers held substantive disagreements and a third senior researcher was called in to moderate. Through this process, consensus was reached on the categorisation of each evaluation in the sample. Throughout the process, researchers kept detailed reflective journals of lessons or questions that emerged through the categorisation process.

Once all evaluations in the sample were categorised, data was analysed according to sector, commissioning agency, and country. Additionally, a thematic analysis of the protocol implementation journals informs the structure of the results presented below.

## Methodological challenges

This research provides a strong empirical basis to argue that evaluations currently undertaken in Africa are overwhelmingly managerial in nature. They serve more of an output monitoring function than a platform for strategic decision-making. However, this piece of research is far from exhaustive and faces a number of limitations that are worth noting.

## The construction of the database itself

The AfrED database was quite exhaustive in its collection of academic research articles on monitoring and evaluation. However, due to the volume and scope of evaluations undertaken in the region, only a subsection of evaluations conducted during the period covered by the database were able to be included. This will be built in the coming years. The database thus does not represent the full population of evaluations undertaken in the region, so the results of this study remain indicative rather than rigorously generalisable. There are important commissioning agencies that are not reflected in the database; they may bring different approaches.

## What is an evaluation?

One unexpected challenge of this study was determining what exactly constitutes an evaluation. While this was dealt with in detail in the methodology chapter for the purposes of database construction, we confronted a number of situations where the database included studies that were not traditional programme evaluation reports. Some examples of these include:

- A national evaluation policy framework;
- A barometer on the national development assistance environment;
- A process document on the development of a monitoring and evaluation system;
- Case studies from evaluative processes; and
- A national level sectoral study.

While all of the above are clearly linked to evaluation in a range of ways and some stem from one or more specific programme evaluations, the products of evaluative processes are not necessarily at the same level. Documents that were not traditional evaluation reports were at times difficult to assess using tools that were designed for evaluation reports only. This did not represent a large portion of the sample, but did make it apparent that while certain organisations have standardised evaluation report structures, in many cases there is a spectrum of products – from research reports to country synthesis studies to barometers. Drawing the line around what exactly is an evaluation was not always a straightforward process.

## Developing a shared conceptual approach to evaluation purpose

One of the most difficult methodological challenges in categorising the evaluation reports was that terminology varied widely across evaluations. There was very little standardisation even within an evaluation report. The stated aims of the evaluation were often not, upon examination, what was done. While this lack of standardisation is common in an emergent profession like monitoring and evaluation, this may be symptomatic of the gap between evaluation supply and demand (Stem, Margoluis, Salafsky & Brown, 2005). Commissioners, who define the purpose of an evaluation in terms of reference, may not successfully develop a shared vision with the evaluation consultant who carries out the evaluation. This results in a disjuncture between the stated aims and the actual work carried out.

## Results

This study clearly demonstrates that across a range of fault lines there are very different views on both the purpose and practice of evaluation. Commissioning agencies differ in their approach to the evaluation function; organisations locate evaluations at different levels of the results chain, and design evaluations for different purposes. Furthermore, different sectors have different norms around evaluation practice and purpose. Health and education, two sectors with a large number of evaluations in the database, seem to target evaluations lower on the results chain than other sectors. Finally, the purpose of evaluations differs by country as well. Interestingly, countries that are leading in evaluation systems development are those that are also targeting their evaluations lower down on the results chain. As country led evaluation systems grow in Africa, understanding these differences and their implications for the evaluation sector, will be important to building systems that are useful and contextually relevant.

## Description of the evaluations

For the purposes of this study, 163 evaluation reports were randomly selected from the database. Figure 2.1 describes the sample by sector. Most of the evaluation reports in the sample are public health focused. This is indeed a true reflection of the AfrED database as public health represents the biggest group of evaluations in the database. Evaluations focusing on governance and formal

education also account for more than 10% of sample proportions (15% and 10% respectively). Health, public policy (including donor strategy) and education are amongst the highest areas benefiting from evaluation studies across the region. Other significant sectors include clinical health – mainly driven by the level of resourcing devoted to the Global Fund on Malaria, Tuberculosis, and HIV/AIDS, in the past decade (World Health Organization, 2014); social cohesion and poverty reduction, which are both mainly driven by high migration levels in the region (UN Department of Economic and Social Affairs, 2016).

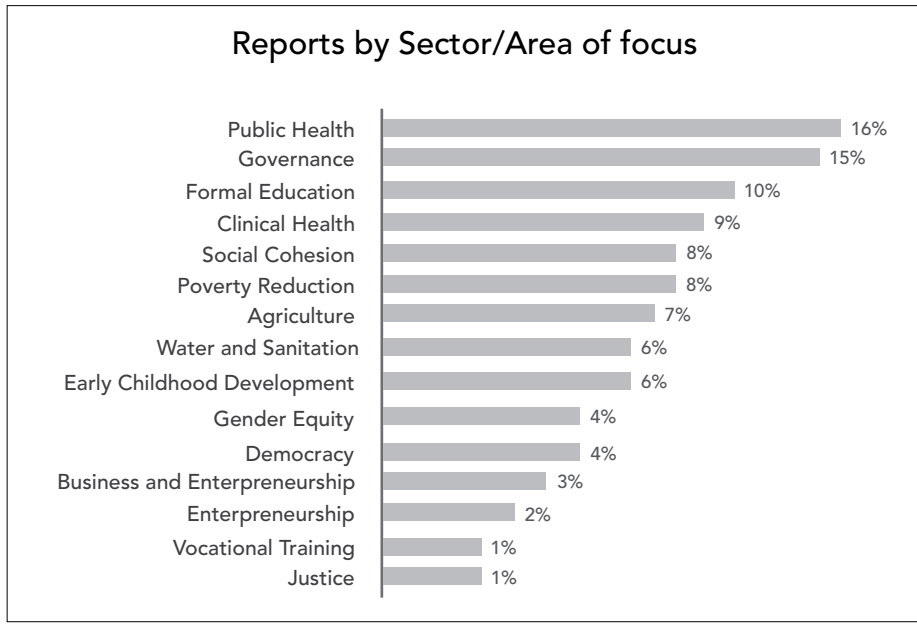



**Figure 2.1** Evaluations by sector (based on AfrED database)

Figure 2.2 shows a breakdown of sample evaluations by evaluating agency. Considering the list of agencies and their respective work in social development in Africa, one can speculate about the purpose of the respective evaluations in the sample. This is especially the case with an organisation such as 3ie which funds many impact evaluations and generally promotes highly scientific evaluation methods for programme effectiveness. The sample also shows evaluation leadership by international agencies broadly. Finally, it is encouraging to observe a relatively significant representation of evaluations received from country departments and various spheres of government (at 14.3% of the sample).

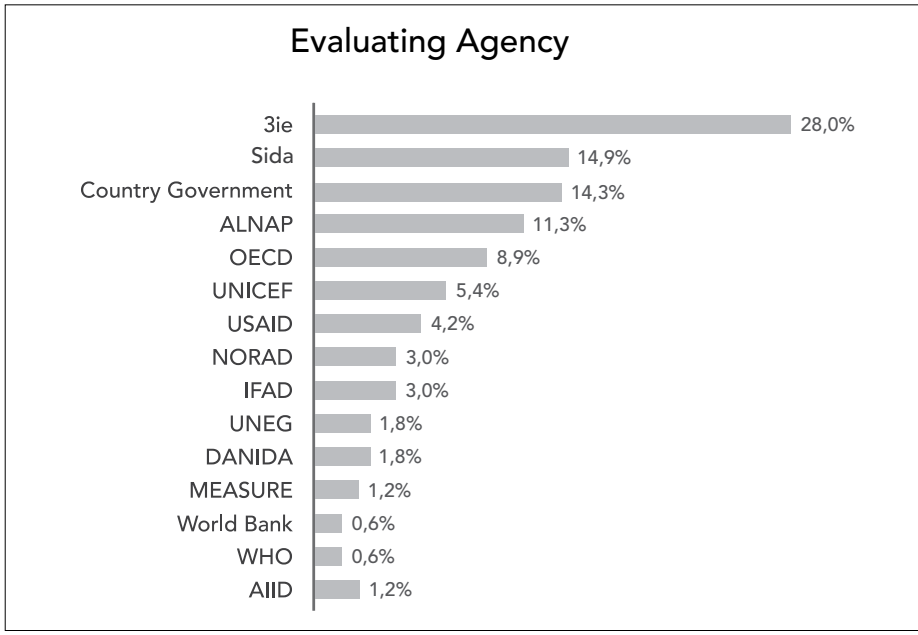



**Figure 2.2** Evaluations by agency (based on AfrED database)

A common assertion in the field of monitoring and evaluation is that the field in Africa remains donor led, and that this comes with a range of related problems of appropriateness, incentive structures, and ownership (Bornstein, 2006). While the initial construction of the database does not yet allow us to make conclusive empirical findings on this, it does provide enough information to begin a conversation that is grounded in research. Figure 2.2 does suggest that this assertion holds merit. Fewer than one in six evaluations in the database are commissioned by a national agency, with all the rest led by bilateral or multilateral donors. Research is still emergent on the ways in which commissioning organisations shape evaluations (Porter & Goldman, 2013). There is a strong push globally to strengthen national ownership of evaluations and evaluation systems (Green, 1988; Kirby et al., 2003).

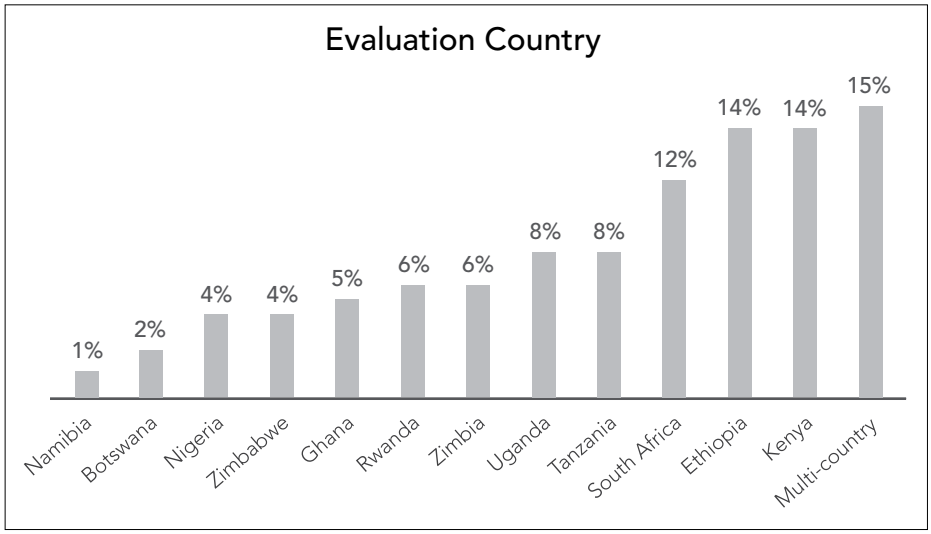



**Figure 2.3** Evaluations by country (based on AfrED database)

Figure 2.3 shows a breakdown of evaluation by evaluation country. Three clear trends emerge from this analysis:

1. There is a growth in popularity of multi-country programmes and programme evaluation. Perhaps this can be attributed to the emerging popularity of contextual studies in evaluation practice, especially by donor agencies.
2. The obvious interest of the donor community in Kenya and Ethiopia as popular destinations for donor funding. These countries have also featured significantly in multi-country evaluations.
3. Finally, it is also interesting to observe evaluation production rates of South Africa and Uganda as the two countries with the most mature national evaluation systems (NES) in the region (Genesis Analytic, 2017). The sample shows that most of the evaluations from the two countries are government commissioned as opposed to donor commissioned as is the case in other countries.

## The purpose of evaluation

Of 163 evaluations, more than half exclusively measure progress towards predetermined results, while only 12% of evaluations have an exclusive focus on the strategic positioning of programmes. This strongly endorses the hypothesis that evaluation practice in the region emphasises progress towards specific results, rather than looking more broadly at programme strategy in context.

Given the existing research indicating that monitoring and evaluation practice in Africa heavily emphasises monitoring (Cloete, 2009), it is important to highlight that all evaluation is not equal. If evaluations are functionally monitoring implementation processes, then even emergent evaluation practice is still not serving the needs of an evaluation system but is rather an extension of the purpose of monitoring.

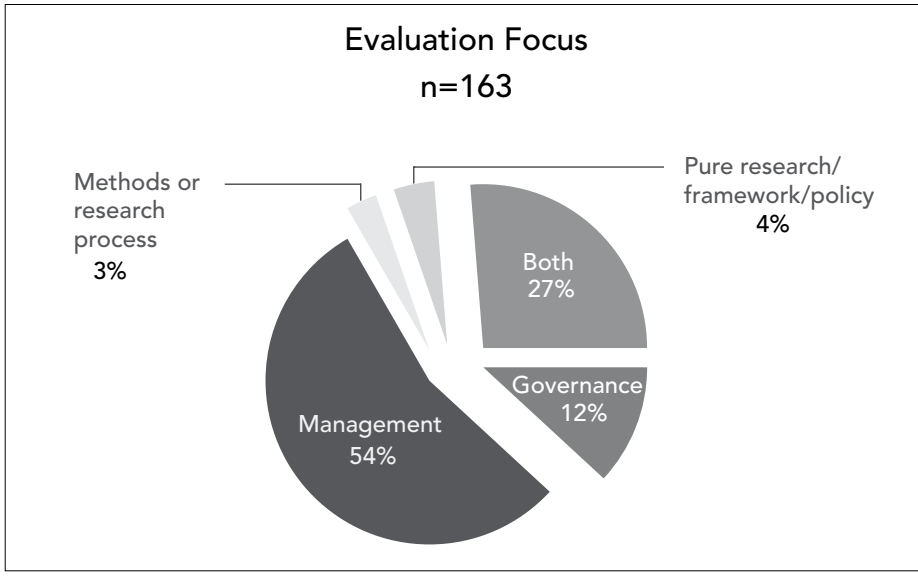



**Figure 2.4** Overall purpose of evaluation (based on AfrED database)

This research does not come with a normative prescription on the ideal balance of management and governance in evaluation practice. However, given the criticism that the field of monitoring and evaluation is not enabling of strategic decision-making, due to a strong compliance focus (Govender, 2013), some sort of balance needs to be struck. At this moment, empirical evidence from the database demonstrates a disproportionate focus on monitoring functions. Clearly there is a need to shift the evaluation function into a more strategic decision-making position.

Looking at the purpose of evaluation by sectoral area, there is variation across sectors. Over 80% of all evaluations in the education sector is in the realm of management, versus less than half that percentage of evaluations with a broad social focus. On one hand, it seems appropriate that health and education lend themselves better to quantitative approaches They are measuring progress on agreed upon outcomes, since these established fields have significant

collective agreement around pre-determined outcomes. More emergent fields appropriately take more exploratory approaches. However, this may also illustrate a limitation in the ability to innovate in the fields of health and education.

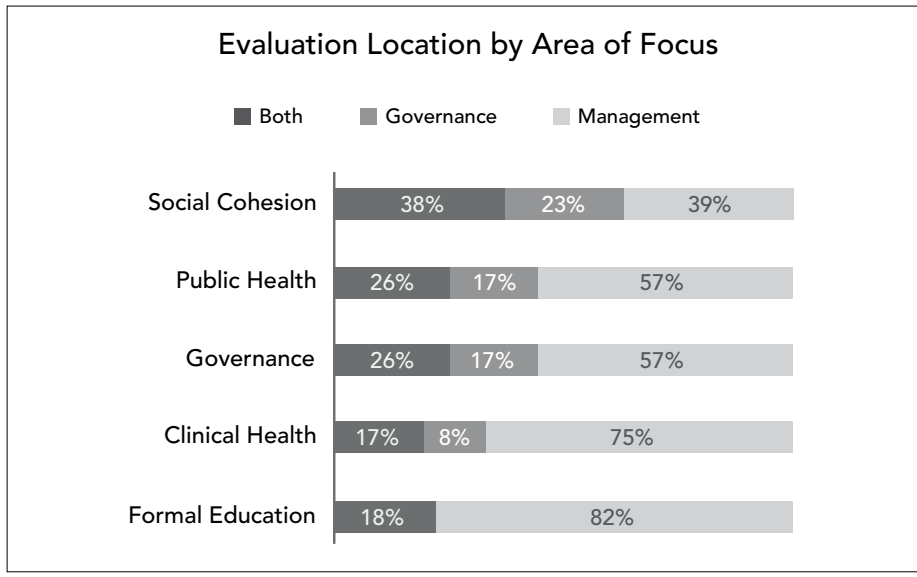



**Figure 2.5** Purpose of evaluation by sector (based on AfrED database)

One important motivation for this study is to understand the current purpose of evaluation in the region in the context of a growing demand for country-led evaluation systems. This is an imperative of the Sustainable Development Goals regime. Progress in developing national evaluation systems in Africa has been uneven, but momentum is growing rapidly (Goldman et al., 2018). As the national imperative for strengthening evaluation systems grows, it is important to understand the variation of evaluation purpose amongst countries in the region.

Figure 2.6 illustrates evaluation purpose by country within the database. It demonstrates two important trends. The first is that South Africa and Uganda, the two countries with the strongest evaluation systems in the region, demonstrate the least strategic evaluation purpose. As the progress continues towards increasing country ownership of evaluation systems, this finding indicates that country ownership in itself is not sufficient for a robust evaluation system. It is but one component of a successful evaluation system, which should include a consideration of the purpose and results of evaluations themselves.

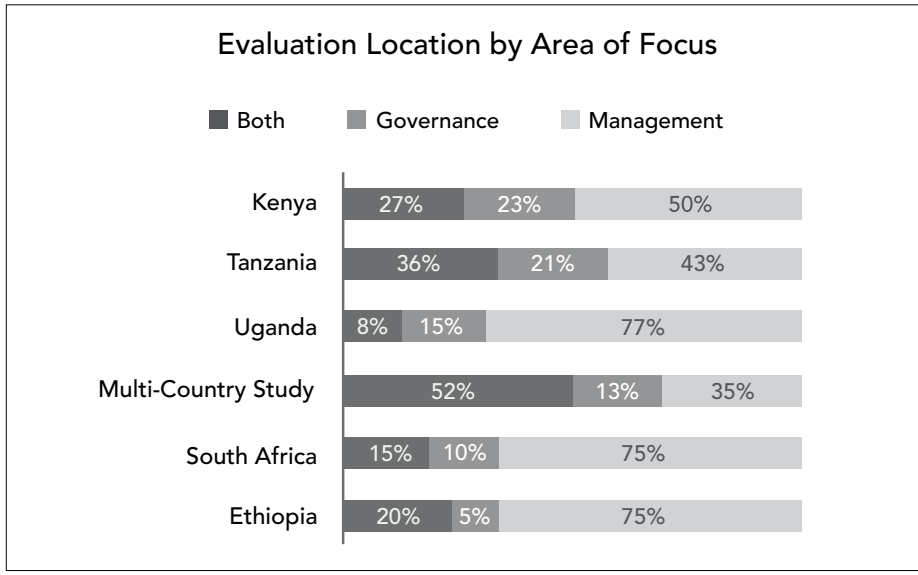



**Figure 2.6** Evaluation purpose by country (based on AfrED database)

A second interesting result, reflected in Figure 2.6, is that multi-country studies are significantly more likely to deal with more emergent issues and use evaluations for strategic decision-making and programme governance. More research is needed to confirm why this is the case, but possible hypotheses could include the need for relatively more groundwork to understand the context across multiple countries. Alternatively, there may be something inherent in the nature of multi-country programmes, whether it is in their funding or institutional arrangements, that makes resources more likely to be available for an evaluation that addresses issues of programme governance.

It is clear from the discussion above that currently evaluation is primarily playing a management function in the region. However, there are notable exceptions, with specific evaluations, commissioning agencies, and sectors which tend to move more into the governance space. It is worthwhile to look at the pockets of innovation that can be found. From our samples, Kenya is doing the most evaluations in the governance space, and this is happening across a range of thematic areas. In this case, public health is an important driver of innovation (as opposed to clinical health research, which is largely less exploratory). While education does not feature, early childhood development does. Evaluations with a broad social remit, such as from fields of poverty reduction and governance, tend to maintain a strategic focus.

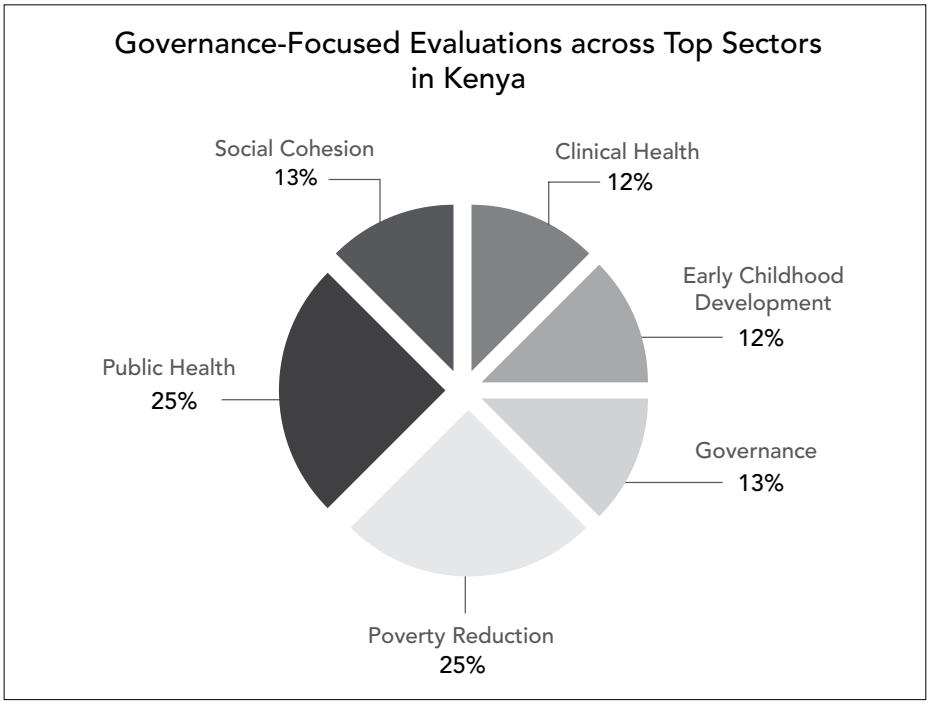



**Figure 2.7** Governance-focused evaluations in Kenya (based on AfrED database)

## Fit-for-purpose methods

The section above clearly demonstrates the lack of consensus amongst various stakeholders about the purpose and scope of the evaluation function. There are variations across organisations and also differing norms in different disciplines. When data collection began, there was an intention to collect data on the methods employed for each evaluation in the sample. However, in keeping with the findings of other authors in this volume, such as in Chapters 4 and 5, evaluations do not often explicitly discuss the methods used, so this approach was dismissed. Rather, methodological findings and trends were noted in the researcher journals, and these were drawn on for discussion below.

Figure 2.8 demonstrates that different organisations have different understandings of the purpose and practice of evaluations. As the authors of Chapter 5 in this volume highlight, this differing understanding of the evaluation functions also translates into different methods employed through evaluation practice. Some of this variation can be understood by considering the nature

of the organisation in question. It seems appropriate that ALNAP, which deals with humanitarian evaluations, should focus on emergent practice and look extensively at strategy in context.

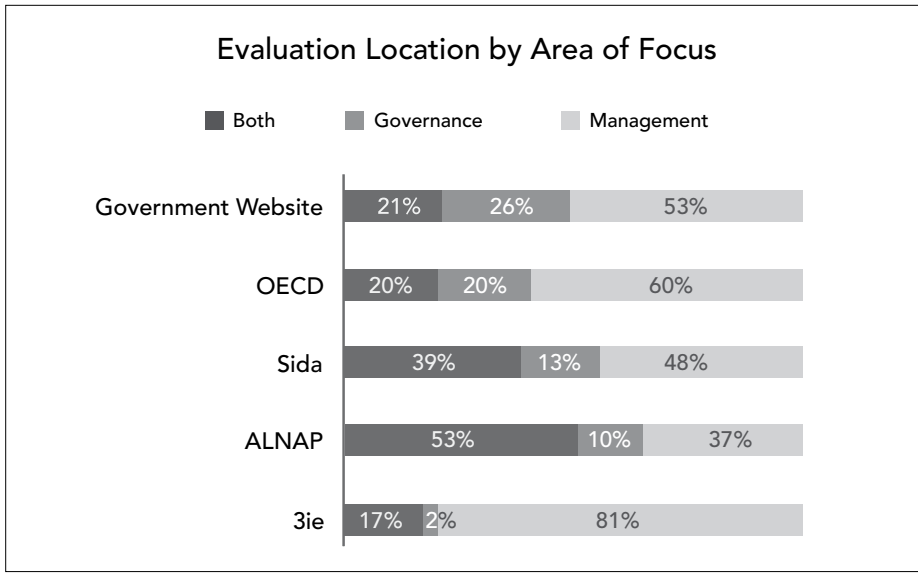



**Figure 2.8** Evaluation purpose by commissioning agency (based on AfrED database)

3ie, on the other hand, is an organisation dedicated primarily to impact evaluations. 'Impact', like other terms in the evaluation sector, is used differently by different stakeholders, creating some confusion around semantics (White, 2010). While the organisation takes a theory-based approach to understanding programmatic contribution to certain results, the organisation also defines impact evaluation as a specific quantitative methodological approach to understanding the attribution of a specific indicator. This single evaluation method answers critical research questions about programmatic contribution, but is explicitly not well equipped to consider questions of context, alternatives or unintended consequences. This is well understood by 3ie and other experienced evaluators, which is why other approaches to impact evaluation are also available (Rogers, 2009), and this specific method is considered one tool in the arsenal of evaluators. However, as a result, nearly every evaluation led by 3ie is measuring progress towards predetermined goals and objectives. There has been a tremendous growth in both the resourcing and

popularity of quantitative, RCT-based impact evaluations of social programmes over the last decade (White, 2010). This is an important shift to recognise and understand in the evaluation landscape, and consider what it means for the role evaluation plays in organisational practice and development.

This raises one area for exploration in fit-for-purpose evaluation methods. There is a large body of research in the field of methods that suggests people 'trust' results stemming from quantitative research more, and that qualitative research is more strongly tied to trust in the researcher as an individual (Kuhn, 2012; Patton, 2002). While there is also ample research presenting a counterpoint around qualitative research approaches, one particular challenge is that with a utilisation-focused approach to evaluation, perceptions may be more critical than validity of data to ensuring ownership and use of resultant evaluations. This study does not presume to enter into a debate on a hierarchy of research methods. However, an important area for future research is to understand the African context around credibility of methods, appropriateness of these methods (which Tirivanhu and Blaser Mapitsa in this volume explore in further detail, see Chapter 7). Empirically grounding this debate may help illustrate some of the challenges evaluation practitioners face in meeting the aspirations of an evaluation that is of strategic importance, with 'credible' methods and approaches.

Similarly, the OECD takes a prescribed approach to evaluations, adhering strictly to the OECD/DAC criteria of relevance, effectiveness, efficiency, impact and sustainability. These criteria have a long history in the evaluation space, but have recently been subject to discussion on areas of improvement and their ongoing contribution to the evaluation space (Chianca, 2008). Comparing these criteria to the components of the protocol originally developed on management and governance, some of the criteria fall into the realm of governance, while others fall into the realm of management This suggests that most OECD/DAC oriented evaluations should be categorised as fulfilling both a management and governance purpose. However, digging deeper into the way the criteria are applied in practice demonstrated that this evaluation approach was applied narrowly in most cases, looking only at programme implementation and achievement of immediate results.

## Supply and demand gaps reflected in the landscape

One of the most striking findings of applying the protocol to the evaluations in the sample is that very often there was significant discrepancy between the purpose as stated in the introduction of the evaluation, and the actual methods and results in the evaluation undertaken. In over half of the evaluations which were not categorised in the same way by researchers in the first round of applying the protocol, it was because the researchers primarily drew on evidence from different sections of the evaluation. Most often, the introductory section which stated the purpose of the evaluation placed the evaluation purpose firmly in the strategic governance space. However, after looking at what was actually done, it was clear that the evaluation could only serve a managerial purpose. This suggests that there is an aspiration by some stakeholders for evaluation to play a governance function, but there are barriers in place in the evaluation process. Future research may need to focus on better understanding the nature of these barriers.

On the demand side, the construction of public sector evaluation systems is such an important feature of the evaluation landscape in Africa at the moment, a closer look at evaluations commissioned by the public sector is useful. While we saw above that, in general, evaluations commissioned by government do not tend to be exploratory in nature, Figure 2.9 demonstrates that this aggregate analysis is not sufficient, and that there is considerable variation in the purpose of evaluations within government. First, the figure reinforces the finding above, that non-state evaluations are better suited for innovative learning. Local government in particular is conducting evaluations that are concentrated lower on the results chain. While further work is needed to explore this, it could be a reflection of the compliance-driven nature of much local government in the region or could demonstrate a reliance on monitoring systems for evaluation practice. Finally, it is interesting to note that within the public sector, evaluations in the education space are more exploratory than evaluations in the education sector as a whole, as reflected in the database. This could be linked to methods employed, or it could indicate that governments in the region have an urgent need to learn and innovate in the education sector.

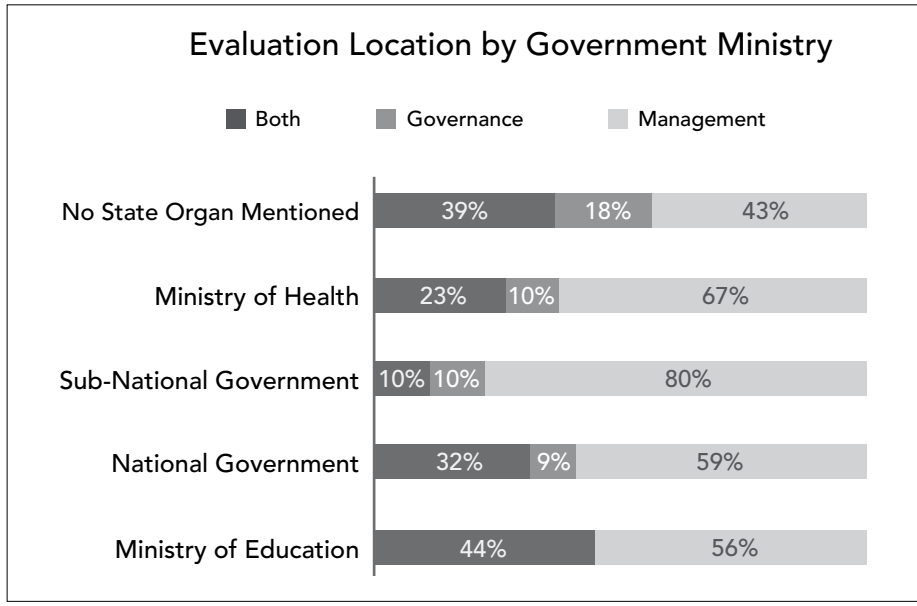



**Figure 2.9** Evaluation purpose by public sector ministry (based on AfrED database)

On the supply side, there is limited empirical data looking at the identity of evaluators in Africa, and perceptions around the credibility of evaluation results and subsequent evaluation use. However, one of the key assertions in the regional evaluation debate is that non-African evaluators do not come with sufficient contextual understanding of the region to generate appropriate and useful evaluations. This is being explored in more detail by Tirivanhu and Blaser Mapitsa (see Chapter 7), but institutional affiliation of individuals is also important to consider. Internal evaluations are one clear manifestation of an approach to bridge the supply and demand gap, but they have been insufficiently interrogated in the African evaluation literature. In the AfrED database, over a quarter of evaluators report an internal affiliation, and nearly a third were some combination of internal and external. Understanding how this shapes the purpose of evaluations and concomitant methods will be an important next step.

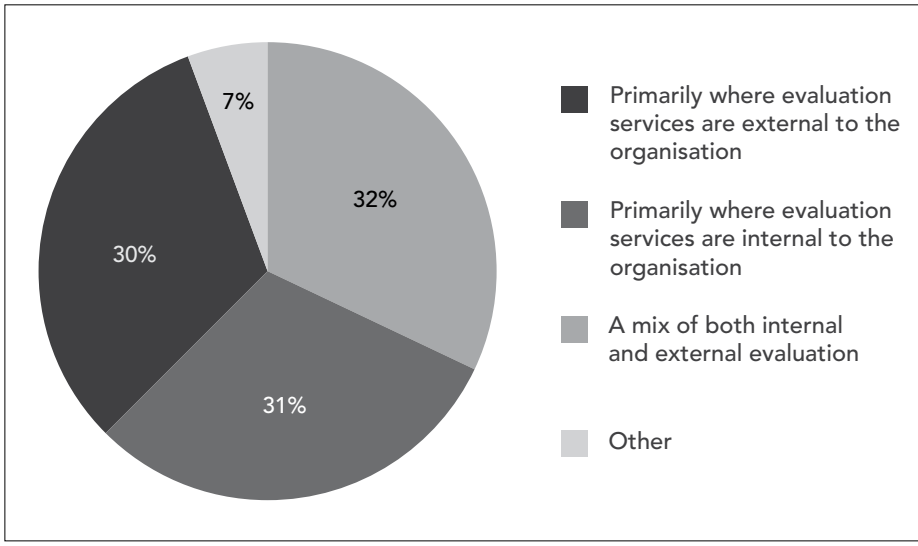



**Figure 2.10** Institutional local of evaluator

Global literature offers some information about the institutional identity of evaluators and perceptions of credibility. Even though internal evaluations are often of equal or superior quality to externally commissioned evaluations, those conducted by someone external to the organisation are usually considered more credible (Conley-Tyler, 2005). Anecdotal information suggests this may also be the case in Africa and raises additional questions about whether there is an outsider-credibility bias. Given the predominance of questions around donor-linked evaluation practitioners in the region (explored by Ngwabi in this volume, see Chapter 4), it is an important area for further exploration.

Additional research is needed to understand these phenomena in detail, but the supply and demand gap in the region reflects a combination of factors, including the differing perspectives of stakeholders on the purpose and scope of evaluations, as well as variation in resources available for the evaluation process. On the one hand, there are some incentives for some evaluation practitioners, particularly those who also identify as researchers, to keep evaluations high on the results chain. At the same time, there are incentives to reduce the scope of work for contracted consultants and keep the evaluation remit narrow and specific. It will be important to better understand how different stakeholders articulate these divergent incentives, but from the data in this study it is apparent that the stated objective is more strategic than the methods as implemented.

When evaluators and authors of evaluation related research were surveyed in the African database about the purpose of the evaluations they conducted, the data presents a dilemma. The most important finding is that accounting to donors is the least common purpose of evaluations, with 109 of 460 respondents never conducting evaluations for accountability. Decision-making and improving practice are the most common purposes. However, organisational learning also remains relatively insignificant, with only 179 respondents frequently conducting evaluations for learning. This may suggest that this gap between intention and application of the evaluation function is a manifestation of the gap between supply and demand.

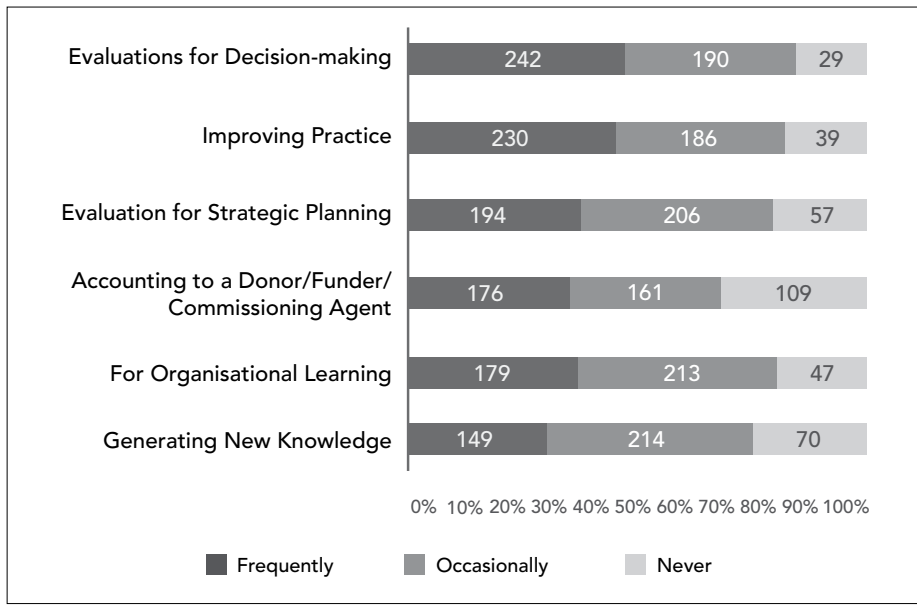



**Figure 2.11** Purpose of evaluations conducted (AfrED database close out report)

What these anecdotal experiences underscored is that there is little consensus about the purpose of monitoring and evaluation, and where on the results chain or organisational practice the monitoring and evaluation function sits. So far, the thrust of the debate has been around professionalisation of evaluation (Basheka & Byamugisha, 2015), taking an angle of evaluation practice. However, the evaluations on the AfrED database provide us with an opportunity to look at the evaluations themselves, and see what the stated purpose of the evaluation was, whether the methods were appropriate to the purpose, and to what extent, from the commissioning side, it appears that evaluations are being undertaken to address results at a low level of the results chain, or a high

level of the results chain. As different role players in the region are exploring the gap between the supply and demand for evaluation (Abrahams, 2015), we hope that this chapter will help define this gap by identifying key questions around the stated purpose for evaluations, whether the methods and evaluation implementation responded appropriately to this purpose, and what implications this may have for both the professionalisation debate and the discussions on evaluation demand and use.

## Conclusion

The empirical results from the database debunk some myths about evaluation purpose and practice in Africa, but confirm truth in others. The primary result of this study is finding that evaluations in the region have a disproportionate focus on the low end of the results chain, focusing on processes of implementation and progress towards predetermined results, rather than broader analyses triangulating context, implementation process, and results. This means that even though evaluation practice in the region is growing, the purpose of evaluation is still very closely tied to monitoring. This both reflects a lost opportunity for learning, but also calls into question whether an expansion of evaluation systems in the region is likely to achieve the transformative affects on culture and programme design that are anticipated. Several core purposes of the evaluation function, such as considering alternative programmes to achieve the same objective and passing judgement on the choice made, are not reflected in current evaluation practice.

Linked to the finding above, is a demonstration that various stakeholders, from evaluation practitioners to commissioners, do not necessarily share a view on the purpose of evaluation, and do not necessarily hold common definitions of the distinction between monitoring, and evaluation. As Jane Davidson, one of the founders of evaluation, said, "the lack of a clear distinction between the two means that evaluation is getting short-changed" (Davidson, 2013). This is certainly the case in the evaluations studied.

The growth of evaluation systems in the region present a unique opportunity to begin to build this consensus around the purpose and practice of evaluation in the region. However, this study has demonstrated that public-sector led evaluations in particular marginalise the evaluatory function, and are instead using evaluation as an extension of monitoring and compliance. As the role

governments have to play in building evaluation systems grows, it is important both to be aware of this dominant ideology, and its implications for the growth of the evaluation sector. It is particularly important to continue the currently dynamic discussion around what the role of evaluation is, and should be, in the context of Africa's public sector management systems.

An additional finding this chapter explored, but was not able to conclude, is the relationship between the purpose of evaluations, and the methods employed. It is appropriate that evaluations that were centered around a large quantitative component focused on generalisability, while evaluations that were looking more at emergent context employed qualitative methods. However, the strength of this linkage, in addition to other components of evaluation methods, such as degrees of participation, and the role of the commissioning organisation, are an important next step in further understanding the purpose and practice of evaluation in the region. This is a particularly important area of exploration given the current global trends towards impact evaluations, randomised control trials, and other methods that may not be best suited to responding to questions of context. Evaluation systems that grow around an assumption that these methods represent a 'gold standard' will, in all likelihood, be seen as a more managerial tool.

There is a need for greater exploration amongst all stakeholders around the purpose of evaluations in Africa, and for the discourse both professionalisation of the sector, contextually appropriate evaluation methods, and purpose of evaluation systems in the region to be grounded in empirical data on current practice. Furthermore, this foundation of data should provide a focal point for stakeholders with diverse views to engage, and, if not reach areas of consensus, at least acknowledge areas of disagreement to minimise tension. At this moment, accusations of poor-quality evaluations, inappropriate methods or unskilled evaluators are often masking a fundamental disagreement around the purpose and potential of evaluations.

## References


Abrahams, M.A. 2015. A review of the growth of monitoring and evaluation in South Africa: Monitoring and evaluation as a profession, an industry and a governance tool. *African Evaluation Journal*, 3(1):1-8. https://doi.org/10.4102/aej.v3i1.142

Bamberger, M.; Rao, V. & Woolcock, M. 2010. *Using mixed methods in monitoring and evaluation: Experiences from international development*. Washington, DC: The World Bank.

Basheka, B.C. & Byamugisha, A. 2015. The state of Monitoring and Evaluation (M&E) as a discipline in Africa. *African Journal of Public Affairs*, 8(3):75-95. https://doi.org/10.4102/apsdpr.v3i2.80

Binnendijk, A.L. 1989. Donor agency experience with the monitoring and evaluation of development projects. *Evaluation Review*, 13(3):206-222. https://doi.org/10.1177/0193841X8901300302

Bornstein, L. 2006. Systems of accountability, webs of deceit? Monitoring and evaluation in South African NGOs. *Development*, 49(2):52-61. https://doi.org/10.1057/palgrave.development.1100261

Chianca, T. 2008. The OECD/DAC criteria for international development evaluations: An assessment and ideas for improvement. *Journal of Multidisciplinary Evaluation*, 5(9):41-51.

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.

Cloete, F.; Rabie, B. & De Coning, C. (eds). 2014. *Evaluation Management in South Africa and Africa*. Stellenbosch: African Sun Media. https://doi.org/10.18820/9781920689513

Conley-Tyler, M. 2005. A fundamental choice: Internal or external evaluation? *Evaluation Journal of Australasia*, 4(1/2):3. https://doi.org/10.1177/1035719X05004001-202

Cronin, G.P. & Sadan, M. 2015. Use of evidence in policy making in South Africa: An exploratory study of attitudes of senior government officials. *African Evaluation Journal*, 3(1):10. https://doi.org/10.4102/aej.v3i1.145

Davidson, J. 2013. Monitoring and evaluation: Let's get crystal clear on the difference. *Genuine Evaluation*. http://genuineevaluation.com/monitoring-and-evaluation-lets-get-crystal-clear-on-the-difference/

Dayton, N.A. 2001. Total quality management critical success factors, a comparison: The UK versus the USA. *Total Quality Management*, 12(3):293-298. https://doi.org/10.1080/09544120120034447

De Janvry, A. & Dethier, J.J. 2012. *The World Bank and governance: the Bank's efforts to help developing countries build state capacity*. The World Bank. https://doi.org/10.1596/1813-9450-6275

Fukuyama, F. 2013. Democracy and the Quality of the State. *Journal of Democracy*, 24(4):5-16. https://doi.org/10.1353/jod.2013.0074

Higgins, J.P. & Green, S. 2005. Cochrane handbook for systematic reviews of interventions. Hoboken, NJ: Wiley-Blackwell.

Goldman, I.; Byamugisha, A.; Gounou, A.; Smith, L.R.; Ntakumba, S.; Lubanga, T.; Sossou, D. & Rot-Munstermann, K. 2018. The emergence of government evaluation systems in Africa: The case of Benin, Uganda and South Africa. *African Evaluation Journal*, 6(1):1-11. https://doi.org/10.4102/aej.v6i1.253

Goldman, I.; Mathe, J.E.; Jacob, C.; Hercules, A.; Amisi, M.; Buthelezi, T., Narsee, H.; Ntakumba, S. & Sadan, M. 2015. Developing South Africa's national evaluation policy and system: First lessons learned. *African Evaluation Journal*, 3(1):1-9. https://doi.org/10.4102/aej.v3i1.107

Govender, I.G. 2013. Monitoring and evaluation systems for enhancing governance and government legitimacy in South Africa. *Journal of Public Administration*, 48 (Special issue 1):811-823.

Jacob, S. & Boisvert, Y. 2010. To be or not to be a profession: Pros, cons and challenges for evaluation. *Evaluation*, 16(4):349-369. https://doi.org/10.1177/1356389010380001

Jesover, F. & Kirkpatrick, G. 2005. The revised OECD principles of corporate governance and their relevance to non-OECD countries. *Corporate Governance: An International Review*, 13(2):127-136. https://doi.org/10.1111/j.1467-8683.2005.00412.x

Kabore, I. 2013. Commentary on country-led monitoring and evaluation systems, better evidence, better policy, better development results. *African Evaluation Journal*, 1(1):1-3. https://doi.org/10.4102/aej.v1i1.46

Khan, M.A. 1998. Evaluation capacity building: An overview of current status, issues and options. *Evaluation*, 4(3):310-328. https://doi.org/10.1177/13563899822208626

Kirby, P.; Lanyon, C.; Cronin, K. & Sinclair, R. 2003. *Building a culture of participation. Involving children and young people in policy, service planning, delivery and evaluation*. London: National Children's Bureau.

Kuhn, T.S. 2012. *The structure of scientific revolutions*. Chicago: University of Chicago Press.

Linimon, D. & Borich, G. 1981. Evaluator as technician versus evaluator as change agent: A continuum of power relationships. Paper presented at the Joint Annual Meetings of the Evaluation Network and the Evaluation Research Society, Austin, Texas.

Marshall, M.N. 1996. Sampling for qualitative research. *Family Practice*, 13(6):522-526. https://doi.org/10.1093/fampra/13.6.522

Mathison, S. 1994. Rethinking the evaluator role: Partnerships between organizations and evaluators. *Evaluation and Program Planning*, 17(3):299-304. https://doi.org/10.1016/0149-7189(94)90009-4

Mayne, J.W. 2017. *Monitoring performance in the public sector: Future directions from international experience*. New York: Routledge.

Mouton, C. 2010. The history of programme evaluation in South Africa. PhD thesis, Stellenbosch University.

Patton, M.Q. 2002. Two decades of developments in qualitative inquiry: A personal, experiential perspective. *Qualitative Social Work*, 1(3):261-283. https://doi.org/10.1177/1473325002001003636

Porter, S. & Goldman, I. 2013. A growing demand for monitoring and evaluation in Africa. *African Evaluation Journal*, 1(1):1-9. https://doi.org/10.4102/aej.v1i1.25

Reeve, J. & Paperboy, D. 2007. Evaluating the evaluation: Understanding the utility and limitations of evaluation as a tool for organizational learning. *Health Education Journal*, 66(2):120-131. https://doi.org/10.1177/0017896907076750

Rogers, P.J. 2009. Matching impact evaluation design to the nature of the intervention and the purpose of the evaluation. *Journal of Development Effectiveness*, 1(3):217-226. https://doi.org/10.1080/19439340903114636

Rutkowski, D. & Sparks, J. 2014. The new scalar politics of evaluation: An emerging governance role for evaluation. *Evaluation*, 20(4):492-508. https://doi.org/10.1177/1356389014550561

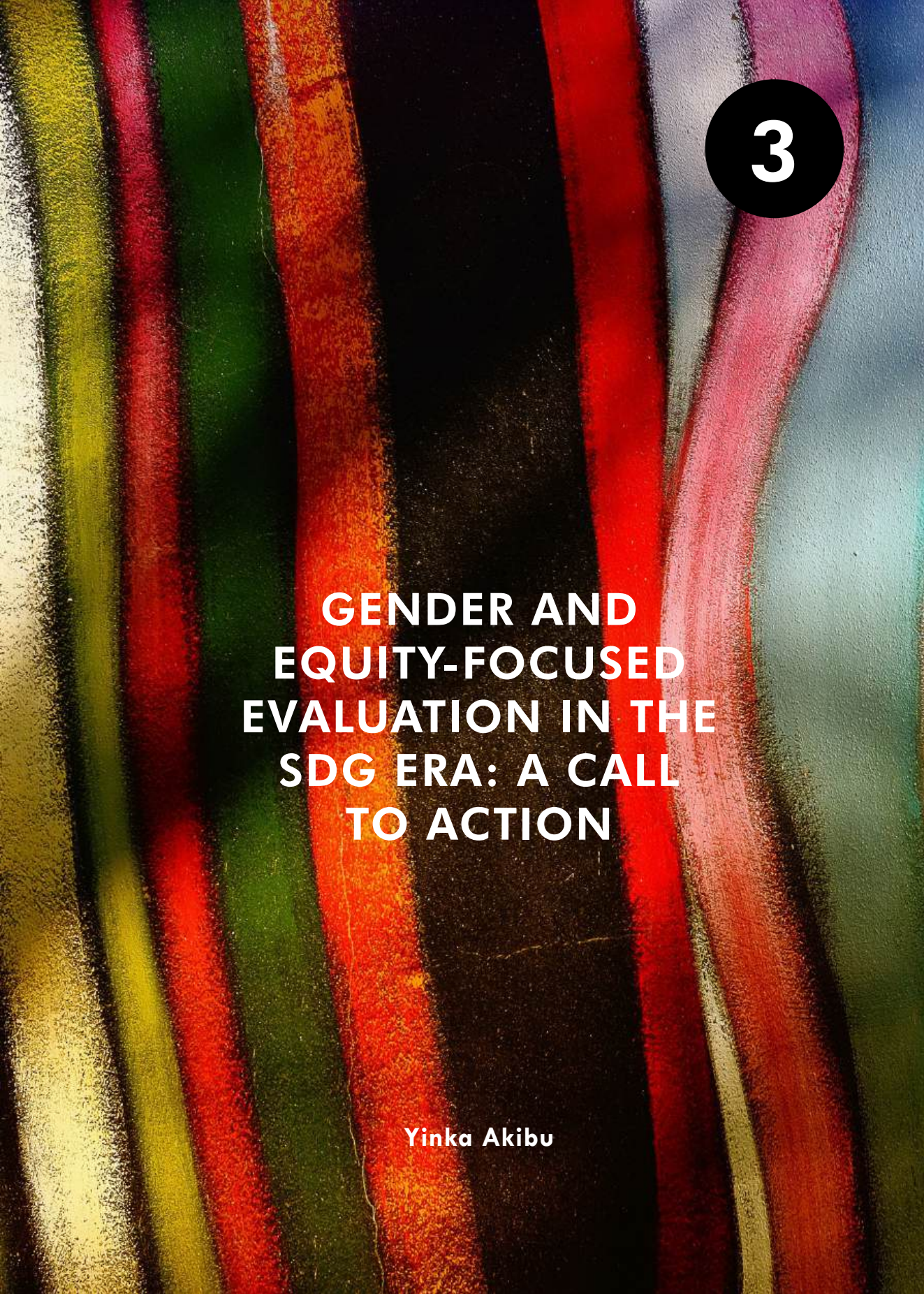
Scriven, M. 1994. Evaluation as a discipline. *Studies in Educational Evaluation*, 20(1):147-166. https://doi.org/10.1016/S0191-491X(00)80010-3

Stem, C.; Margoluis, R.; Salafsky, N. & Brown, M. 2005. Monitoring and evaluation in conservation: A review of trends and approaches. *Conservation Biology*, 19(2):295-309. https://doi.org/10.1111/j.1523-1739.2005.00594.x

Taylor, D. 2005. Governing through evidence: participation and power in policy evaluation. *Journal of Social Policy*, 34(4):601-618. https://doi.org/10.1017/S0047279405009177

White, H. 2010. A contribution to current debates in impact evaluation. *Evaluation*, 16(2):153-164. https://doi.org/10.1177/1356389010361562

World Heath Organization. 2015. *The health of the people: What works – the African Regional Health Report 2014.* Brazzaville: Regional Office for Africa, World Health Organization.



3

# GENDER AND EQUITY-FOCUSED EVALUATION IN THE SDG ERA: A CALL TO ACTION

Yinka Akibu

## Introduction and background

Equity and gender equality are the core challenges of contemporary development, while evaluation serves as a means to assess the extent to which different social groups have equitable access to development outcomes. For African countries to be on track in achieving the Sustainable Development Goals (SDGs) by 2030, their evaluation frameworks and policies need to include and consistently track the gender and equity (G&E) dimensions in all commissioned evaluations. However, a content analysis of 36 evaluation reports, guided by the United Nations Evaluation Group (UNEG), found that 78% of the evaluations were unsatisfactory or highly unsatisfactory in their approach to gender and equity. Similarly, a qualitative approach revealed a unanimous agreement from all key stakeholders that the primary reason for the non-use of gender and equity dimensions are: (i) gender and equity issues are not a requirement in the national evaluation policy for countries who have one; (ii) the capacity to conduct gender and equity evaluation is weak, and (iii) funds to conduct gender and equity-focused evaluation are not readily available. This questions the relevance of national evaluation systems to respond to current development needs.

The evolving development context necessitates that evaluations be focused on informing policy decisions, plans and programmes, and for holding state actors accountable for development results. The way development evaluations are rolled out and conducted needs to reflect the commitment countries made in achieving the SDGs. This has resulted in a push by development and evaluation experts for the harmonisation and better alignment of the evaluation criteria to meet the information needs of all stakeholders, donors and beneficiaries alike. The OECD/DAC Standards for Development Evaluation have served as a good practice in evaluation built through international consensus. However, with the complexity and the interconnectedness required in delivering sustainable development based on the agenda 2030, it becomes necessary that the five OECD/DAC criteria (relevance, effectiveness, efficiency, impact and sustainability) are expanded to integrate other important dimensions like gender and equity. Ofir (2017) succinctly captured the argument around the DAC criteria expansion, describing them as necessary but not sufficient to determine the worth, significance or merit of a development intervention. She further argued that some important aspects are absent, creating a false sense of confidence about the extent to which they can lead to a sound, defensible

synthesis judgement about contributions to development (Ofir, 2017). The work of Reeves and Baden (2000a:11) provides a satisfactory framework for integrating other essential components into the evaluation criteria. The framework advocates for the squaring of gender/equity indicators (ratio of access, participation, inclusion, non-discrimination and empowerment) within the five evaluation criteria.

The Fourth World Conference on Women, which took place in Beijing in September 1995, focused on Action for Equality, Development and Peace. The conference resolution urged countries to adopt strategies for mainstreaming gender perspectives into all aspect of development in order to close the gender gap. However, 23 years since then not much progress has been made. Globally, women earn less than men and are more likely to work in poor quality jobs. A third suffer physical or sexual violence in their lifetime. Gaps in reproductive rights and health care leave 800 women dying during childbirth each day (UNWomen, 2017).

In Africa, women and girls continue to face a bleak future as boys are, for example, still more likely to benefit from a secondary school education than girls. Furthermore, complications due to pregnancy or childbirth leaves African women with a 1 in 3 chance of dying, compared to a 1 in 4 300 chance in the developed world. Women's agency, movement and political voice remains restricted, sustained by customary belief and practices in most African countries (World Bank, 2016). While some significant progress was recorded in developed countries, most African countries are yet to show desirable progress in closing the gender and equity gap.

As African countries enter the fourth year of implementing plans, programmes and policies to achieve the Sustainable Development Goals, an accelerated strategy is required to advance progress, specifically on the transformative result of ending gender disparities, inequality and all forms of social injustice. The analysis of Engela and Ajam (2010) supports the evidence of a severe paucity of information around government owned and led evaluations in Africa, with most African countries lacking gender and equity requirements in their national evaluation policies. However, African countries like South Africa, Uganda and Benin have shown remarkable leadership but still have a significant proportion of their evaluations driven by donors and development partners.

There is the need to evaluate the focus and quality of evaluations from a gender and equity perspective to ensure that evaluation is serving a developmental function. If the current global inequality and exclusion gap is to be closed, gender and equity-focused evaluations need to be embraced holistically by government agencies, private institutions, universities and sundries. This approach underscores the commitment made by 193 countries in the wake of the 2030 agenda for sustainable development set in 2015. Specifically, goal 5 of the SDG advocates for gender equality and calls for the elimination of all forms of discrimination against women while goal 10 explicitly calls for the reduction in all forms of inequality. Similarly, the SDG targets place a strong emphasis on data disaggregation, i.e. providing data that distinguishes on grounds of gender, age and other relevant characteristics. Ending all forms of discrimination against women and girls is not only a basic human rights issue, but it is also a precursor for accelerating sustainable development. Evidence has it that the empowerment of women and girls has a multiplier effect and helps drive up economic growth and development across the board (Klasen, 2000).

Reiterating the criticality of gender and equity perspectives in development evaluation, an in-depth analysis of the 230 SDG indicator reveals that 32% of them are gender specific while 21% explicitly requires sex disaggregation. Twelve out of the seventeen goals speak to gender equality. Approximately 34% (58 out of 169 targets) implicitly address gender equality and empowerment of women and girls. Overall, the strong emphasis is on disaggregated data by sex and other relevant characteristics to capture the intersecting inequalities (Segone, 2013). Therefore, the principal focus of this study is to promote the correct and consistent use of gender and equity-focused evaluations to drive the achievement of sustainable and inclusive development.

## The case for equity and gender perspectives

Based on the aforementioned gaps and challenges in the introductory section, the account of Robert Picciotto in *The 5th wave* gives a more concise global picture of the expanding scourge of inequality and inequity driven by the unprecedented growth of the private sector in the global economy. By his account, large parts of sectors such as social services and infrastructure (which in the past usually was the reserve of government) is now dominated by the public-private partnership (Picciotto, 2015). This situation is more pronounced in African countries where systems and institutions have remained weak (UNDP, 2014). Similarly, *Forbes Magazine* (2015) aptly captured the extent

of the global inequality and inequity stating that the three richest people in the world own wealth that is equivalent to the combined GDP of the world's poorest 48 countries. Families, communities, companies and countries are all paying a tremendous price for this inequality, with the global gender pay gap sitting at a staggering 59% (Mahmud, 2018). Recent figures estimate that women are being paid less than men, and their concentration in the unpaid, minimum wage and insecure work costs the global economy $9 trillion a year (McCoy, 2017).

To clarify terms, "inequity" refers to the disparities between population groups that are avoidable and unfair, and which keeps a particular group of society in deprivation; "gender inequality" refers to the unequal treatment or perceptions based on gender, arising from differences in gender roles (UNICEF, 2013).

The account of Segone (2013) gives a glimpse into the prevalence of inequity in Africa, where about 90% of children with disabilities do not attend school. Furthermore, the majority of out-of-school children are girls (53%); thus, achieving gender parity would mean 3,6 million more girls in primary school (Segone, 2013). Wide and often mutually reinforcing inequalities are evident across developing countries, including inequality as per household wealth, rural/urban disparities, disability, indigenous people, ethnic minorities, those with migrant status, and gender (SDG Compass, 2015). Social equity and gender equality in human development are gaining the required support, more so with the SDG principle demanding that the furthest behind is reached first and that no one should be left behind. However, the resulting challenge is keeping the focus on achieving equitable development results through the evaluation function. The capacity of governments, civil societies and communities to evaluate the effectiveness of interventions on equitable outcomes for women and other marginalised populations remains the issue of concern.

Gender and equity-focused evaluations is an assessment of relevance effectiveness, efficiency, impact and sustainability of interventions on equitable development results (UNICEF, 2013). Gender and equity-focused evaluation employs a rigorous, systematic process in the design, analysis and interpretation of specific questions on what works or does not work for worst-off groups compared with the "better-off" groups, by providing evidence on what works in reducing inequality (Segone, 2013). Studies have shown that equity and gender-focused evaluations can be empowering for the worst-off groups as it provides evidence for advocacy, for holding duty-bearers accountable, and for demanding change (Round & Paulden, 2017).

This study adopted a set of criteria of UNEG to assess the equity and gender component of the evaluations reports in AfrED. Similarly, the argument of Reeves and Baden (2000b:11) was taken into cognisance, i.e. that a combined strategy of gender and equity-focus in programmes, and its integration into the evaluation criteria, is an efficient framework for evaluating social issues.

The evaluation practices and function are rapidly emerging in Africa, though largely donor-driven, necessitated by the need for accountability for the utilisation of donor funds (Rosenstein, 2013). Indeed, the funding and technical support from development partners have ushered in capacity and knowledge for the collection, analysis and use of evidence for decision-making in Africa. South Africa is one of the first African countries to develop and operationalise a national evaluation policy which is on par with international standards, with other African countries, like Uganda, Benin, Kenya and Ghana, achieving similar feats (DPME, 2014). It is worth mentioning, that the African Evaluation Association (AfrEA) established in 1999 has recorded giant strides, e.g.: (i) the hosting of the eighth African international conference on evaluation, (ii) the establishment of the *African Evaluation Journal* and the provision of oversight function to voluntary organisations for professional evaluation (VOPEs) in different African countries.

However, the analysis of Engela and Ajam (2010) reveals a severe paucity in government-led and solely funded gender and equity-focused evaluations in Africa. An enquiry is necessary to unravel the institutional root causes responsible for the over dependence by African countries on donor founded evaluations.

Given the above realisation, the study sought to:

1. Assess the extent to which evaluation reports published in AfrED (between 2005–2015) incorporated gender and equity dimensions in the evaluations conducted.
2. The factors (secondary or primary) hindering the inclusion/use of gender and equity in evaluations conducted in selected West African countries.
3. Lessons learned in practice and recommendations on how to overcome institutional challenges around the conduct and use of gender and equity-focused evaluation in different scenarios.

## The incorporation of gender and equity in evaluations

A critical first step to understand the extent to which evaluations are responding to the need for gender responsiveness is to describe the current practice in evaluation reports. To do this, the study was sectioned into three: (i) the desk review/content analysis of the evaluation reports, (ii) qualitative investigation on the causes of non-inclusion and poor utilisation of gender and equity-focused evaluations, and (iii) an appreciative enquiry on recommendations and lessons learned for the conduct of gender and equity focused evaluations.

The review of the content analysis of the evaluation reports was sub-divided into two broad types of evaluation: (i) 'Thematic evaluations' refers to evaluation reports with specific intent to evaluate gender and equity, and (ii) the "General evaluations" represents development work which may not be specifically designed to evaluate gender and equity components.

Emphasis was placed on the type of evaluation, evaluation design, evaluation focus, and evaluation findings and recommendations. This is because an evaluation type should be applied first to establish whether the evaluation report considered gender and equity dimensions in the evaluation process or not. The two categories of general and thematic evaluations were further sub-divided into four niches for the purpose of managing the analysis:

a. Project evaluations;
b. Clustered projects;
c. Policy evaluations; and
d. Evaluation of gender equity funds (see Tables A.1 and A.2 in the Annexure).

## Methods

For the first section of the study, a cluster sampling approach based on the aforementioned niches was used to select two evaluation articles each for the "General evaluation" category, and one evaluation report each under the "Thematic category". This was necessary because not all the evaluation reports in the database has a gender focus project, a necessary inclusion criterion for the thematic category. Twelve thematic evaluations and 24 general evaluation reports were selected from 492 evaluation reports captured in the AfrED database between 2015 and 2017.

For the second segment of the investigation (qualitative enquiry section), the study used a combination of purposeful sampling methods as conceptualised by Park and White (2017), which include purposeful sampling and sampling important expert groups. The "sampling of important expert groups" enabled the study to harvest relevant insight into the policy, practices and the political context hindering the utilisation of gender and equity components in evaluations from a donor and a policy maker's perspective. The study also examined the perspectives of respondents in national and sub-national settings, to uncover any variation in respondent's views based on location. The above sampling strategy examined 40 respondents, in line with the grounded theory methodology construed by Creswell (1998) and Morse (1994). The theory suggests that 30-50 interviews are enough to achieve theoretical saturation. Similarly, the sampling strategy proposed by Morse (2016) revealed that a combination of data collection approaches including key informant interviews, focus groups, and desktop reviews leads to higher qualitative study validity.

Two persons each from the stakeholders identified above using purposeful sampling were selected as construed by Goldstein and Reinert (1997). A respondent sample size of n=40 was adopted and more than ten years of evaluation experience regarded as an expert (on the premise that this period of cogent working experience in evaluation qualifies a respondent as such). Equally, the 'Delphi-type' key informant interview method was employed to identify the factors (primary or secondary) that hinder the implementation of gender and equity focus in Evaluation Delphi-technique as an appropriate method for reviewing plans and policies, especially when the judgement and opinions of experts are paramount (Hsu & Sandford, 2007).

## Data collection

The study used a semi-structured interview guide, containing a series of open-ended questions and probes. Similarly, key informant interviews (KIIs) was utilised to collect information from experts, policy makers and heads of non-governmental organisational with over ten years experience in managing projects pertinent to gender and equity focused evaluations. Questions in in-depth interviews (IDIs) focused on identifying the challenges and documented recommendations on how to improve the conduct of gender and equity-focused evaluations.

Similarly, the application of gender equality in selecting respondent was utilised to prevent over-representation from a sex category which could skew the study findings towards a specific gender perspective. The research questionnaire was streamlined to answer the following research questions:

1. What were the reasons for the low utilisation of the gender and equity dimension in evaluation?
2. What were the lessons learned and recommendations made for improving the use of the gender and equity dimensions in evaluations?

Four countries in the West African region were selected purposively. Nigeria and Ghana represented the Anglophone countries (English-speaking), while Benin and Togo represented the Francophone countries (French-speaking) respectively. Important stakeholders like VOPEs, policy makers, national evaluators, representative from donor organisations and heads of NGOs were selected. The MAXQDA analysis tool was the tool of choice because of the advantage it has in analysing open-ended survey questions over NVivo, ATLAS.ti, and QDA Miner. Besides, it can graphically display a weighted frequency analysis of issue-based themes and at the same time show the weighted responses of different respondents to questions, using an infographic display feature (Franzosi, Doyle, McClelland, Rankin & Vicari, 2013).

## Analytical framework

The research analytical framework was based on the integration model proposed by Reeves and Baden (2000b:11) founded on squaring gender/equity indicators (ratio of access, participation, inclusion, non-discrimination and empowerment) within the five evaluation criteria.

**Table 3.1** Integration Evaluation Model (gender and equity components in the evaluation criteria)

| Project/ Programme evaluation criteria | Mainstreaming of gender and equity dimensions/criteria | | | | |
|---|---|---|---|---|---|
| | Ratio of Access | Participation | Inclusion | Empowerment | Non-discrimination |
| Relevance | | | | | |
| Effectiveness | | | | | |
| Efficiency | | | | | |
| Impact | | | | | |
| Sustainability | | | | | |

*Sources:* Martha de Waal (2006); Reeves & Baden (2000b:11); Segone & Rugh (2013)

A comprehensive qualitative assessment was undertaken for each evaluation report selected. A data sheet was developed and tested using a sample of both "thematic" and "general" evaluations. The extent to which gender and equity were included in the evaluation report was assessed based on the parameters in the checklist below:

**Table 3.2** UNEG quality checklist for evaluation

| United Nations Evaluation Group quality checklist for evaluation | |
|---|---|
| Evaluation design | The report illustrates the extent to which the design and implementation of the objective, the assessment of results and the evaluation process incorporate equity and gender-based approach. |
| Critical questions | Is the project goal SMART and the beneficiaries identified?<br>Is the intervention logic or theory, including underlying assumptions and factors affecting the success of the intervention present? |
| Evaluation Indicators | The report uses equity and gender-sensitive language throughout, including data disaggregated by sex, age, disability, income, etc. |
| Critical questions | To what extent was equity and gender-sensitive indicators used?<br>Was the data collected disaggregated by sex and other parameters? |
| Evaluation analysis | The evaluation approach and data collection and analysis methods are responsive and appropriate for analysing the equity and gender equality issues identified in the scope. |
| Critical questions | Was there depth and quality in the equity and gender analysis conducted?<br>Did the analysis show the disparities (who is getting what by sex, location, etc.) |
| Evaluation findings | The report assesses if the design of the object was based on sound equity and gender analysis and implementation for results were monitored through equity and gender frameworks, as well as the actual results on equity and gender. |
| Critical questions | Did the findings show the impact to beneficiaries as planned?<br>Was attention given to how the resources were directed and spent? |
| Evaluation recommen-dations | Reported findings, conclusions, recommendations and lessons provide adequate information on equity/gender equality rights |
| Critical questions | Are the recommendations relevant and reinforcing gender and equity positions?<br>Are the recommendations specific and actionable with a focus on resources required to close the gender and equity gap? |

*Sources:* Adapted from UNEG (2016); European Commission (2010); CSW (2017)

Similarly, Table 3.3 gives the rationale for scoring the evaluation reports into four categories: (i) highly satisfactory, (ii) satisfactory, (iii) unsatisfactory or (iv) highly unsatisfactory, based on the predetermined criteria ranging from 0 (which represents highly unsatisfactory) to 5 (which represents highly satisfactory):

**Table 3.3** Scoring criteria and rating

| Evaluation design | Evaluation indicators | Evaluation analysis | Evaluation findings | Evaluation recommendations | Rating | Criteria |
|---|---|---|---|---|---|---|
| 4 | 4 | 4 | 4 | 4 | Highly Satisfactory | Up to four or more coherent G&E statements in each Evaluation component. |
| 3 | 3 | 3 | 3 | 3 | Satisfactory | Up to three but not less than 2 coherent G&E statements in each Evaluation component. |
| 1 | 1 | 1 | 1 | 1 | Unsatis-factory | At least one coherent statement in any two of the Evaluation components. |
| 0 | 0 | 0 | 0 | 0 | Highly unsatis-factory | No statement on G&E in the Evaluation component. |

*Source:* Adapted from UNEG (2016)

## Findings on general evaluations

The findings in Table 3.4 presents the overall analysis of the extent gender and equity were included in the "general evaluation" category. A total of 24 project evaluation reports were assessed in the general category. The country report codes (NG2, ZM1, RW2, etc.) represent the evaluation reports selected from the AfrED (see Table A.1 in the Annexure for a full description of reports).

**Table 3.4** Overall analysis for general evaluation category

| Highly Satisfactory attention to gender and equity issues | Satisfactory attention to gender and equity issues | Unsatisfactory attention to gender and equity issues | Highly Unsatisfactory attention to gender and equity issues |
|---|---|---|---|
| | NG2<br>RW1<br>BW1<br>TZ2<br>ZA1 | NA1<br>ZW2<br>NG1<br>ZM1<br>RW2<br>BW2<br>UG2<br>GH2 | NA2<br>ZW1<br>ZM2<br>TZ1<br>UG1<br>GH1<br>ZA2<br>KE1<br>KE2<br>ET1<br>ET2 |
| 0 | 5 | 8 | 11 |

1. Eleven evaluations (45%) are highly unsatisfactory in their treatment of gender and equity issues. The evaluation reports did not examine the benefits or impact on women, neither did they assess women's participation. In most cases, little or no data was presented as evidence in the analysis although it was the intention In some of the evaluation reports reviewed, the agency policy commitment to gender equality and equity was described and recommendations made that more gender or equity-related analysis was needed. The evaluation design, indicators, questionnaire, findings and recommendations did not consider gender and equity issues.

2. Eight evaluations (33%) were rated as having an unsatisfactory gender and equity component in their analysis. The evaluation reports measured in a limited way the participation of socially disadvantaged groups, including women, in project or programme activities. There was no clear information on how the socially marginalised groups, including women, have benefitted. Nor was there any statements about the impact of the project on women or

any other category of socially excluded individuals. The evaluation design, indicators, questionnaire, findings, and recommendations had very weak and incoherent references to gender and equity issues.

3. Five evaluations (20%) were rated as having a satisfactory gender analysis. The reviewed reports measured the participation of women in activities and provided at least two pieces of information on how the disadvantaged groups, including women, have benefitted from programme or project activities. It also gave at least two accounts on the impact of activities on the socially disadvantaged groups, including women. The evaluation design, indicators, questionnaire, findings and recommendations included clear gender and equity dimensions. A few of the evaluation reports captured in detail an assessment of changes in gender relations.

In conclusion, none of the evaluation reports reviewed under the "general category" scored a highly satisfactory rating (the gold standard) of having up to four or more coherent gender and equality statements in each of the evaluation components. However, it can be understood that these projects (under the general category) did not include gender and equity dimensions in their design from the unset as they were not designed to deal with gender or equity issues.

## Findings on thematic evaluations

The findings from Table 3.5 presents an overall analysis of the extent to which gender and equity were included in the 12 "thematic evaluation" reports reviewed. The country report codes (UG0, ZAO, TZ0, etc.) represent the evaluation reports selected from the AfrED (see Table A.1 in the Annexure for a full description of reports).

**Table 3.5** Overall analysis for thematic evaluation category

| Highly Satisfactory attention to gender and equity issues | Satisfactory attention to gender and equity issues | Unsatisfactory attention to gender and equity issues | Highly Unsatisfactory attention to gender and equity issues |
|---|---|---|---|
| UG0<br>NG0<br>RW0 | ZA0<br>GH0<br>ET0<br>KE0<br>ZM0<br>BW0 | TZ0<br>ZW0<br>NA0 | |
| 3 | 6 | 3 | 0 |

1. Three evaluations (25%) were rated as having an unsatisfactory gender and equity component in their analysis, though a gender specific intervention. The evaluation reports assessed the participation of socially disadvantaged groups, including women, in project or programme activities in a limited way. There was no clear distinction of information provided on how the socially marginalised groups including women have benefitted, nor was there any statements about the impact of the policy on women or any other category of socially excluded persons. The evaluation design, indicators, questionnaire, findings and recommendations had a weak statement on gender and equity issues.

2. Six evaluations (50%) were rated as having a satisfactory equity/gender analysis. The reports measured the participation of women and the targeted excluded population in project implementation activities. The reports also provide at least two pieces of information on how the disadvantaged groups, including women, have benefitted from programme or project activities and gave two accounts on the impact of activities on the socially disadvantaged groups, including women. The evaluation design, indicators, questionnaire, findings and recommendations had clear gender and equity dimensions. A few of the evaluation reports captured in detail an assessment of changes in gender relations.

3. Two evaluations (25%) were rated as having a highly satisfactory gender analysis. These reports gave a good account of the participation of women and the targeted excluded population in the project implementation activities. The project possessed a clear logic intervention (theory of change). The reports also provide more than three items of information on how the disadvantaged groups, including women, have benefitted from programme or project activities and give more than three accounts on the impact of activities on the socially disadvantaged groups, including women. The evaluation design, indicators, questionnaire, findings and recommendations had clear gender and equity dimensions. The evaluation reports gave a detailed analysis of changes in gender relations, analysis of the inequality differential and a relevant recommendation for improvement.

In conclusion, it is rather improper and unacceptable for projects and programmes specifically designed to close the gender and equity gap to score less that average in the gender and equity focus of such evaluation reports.

This calls for more in-depth research on why a commissioned equity/gender focused projects can deviate and fail to achieve their purpose.

## Understanding the exclusion of gender and equity

Several factors ranging from capacity, competence to policy considerations present as determinants influencing the non-inclusion and use of gender and equity perspectives in evaluations. An analysis of the appreciative enquiry conducted amongst ten different stakeholders from different backgrounds, sex, expertise and experiences (policy makers, project implementors, evaluators and donors, etc.) revealed eleven broad themes which encapsulated the reason (primary and secondary) why equity and gender perspectives in evaluations were not included or used in the selected African countries. Table 3.6 highlights the eleven reasons why gender/equity dimensions in evaluations are not conducted or in low demand.

**Table 3.6** Reasons for non-/low utilisation of G&E-focused evaluations

| S/N | Reasons for non-/low utilisation of gender and equity-focused evaluations |
|---|---|
| 1 | Not a mandatory requirement in the national evaluation policy |
| 2 | Low capacity to conduct G&E-focused evaluation |
| 3 | Lack of funds to conduct G&E-focused evaluation |
| 4 | Time-consuming and difficult to measure |
| 5 | Low demand for G&E-focused evaluations |
| 6 | Donors not doing enough (funding + demand) for G&E-focused evaluation |
| 7 | The concept is not well understood |
| 8 | Lack of disaggregated data by sex, location, income, etc. not readily available |
| 9 | Non-inclusion of gender and equity component in the evaluation terms of reference |
| 10 | The project/programme is not about gender and equity so there is no need for G&E-focused evaluation |
| 11 | Political and religious issues around gender and equity perspectives |

Similarly, Table 3.7 shows a combined frequency analysis (KII and IDI) from each stakeholder's view of what constitutes the primary or secondary reasons for the non-/low utilisation of G&E dimensions in evaluations.

**Table 3.7** Reasons for poor utilisation of gender and equity perspectives in evaluations

Legend:

Primary reason for low G&E-focused evaluation

Secondary reasons for low G&E-focused evaluation

| # Reasons for poor utilisation of Gender & Equity perspectives in evaluations | Policy Makers (National level) Female | Policy Makers (sub-National level) Male | VOPEs (National level) Female | VOPEs (sub-National level) Male | Evaluation experts (10+ years' experience (Male) | Evaluation experts (10+ years' experience (Female) | A representative from donor organisation (Male) | A representative from donor organisation (Female) | Head of NGO with 10+ years' experience in managing project (Male) | Head of NGO with 10+ years' experience in managing project (Female) | # Respondents |
|---|---|---|---|---|---|---|---|---|---|---|---|
| #1. Not a requirement in the National Evaluation Policy | | | | | | | | | | | 40 |
| #2. Low capacity to conduct G&E-focused evaluation | | | | | | | | | | | 40 |
| #3. Lack of funds to conduct G&E-focused evaluation | | | | | | | | | | | 40 |
| #4. Time consuming and difficult to measure | | | | | | | | | | | 40 |
| #5. Low demand for G&E-focused evaluations | | | | | | | | | | | 40 |
| #6. Donors not doing enough (funding and awareness) of G&E evaluation | | | | | | | | | | | 40 |
| #7. The concept is not well understood | | | | | | | | | | | 40 |
| #8. Disaggregated data by sex, location, income, etc. is not readily available | | | | | | | | | | | 40 |
| #9. Non-inclusion of G&E in the evaluation terms of reference | | | | | | | | | | | 40 |
| #10. The project/programme is not about gender and equity so there is no need for G&E evaluation. | | | | | | | | | | | 40 |
| #11. Political and religious issues around G&E perspectives | | | | | | | | | | | 40 |
| # Primary cause | 4 | 4 | 6 | 6 | 5 | 5 | 6 | 6 | 7 | 7 | |
| # Secondary cause | 7 | 7 | 5 | 5 | 6 | 6 | 5 | 5 | 4 | 4 | |

## Primary causes limiting use

From the frequency analysis above, all respondents considered the lack of policy requirement mandating the use of G&E focus in evaluations, the low capacity to conduct G&E-focused evaluation and the lack of funds as the primary reason for low G&E-focused evaluations. The next compelling primary reason highest amongst respondents, accounting for the poor utilisation of gender and equity perspectives in an evaluation, is the low demand for G&E evaluations. Below are some selected passages, revealing a more in-depth perspective of different stakeholders around the primary reasons for the non-inclusion and poor use of gender and equity perspectives in evaluation:

> Gender and equity evaluations are time-consuming and difficult to measure. Considering conducting a gender and equity evaluation at the middle of a project or program is a difficult thing to do. (IDI with VOPEs at national level, Togo)

> From my experience, end-of-project evaluations are not that difficult especially when you are not focusing on gender and equity issues. Gender and equity data are usually not readily available. (IDI with VOPEs at sub-national level, Ghana)

> Not many projects request a gender or equity dimension in the evaluation and you as an evaluator are strictly bound by the terms of reference of the job. It is not professional to add G&E perspectives in an evaluation when it is not required. (KII with male evaluation expert with 10 plus years' experience, Benin)

> The capacity to conduct gender and equity focused evaluations should be considered as the first primary cause. It is a worrisome challenge for Africa. Very few African universities have a comprehensive curriculum that strengthens the knowledge and practice of gender and equity focused evaluations. (KII with female representative from donor organisation)

## Additional factors limiting use

From the frequency analysis above, all respondents except for VOPEs (male and female) consistently agreed that a major secondary cause for low gender and equity evaluation is that it is time-consuming and difficult to measure. Second, all respondents except for heads of NGOs (male and female) believed that the non-inclusion of G&E in the evaluation terms of reference is a secondary causative reason. Third, all respondents, with the exception of the voluntary organisations of professional evaluators (VOPEs), agreed that projects and programmes not focused on achieving gender and equity evaluations from the start, should be required to use a lens of gender and equity in their evaluation.

Lastly, all respondents, excluding evaluators with over ten years' experience, agreed that the lack of readily available data, disaggregated into sex, location, income, etc., constitutes a secondary cause for the poor utilisation of gender and equity perspectives in evaluation.

The selected passages below reveal an in-depth perspective of some stakeholders regarding secondary reasons for the poor utilisation of gender and equity perspectives in evaluation:

> The concept and ideology of gender and equity-focused evaluation is still an emerging concept, though not entirely new. The call for it became stronger with the introduction of the Millennium Development Goals (MDGs) in 2000. The concept still needs a lot of sensitization and awareness raising amongst the civil society, evaluators, academia, etc. (KII with heads of NGOs, Benin)

> For evaluations that require gender and equity perspectives, the absence of disaggregated data by sex, age, income and location, etc. can make the evaluation process impossible. This is often the case in many projects. Even when there is a routine system capturing disaggregated data by sex etc. it is of poor quality. (IDI with VOPEs at national level, Togo)

Evaluators are not magicians; we do not manufacture data. We evaluate based on the evidence provided (IDI with VOPEs, Togo). Stakeholders from Togo and Benin reaffirmed the need for further sensitisation and awareness on the concept of gender and equity and the need to have a routine system for capturing disaggregated data.

> There are more pressing issues the government if trying to tackle now, for instance corruption, poverty reduction, and humanitarian issues. Gender and equity issues can be addressed later, as it is not so urgent. (KII with policy makers at national level, Nigeria)

> How religion influences the unequal power relation amongst sexes in the African society is underestimated. The push for gender equality is been slowed down, even by women groups due to religious and cultural beliefs. There are no gaining grounds without the active participation of our religious and traditional gatekeepers. (KII with policy makers at sub-national level, Ghana)

> Though a lot has been achieved over time, the concept of gender equality and equity is still being viewed as a western agenda even by highly educated top government personalities. Though public declarations are made in favour of gender equality, the political will to follow-up on these commitments is weak. (KII with a female representative of donor group, Nigeria)

The perspectives from Nigeria and Ghana reveal that making gender and equity government priority is critical and that religious and cultural norms still remain the single most important determinant to overcome.

In summary, eleven distinct yet interconnected reasons were identified for the poor utilisation of gender and equity perspectives in evaluations. Respondents categorised differently the reasons for the non or poor utilisation of gender and equity issues in evaluation. However, there was full agreement from all the respondents regarding the primary reasons for low gender and equity focus in evaluations, namely (i) the lack of policy requirement mandating this focus; (ii) the low capacity to conduct these focused evaluations; and (iii) the lack of funds. This explains the institutional weakness present in the selected African countries examined. The context in which gender and equity-focused evaluations are conducted, the perception of different stakeholders on why it is commissioned are important factors that should not be overlooked. Other variables of importance are the political will, cost and legal issues.

## Learning from practice – challenges institutionalising a gender and equity focus

The previous section identified the interconnected factors that influence the successful implementation of gender and equity perspectives in current evaluation practice. The review of what works and lessons learned (including good practices in different scenarios) is essential, particularly for most African countries operating in resource-constrained settings with weak systems and different institutional capacities. This section offers practical solutions on what to do when faced with five of the scenarios as described below.

### *Scenario 1 Government non-supportive of gender and equity-based evaluations*

In a situation where the government opposes or is not favourably disposed to gender and equity-focused evaluations, it is important for evaluators to attempt to understand the cause of the problem or what is driving the resistance. The resistance may be rooted in religious perspectives that refuse to acknowledge women as equal to men in all ramification, though the country's constitution may support gender equality. Sometimes there is a deliberate lack of support of a certain group for not voting for the incumbent government during the election, or it may be that the government wants to restrict the movement of

a certain group from a specific region (i.e. migration issues). Several legitimate or illegitimate reasons may underscore why a government may not support gender and equity-focused evaluations. Be that as it may, evaluators need to be committed to their professional ethos to ensure that evaluation reports reflect the true situation. These issues are difficult to address as they are multi-faceted (financial, political, religious, etc.). As a way forward, evaluators can:

1. Partner with other donor and development agencies who support social change and transformation. Identify areas were gender and equity interventions are likely to be relatively non-confrontational and can produce some acceptable gains that are immediate. For example, the provision of subsidised transportation for pregnant women can increase the number of women from poor communities that access antenatal care. The evaluation of such interventions could show a positive benefit and will find acceptance with government (Bamberger, 2011).
2. Gaining acceptability from a resistant government by starting with non-confrontational findings is a good starting point. For example, by choosing to focus on the identification of cost-effective evaluations using the bottleneck supply and demand-side approach to identify potential areas that can easily be resolved. A good starting point is to examine factors affecting the ability of low-income households, and particularly mothers, to access services. Often this will identify issues such as inconvenient opening hours or locations, long hours of patient waiting time, lack of transport, etc., that could be addressed relatively easily.
3. The government can be subtly engaged through high-powered advocacy led by international NGOs. A reasonable shift in favour of women's rights can be demanded, for instance, a concerted approach by the monitoring and evaluation technical working groups to ensure that uncontroversial components of gender and equity are mainstreamed in the national evaluation checklist for the collection of gender and equity-related information.
4. The confrontational approach, though not the best option, also has its place. This is especially the case when the generated equity/gender data gap (evidence) is coming from the government-led statistical bureau. Speaking the truth to power using hard evidence through credible champions can be effective.

### *Scenario 2 Project or programme to be evaluated lacks or has no comprehensive theory of change or logical framework*

First, a proxy theory of change can be reconstructed for a project without one. This can be achieved through a strategic discussion with the relevant stakeholders on the goal the project or programme was designed to achieve. This approach is based on the thoughts of the project designers and actors on what the project was meant to deliver and the pathway for the desired change to be achieved. In retrospect, it requires reflecting on the agreed outcome and how it should be done.

Second, a review of strategic documents, observation of the decision-making process, and consultation with relevant managers is also an effective method (Chen, 2005). The observational approach on how the project is implemented, what is critical and what is not in the project implementation, is explained to new clients. Using this approach allows you to decipher the difference between what managers say and what is observed in the actual field implementation. The observation of those who benefit the most from the project is the correct indication of the level of equity in the programme and not what was claimed.

Third, a new theory of change can be developed or the existing one enhanced by drawing on the knowledge of the programme executors. This can be done either through a forward-vision look at the intervention and proposing the desired outcomes, or through a backward visioning, i.e. by focusing on the desired outcome and looking backwards at the intervening factors and the identification of the critical determinants and pathway for change.

Lastly, a credible reality test to confirm the extent of the project's focus on gender and equity issues is by reviewing three critical issues. The first is the extent to which resource allocation reflects stated programmatic priorities. The second is identifying the extent to which information is available; a lack of data can reflect that an issue has not been prioritised. Finally, allocation of staff time and reflect programmatic focus. If an issue area is a stated priority, but staff time has not been meaningfully dedicated to it, this may reflect true programmatic prioritisation (Weiss, 2001).

### *Scenario 3 Project or programme to be evaluated is time and budget constrained*

Evaluation design can be enhanced and result delivered by simplifying the design and prioritising the essential component of the data-collection points without compromising the validity. First, by using any available secondary data for baseline and as a comparison group. Second, the information to be collected can be simplified to the precise and relevant ones bearing in mind what the stakeholders and project owners consider vital credible evidence (e.g. scale of sample size, focus group discussions with beneficiaries, and case studies) The understanding of the different perspectives of stakeholders on what is credible will determine the type of information that ought to be collected despite fund constraints (Mabry, 2012).

Third, the cost of data collection can be shared with other organisations that have similar interest, through the ART of networking and partnership-building. With a good negotiation, some agencies may be willing to support a similar household survey by adding a few additional questions to their planned household survey. In addressing time constraints posed by the budget, etc., reduction in the amount of data to be collected and analysis can significantly reduce the budget and time constraint, bearing in mind that not all that can be counted counts. Skyping and video conferencing are reliable means of saving cost, especially during data collection and analysis and planning phases.

Household data collection can be costly and expensive when hiring experienced and seasoned data collectors. However, the use of less expensive data collectors that have general competence in the survey subject matter will make a good consideration, for example, the use of certified nurses for medically related surveys can lower the cost of hiring commercial researchers. The use of direct observation can be a more valid data collection method than the application of surveys, depending on the research needs. For example, the observation of patients arriving at a health facility can be a more reliable way of determining means of transportation to a health facility than conducting a general survey to determine this.

Significant cost can also be saved using technology and modern survey methods such as using of focus group discussions instead of individual interviews. Likewise, the use of data through cell phones (GPS, Internet-based surveys) is a faster and cheaper way of collecting and analysing bulk information at the same time.

### *Scenario 4 Substantial baseline information is not available to guide gender and equity-focused evaluations*

The absence or insufficient baseline data in a project hampers the estimation of projects and programmes. Different scenarios play out that lead to the absence of a project baseline, for example, (i) a baseline may have been conducted but did not include a gender and equity dimension; (ii) a project may have been in operation for a while before a baseline was conducted; (iii) poor quality of baseline conducted; and (iv) planned baselines that were never conducted.

Any of the scenarios listed above can be salvaged by reconstructing a new baseline based on proxy measures for the gender and equity component of the evaluation. First, data from other organisations implementing similar projects

can be used to construct a comparison group, for example, data on school enrolment and health facilities. Second, baseline data can be determined from the routine data collected from the project, including administrative data. Third, data from the national datasets, for example, demographic health or multiple indicator health surveys) can also suffice. Fourth, the recall strategy can be an effective way of determining baseline data. However, it is subject to a high level of respondent bias and is very difficult to correctly adjust for the magnitude of bias. The technique requires, for example, that respondents reflect on their household income and expenditure or on the number of pregnant women who visited traditional birth attendants before a new health facility was constructed. More specifically, techniques like focus group discussions, key informant interviews and participatory rapid appraisal (PRA) are effective techniques to estimate baselines for underserved populations. However, it is difficult to determine bias because of loss of memory, intentional distortions, etc., since recall is the only source of information on the past.

The PRA approach is helpful in identifying the social categories through classifications like wealth quintiles or ranking, and for social mapping. It can also assist to determine the social map of a community by creating a standard for ranking a household based on certain criteria, for example, ranking households with a motorcycle and television as 'rich' and those without it as 'poor', or using other vulnerability characteristics (Bamberger, 2010).

### *Scenario 5 Poor demand and use of evaluation reports (a systemic approach)*

The need to capacitate and empower civil society to demands that public policies and programmes are designed and implemented based on evidence cannot be overemphasised. Equity issues are human right issues and probably requires class actions to effect the needed change. The use of incentives can inspire managers to demand and use gender and equity perspectives in evaluations. Governments could also use sanctions for those who do not comply.

It is imperative for governments to invest in the required infrastructure for sustaining the monitoring and evaluation systems. Long-term training and capacity building (informed by a thorough capacity-gap assessment) are required. This strategy will, in the long run, be effective in building the needed political and financial support for sustaining the national M&E system and

structures. Similarly, strengthening the gender and equity indicators in the national data collection system is an important consensus to achieve in the periodic review of the national indicators. This strategy invariably guarantees the generation of important equity analysis like multidimensional poverty trend analysis, and benefit incidence analysis, which are essential for decision-making.

## Conclusion

The use of the integrated evaluation model is effective for assessing the gender and equity focus of evaluations, through a checklist that integrates the gender/equity indicators (ratio of access, participation, inclusion, non-discrimination and empowerment) within the five evaluation criteria. The content analysis of 36 evaluation articles (guided by a set of quality criteria), showed that under the "general report category" (evaluation of projects not focused to address gender or equity) only 20% of the evaluation reports reviewed were classified as satisfactory, 33% unsatisfactory and 45% highly unsatisfactory in their gender and equity focus. Similarly, under the "thematic report category" (evaluation of reports focused on addressing gender and equity) the review found 25% of the evaluation reports highly satisfactory, 50% satisfactory and 25% unsatisfactory with regards to the gender an equity focus. This landscape may change if commissioners of evaluations would prioritise a gender and equity focus in their evaluations.

From the qualitative appreciative enquiry section, it is clear that more needs to be done by African countries towards the inclusion, institutionalise and use of equity and gender perspectives in routine evaluations. First, the need to develop a comprehensive national evaluation policy that incorporates gender and equity perspectives, and the enforcement of such policies to assess the gender and equity outcomes in all commissioned evaluations. Second, the need for governments to provide the leadership, human capacity and funding (including sanctions or incentives) to ensure compliance at all levels. Third, the need to conduct quality checks and reviews on all project evaluations to ensure they attain a good rating through coherent incorporation of equity and gender perspectives in the design, indicators, findings and recommendations, regardless of whether the project or programme was focused on closing the gender or equity gap or not.

The inclusion of gender and equity dimensions in evaluation is not an end, but a means. The end is about achieving equitable development results for all citizens, with a focus on the improvement of the quality of lives of the excluded and the underserved populations. The dream of achieving the SDG agenda by 2030 will be an impossible dream for most African countries, without the timely provision of evidence to inform policies, programmes and plans directed at reducing the gender and equity gaps.

There is a need for African countries to heed the call and address all systemic issues identified in this study that undermine progress. Frameworks and policies that include, consistently track, and use results of gender and equity-focused evaluations would help African countries be on track for achieving the SDGs by 2030. This approach supports the equitable delivery of projects and programmes in compliance with the SDG approach of leaving no one behind, and the call to reach the furthest behind first.

## References


Bamberger, M. 2010. *Reconstructing baseline data for impact evaluation and results in measurement*. No. 4. The Nuts and Bolts of M&E Series. Poverty Reduction and Equity Department. The World Bank. http://siteresources.worldbank.org/ INTPOVERTY/Resources/335642-1276521901256/premnoteME4.pdf

Bamberger, M. & Segone, M. 2011. *How to design and manage equity-focused evaluations*. New York: UNICEF Evaluation Office.

Bustelo, M. 2003. Evaluation of gender mainstreaming: Ideas from a meta-evaluation study. *Evaluation*, 9(4):383-403. https://doi.org/10.1177/1356389003094002

Chen, H. 2005. *Practical Program Evaluation: Assessing and improving planning, implementation, and effectiveness.* London: SAGE.

Creswell, J.W. 1998. *Qualitative inquiry and research design: Choosing among five traditions*. London: SAGE.

CSW (United Nations Commission on the Status of Women). 2017. *Toolkit on mainstreaming gender equality*. New York: CSW.

Denford, S.; Lakshman, R.; Callaghan, M. & Abraham, C. 2018. Improving public health evaluation: a qualitative investigation of practitioners' needs. *BMC Public Health*, 18(1):190. https://doi.org/10.1186/s12889-018-5075-8

DPME (Department of Planning, Monitoring and Evaluation). 2014. *Strategic Plan 2015–2020*. Pretoria: Government Printing Works.

Engela, R. & Ajam, T. 2010. *Implementing a government-wide monitoring and evaluation system in South Africa*. Washington, DC: The World Bank.

European Commission. 2016. Quality Assessment for Evaluation Reports. Brussels. https://ec.europa.eu/transport/sites/transport/files/facts-fundings/evaluations/doc/2016-03-16-evaluation-study-application-cross-border-quality-assessment.pdf [Accessed 9 June 2017].

*Forbes Magazine*. 2015. The World's Billionaires. 9 March.

Franzosi, R.; Doyle, S.; McClelland, L.E.; Rankin, C.P. & Vicari, S. 2013. Quantitative narrative analysis software options compared: PC-ACE and CAQDAS (ATLAS.ti, MAXqda, and NVivo). *Quality & Quantity*, 47(6):3219-3247. https://doi.org/10.1007/s11135-012-9714-3

Hsu, C.C. & Sandford, B.A. 2007. The Delphi technique: Making sense of consensus. *Practical Assessment, Research & Evaluation*, 12(10):1-8.

Jost, C.; Kyazze, F.; Naab, J.; Neelormi, S.; Kinyangi, J.; Zougmore, R.; Aggarwal, P.; Bhatta, G.; Chaudhury, M.; Tapio-Bistrom, M.L. & Nelson, S. 2016. Understanding gender dimensions of agriculture and climate change in smallholder farming communities. *Climate and Development*, 8(2):133-144. https://doi.org/10.1080/17565529.2015.1050978

Kabeer, N. 2003. *Gender mainstreaming in poverty eradication and the millennium development goals: A handbook for policy-makers and other stakeholders.* London: Commonwealth Secretariat. https://doi.org/10.14217/9781848598133-en

Klasen, S. 2000. *Does gender inequality reduce growth and development? Evidence from cross-country regressions.* Sonderforschungsbereich 386, Discussion Paper 212.

Lincoln, Y.S. 2003. Constructivist knowing, participatory ethics and responsive evaluation: A model for the 21st century. In: D.L. Stufflebeam & T. Kellaghan (eds), *International handbook of educational evaluation.* Netherlands: Springer. 69-78. https://doi.org/10.1007/978-94-010-0309-4_6

Mabry, R.M. 2012. Addressing physical inactivity in Omani adults: Perceptions of public health managers. *Public Health Nutrition*, 17(3):674-681. https://doi.org/10.1017/S1368980012005678

Mackay, K. 2007. *How to build M&E systems to support better government.* Washington: The World Bank. https://doi.org/10.1596/978-0-8213-7191-6

Mahmud, R.L. 2018. Unlocking American economic potential: Bridging the gender wage gap for women of all races and ethnicities. MA in Public Management dissertation, Johns Hopkins University.

Maramba, P. & Bamberger, M. 2001. A gender responsive monitoring and evaluation system for rural travel and transport programs in Africa. SSATP (Sub-Saharan Africa Transport Policy Program, the World Bank and Economic Commission for Africa) Working Paper, 55.

McCoy, R. 2017. *World view of women in leadership – A discussion of global gender parity in business and government with a review and recommendations of improvement opportunities.* Management Seminar Paper: Ligs University.

Morse, J.M. 1994. Designing funded qualitative research. In: N.K Denzin & Y.S. Lincoln (eds), *Handbook of Qualitative Inquiry*. Thousand Oaks: SAGE. 220-235.

Morse, J.M. 2016. *Mixed method design: Principles and procedures.* New York: Routledge. https://doi.org/10.4324/9781315424538

Ofir, Z. 2017. Evaluation for development focus on Global South: http://zendaofir.com/updating-dac-evaluation-criteria-part-5/ [Accessed 15 January 2018].

Oxfam International. 2015. Press release. oxfam.org/en/pressroom/pressreleases/2015-01-19/richest-1-will-own-more-all-rest-2016

Park, C.M.Y. & White, B. 2017. Gender and generation in Southeast Asian agro-commodity booms. *The Journal of Peasant Studies*, 44(6):1103-1110. https://doi.org/10.1080/03066150.2017.1393802

Picciotto, R. 2015. The 5th wave: Social impact evaluation. Working paper. New York: Rockefeller Foundation.

Reeves, H. & Baden, S. 2000a. *Gender and development: Concepts and definitions* Bridge Report No. 55). Brighton: Institute of Development Studies. https://www.bridge.ids.ac.uk/reports/re55.pdf

Reeves, H. & Baden, S. 2000b. *Gender and development: Frequently asked questions*. Bridge Report No. 57. Brighton: Institute of Development Studies. https://www.bridge.ids.ac.uk/reports/re57.pdf

Rosenstein, B. 2013. Mapping the status of national evaluation policies. Parliamentarians Forum on Development Evaluation in South Asia and EvalPartners. Unpublished consultancy report.

Round, J. & Paulden, M. 2017. Incorporating equity in economic evaluations: A multi-attribute equity state approach. *The European Journal of Health Economics*, 19(4):489-498. https://doi.org/10.1007/s10198-017-0897-3

SDG Compass. 2015. *The guide for business action on the SDGs*. Geneva, Switzerland: World Business Council for Sustainable Development (WBCSD). https://sdgcompass.org

Segone, M. & Rugh, J. 2013. *Evaluation and civil society: Stakeholders' perspectives on national evaluation capacity development*. New York: UNICEF, IOCE and EvalPartners.

UNDP (United Nations Development Programme). 2014. *Human development report 2014: Sustaining human progress: Reducing vulnerabilities and building resilience*. New York: United Nations.

UNEG (United Nations Evaluation Group). 2016. *2005 Standards for evaluation in the UN System*. http://www.unevaluation.org/document/detail/22 [Accessed 12 January 2018].

UNICEF (United Nations International Children's Emergency Fund). 2011. *Equity and reaching the most marginalized: Selected innovations and lessons from UNICEF programmes*. New York: UNICEF.

Weiss, C.F. 2001. Theory based evaluation: Theories of change for poverty reduction programs In: O. Feinstein & R. Piccitto (eds), *Evaluation and Poverty Reduction*. New Brunswick: Transaction. 103-114. https://doi.org/10.4324/9781351325325-16

# ANNEXURE

**Table A.1** The type of evaluation reports selected under "thematic evaluation" category

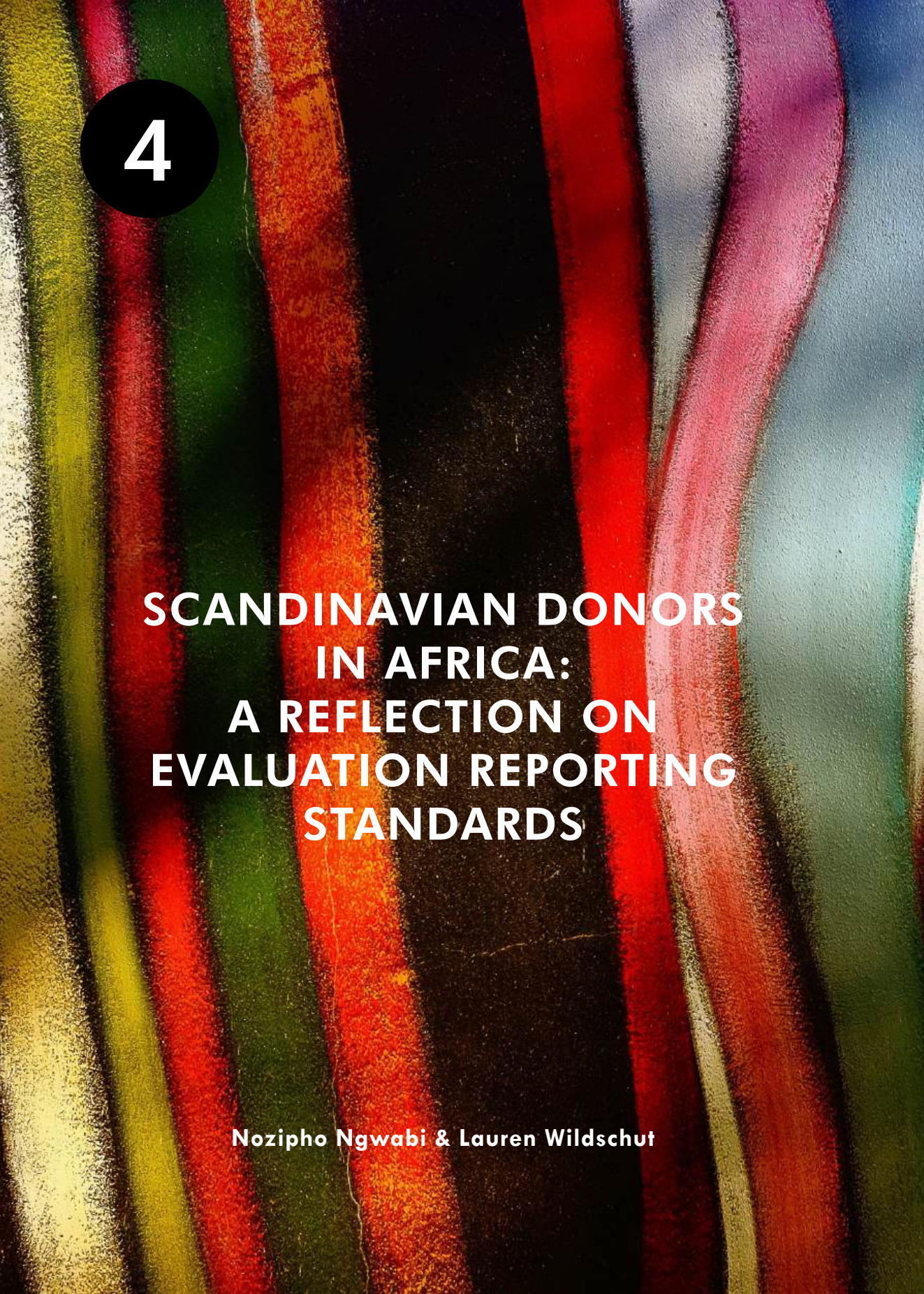
| S/N | Country | Thematic: title of evaluation | Code | Donor support | Category of evaluation |
|---|---|---|---|---|---|
| 1 | Ethiopia | Changing Gender Norms & Reducing Intimate Partner Violence (2015) | ETO | 3ie | Project |
| 2 | Kenya | Changing Gender Norms and Reducing Intimate Partner Violence | KEO | 3ie | Project |
| 3 | South Africa | Evaluation Report of the technogirl Programme in South Africa | ZAO | UNICEF | Project |
| 4 | Ghana | Sustainable Fisheries Management Project (SFMP) Gender Needs Assessment Report | GHO | USAID | Project |
| 5 | Uganda | Govt of Uganda, 2015. Water and Sanitation for the Poor and Bottom 40% – Since 2006 | UGO | World bank | Cluster |
| 6 | Tanzania | The impact of Tanzania's Joint Forest Management Programme on Livelihoods, Governance and Forests | TZO | Sida | Cluster |
| 7 | Botswana | Botswana Performance Evaluation – Tsela Kgopo Gender and OVC Project Final Report | BWO | USAID | Project |
| 8 | Rwanda | Evaluation of HED/USAID Women's Leadership programme, Rwanda (2012–2015) | RWO | USAID | Cluster |
| 9 | Zambia | 2016 Midterm evaluation of stamping out and previous gender-based violence (GBV), Zambia | ZMO | USAID | Project |
| 10 | Nigeria | Explaining Gender Differentials in Agricultural Production in Nigeria | NGO | World bank | Policy |

| S/N | Country | Thematic: title of evaluation | Code | Donor support | Category of evaluation |
|---|---|---|---|---|---|
| 11 | Zimbabwe | Evaluation of Support to Zimbabwe under ZUNDAF (2012–2015) Enhanced Economic Management and Pro-poor Development Policies and Strategies | ZWO | UNDP | Policy / Cluster |
| 12 | Namibia | Effects of Community-based Natural Resource Management on Household Welfare in Namibia | NAO | 3ie | Project |

**Table A.2** The type of the evaluation reports selected under "general evaluation" category.

| S/N | Country | General evaluation: Title of evaluation | Code | Donor support | Category of evaluation |
|---|---|---|---|---|---|
| 1 | Ethiopia | USAID Ethiopia 2016 Ethiopia Strengthening Land Tenure and Administration | ET1 | USAID | Cluster |
| | | Opening Doors: A Performance Evaluation of the Development Credit Authority (DCA) in Ethiopia | ET2 | USAID | Cluster |
| 2 | Kenya | Peace Initiative Kenya: Final Performance Evaluation | KE1 | USAID | Project |
| | | USAID Kenya 2015 Parliamentary Strengthening Programme | KE2 | USAID | Project |
| 3 | South Africa | South Africa 2015 Report-Design & Implementation Evaluation of the Urban Settlements Development Grant | ZA1 | Govt | Project |
| | | Evaluation of the DFID-funded Safer South Africa Programme | ZA2 | UNICEF | Cluster |
| 4 | Ghana | Ghana local Government Decentralisation Program (LOGODEP): Final Performance Evaluation Report | GH1 | USAID | Cluster |
| | | Estimating the Impact and Cost-Effectiveness of Expanding Secondary Education in Ghana | GH2 | USAID | Policy Project |
| 5 | Uganda | Evaluation of Swedish Support to the Diakonia Uganda Programme 2008–2014 – Final Report | UG1 | Sida | Cluster |
| | | The Effectiveness and Efficiency of Implementing the Chronic Care Model for HIV Care in Uganda | UG2 | USAID | Cluster |
| 6 | Tanzania | Tanzania 2015 The Tanzania Social Marketing Project – A Performance Evaluation | TZ1 | USAID | Project |
| | | Tanzania 2015 External evaluation. Amref Health Africa Project on Sex. Reprod. Health Right | TZ2 | Sida | Project |

| S/N | Country | General evaluation: Title of evaluation | Code | Donor support | Category of evaluation |
|---|---|---|---|---|---|
| 7 | Botswana | An Information System for Gender-Based Violence Care and Support: Botswana | BW1 | USAID | Project |
| | | UNEG Botswana 2015_GoB-UN POP Final Evaluation Report 23 October 2015 | BW2 | UNEG | UN programme |
| 8 | Rwanda | Rwanda 2015 Evaluating the Cost and Quality of Service Delivery for Orphans and Vulnerable | RW1 | USAID | Project |
| | | Rwanda 2015 Using Provider Performance incentives to increase HIV testing & counselling Services in Rwanda | RW2 | 3ie | Project |
| 9 | Zambia | Measuring the Impact of SMS-Based Interventions on VMMC Uptake in Lusaka Province, Zambia | ZM1 | 3ie | Project |
| | | Evaluation of the Zambian Governance Foundation (ZGF) Joint Financing Agreement | ZM2 | Sida | Cluster |
| 10 | Nigeria | Targeted States High Impact Project (TSHIP): End-of-project Evaluation Report | NG1 | USAID | Cluster |
| | | Strengthening Integrated Delivery of HIV/AIDS Services (SIDHAS) Project: Mid-term Evaluation Final | NG2 | USAID | Cluster |
| 11 | Zimbabwe | Performance Evaluation of the Strengthening Private Sector Services roject in Zimbabwe | ZW1 | USAID | Cluster |
| | | Zimbabwe Livestock for Accelerated Recovery and improved resiliency (ZRR): Final evaluation report | ZW2 | USAID | Project |
| 12 | Namibia | Endline Performance Evaluation of the Strengthening HIV/AIDS Responses in Prevention and Protection | NA1 | USAID | Cluster |
| | | Effects of Community-based Natural Resource Management on Household Welfare in Namibia | NA2 | 3ie | Project |



4

# SCANDINAVIAN DONORS IN AFRICA: A REFLECTION ON EVALUATION REPORTING STANDARDS

Nozipho Ngwabi & Lauren Wildschut

## Introduction

Evaluations in Africa are largely commissioned by non-African stakeholders, international donors or development agencies that run or fund development programmes on the continent (Bamberger, 2000; Mouton, Rabie, Cloete & De Coning, 2014; Tirivanhu, Robertson, Waller & Chirau, 2018). Evaluation has been institutionalised as a regular part of aid activities, but relatively little attention has been given to the nature or quality of evaluation reports generated by international development agencies, specifically the methodological aspects (Rebien, 1997; Berlage & Stokke, 1992). The importance of the methodological aspects of evaluation reports cannot be overemphasised because the credibility of an evaluation is often based on the methodology used to assess interventions (Alkin, 2004). Rebien (1997) indicates that if an evaluation is not explicit about the designs and methods used to assess and make judgement on an intervention, it becomes difficult to assess the validity of the findings. Donors' information priorities and the evaluation methodologies they use, continue to have considerable influence on how evaluation is both practised and used (or not) by those being evaluated.

Whilst the importance of quality evaluation reports has been strongly foregrounded, some of the challenges in current donor evaluation reporting have also been identified, especially the lack of standardised reporting criteria (Montrosse-Moorhead & Griffith, 2017). There is also a critical need to explore the nature of the power dynamics between donors and African countries, which is enacted through evaluation reporting. The Organisation for Economic Cooperation and Development (OECD) indicates that, "at its best, monitoring and evaluation enables donors and partners to understand their agency and to learn from experience to more effectively influence change on the ground. At its worst, monitoring and evaluation reinforces donors' exercise of power – and disempowers others – in their relationship with in-country stakeholders" (OECD, 2012:267).

This chapter describes the findings of a review of evaluation reports commissioned by three Scandinavian donors – the Danish International Development Agency (DANIDA), the Norwegian Agency for Development (NORAD) and the Swedish International Development Cooperation Agency (Sida). The 142 reports (from 2005 to 2016) examined in the review, focused on African interventions and were published on the donors' respective websites.

The chapter focuses primarily on the methodological elements of the evaluation reports in order to examine the transparency of the evaluation processes, which in turn can be linked to judgements about the quality, credibility, and validity of the evaluations carried out.

The review examined the reports from all three donors to answer the following questions:

1. Are the reporting of evaluation methods standardised across the donors?
2. How do the methodology reporting trends, patterns and gaps differ across the three donors?
3. Who are the evaluators carrying out these evaluations in Africa?

In answering these three questions, we address in part the element of power dynamics in these evaluations commissioned outside of Africa.

## Methodology

The review discussed in this chapter focused on 142 evaluation reports published by DANIDA, NORAD and Sida between the years 2005 to 2015. The evaluation reports were selected purposively based on the following three criteria:

1. The evaluation reports had to be based on interventions in Africa.
2. The reports had to be commissioned by DANIDA, NORAD or Sida.
3. The timeframe of the reports was 2005–2015.

The research used DANIDA, NORAD and Sida web links from the African Evaluation Database (AfrED), the Centre for Learning and Results Anglophone Africa (CLEAR-AA) and the Centre for Research on Evaluation, Science and Technology (CREST) database. The links resulted in the downloading of 306 possible documents. These documents were reviewed and those documents that were not evaluation reports or did not involve African interventions were excluded. This selection process led to 142 reports (DANIDA=32; NORAD=41; and Sida=69).

The 142 selected reports were loaded into ATLAS.ti (a computer-assisted qualitative data analysis programme) for analysis. The review process is illustrated as follows.

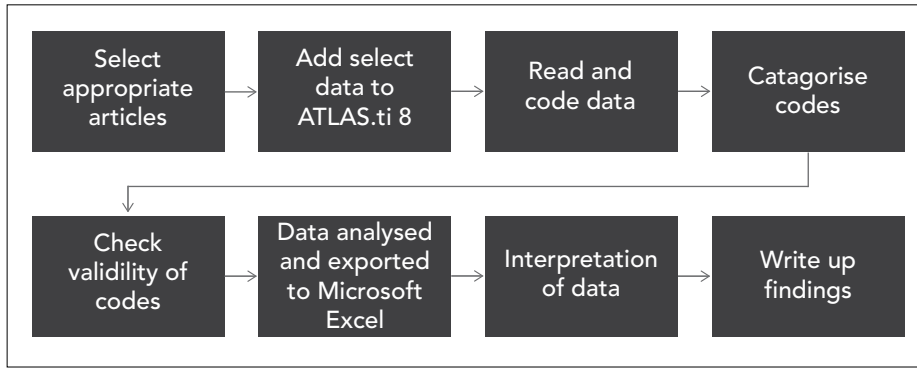



**Figure 4.1** Research process

The above diagram illustrates that 142 sampled documents were added into ATLAS.ti. These were coded based on a code list developed jointly by (CLEAR-AA), the Centre for Research on Evaluation, Science and Technology (CREST) and postgraduate students who also reviewed reports from the AfrED database. The coding list was developed through reviewing evaluation reporting standards listed in the background section of the chapter. The main categories for the codes included: commissioner, evaluation information, evaluation designs, evaluation methods, evaluation methodologies, evaluation purpose, programme information and sectors. The developed codes were compared to the Organisation for Economic Cooperation and Development's Development Assistance Committee (OECD/DAC) quality standards to ensure the research fully measures the reports against their stipulated standards. Both inductive and deductive coding was used. The accuracy of the coding was validated through repeatedly checking the coded documents, which were exported to Microsoft Excel for interpretation. This process was concluded by a write up of the findings. It is important to note that most of the evaluations had more than one country evaluated in each report and more than one consultancy team for some reports, hence the high number of countries and evaluators in the findings section.

## Limitations

The study had a number of limitations, which included:

1. The possibility of bias in selection. The AfrED database, which was used for the selection of the reports, had limitations in scope. Thus there is a possibility of an incomplete search of the evaluation reports.

2. There was a language barrier, because the research was restricted to English-language documents on the websites of each of the three donors Thus the literature review information on all three donors could have been missed in the documents written in each donor's own language.
3. There may have been data-coding gaps. The evaluation reports use different terms for similar concepts and this non-standardisation of terminology left it to the researcher's discretion to code terms and words.
4. The section on the 'evaluators of the reports' was coded through Google searches to check the location of each of the companies. This is not without its challenges as consultants and companies may relocate.
5. Lastly, only a review method was used and this narrows the methods used.

## Background

According to Weiss (1998) and Freeman and Lipsey (2004), programme evaluation is the use of social research methods to systematically assess the process and outcomes of a programme or policy as a means of making value judgements and improving social interventions. From the origins of programme evaluation in the eighteenth century, its aim in development has mainly been to make sure that donor aid is effective in supporting economic growth and reducing poverty (OECD, 2013).

As international development agencies continue to address poverty through aid, there is an increasing need to utilise the lessons of evaluations, to ensure both 'evidence-based' interventions and policy making (Berhane, 2015; Picciotto, 2003).

Many evaluations (not only those carried out by international development agencies) currently aim to align with either the evaluation criteria of utility, feasibility, accuracy, and accountability (Joint Committee on Standards of Education Evaluation, 1994) or the OECD criteria of relevance, effectiveness, efficiency, impact and sustainability (OECD, 2010). This impetus, together with the "growing, global demand for evaluation, has placed the field's longstanding concerns with quality, validity, and credibility once again in the spotlight" (Montrosse-Moorhead & Griffith, 2017:577). The quality, validity and credibility of evaluations is to a great extent measured through the reports they produce, and "without consistent, complete and transparent reporting… the quality of evaluations may not be clear to anyone other than the primary evaluators" (Montrosse-Moorhead & Griffith, 2017:580) and this threatens the credibility of evaluations. The American Evaluation Association (AEA) guidelines highlight the

methodological dimension of evaluation, and explains that it is "important for evaluators to communicate their methods and approaches accurately and in sufficient detail to allow others to understand, interpret and critique their work" (Morris, 2008; AEA, 2018). Montrosse-Moorhead and Griffith (2017) concur that no evaluation would be complete without describing the procedures used.

An early step in the review process was to explore whether evaluation guidelines assist both commissioners and evaluators in terms of reporting standards and whether these guidelines stipulate that evaluators must report on methodology. Eleven sets of guidelines, including those of the three donors, were examined. Let us first consider the notion of 'standards'.

## Evaluation reporting standards

According to Podems and Cloete (2014:315) evaluation standards are "generally agreed upon criteria with which evaluation practices must comply in order to be accepted as valid and credible … Whilst evaluation standards describe what comprises a quality evaluation there are no universally agreed upon criteria for evaluation standards".

When examining evaluation guidelines or standards, it is evident that there are four disparate groups developing guidelines for evaluators (and commissioners): voluntary organisations for professional evaluation (VOPEs), international development agencies, government departments, and evaluation scholars and practitioners themselves (particularly in evaluation research).

The titles of the guiding documents which were reviewed included the descriptors 'guidelines', 'standards', 'manual' and indicated a range of purposes: *to improve the quality of evaluations; to improve credibility of evaluations; to support partnerships; and to foster dialogue for accountability, for improved effectiveness of projects and for learning*. With these different purposes also come varying degrees of expected compliance to the standards set and in some cases the concept of 'standard' was diluted into 'guidance'.

In a bid to understand and determine the degree of structure and standards (or guidance) provided to evaluators on how to write evaluation reports, documents from three VOPEs, one network, five donors, one international organisation and one government department were reviewed. The findings are shown in Table 4.1.

**Table 4.1** Guidance on evaluation reporting provided in documents

| Organisation | Inclusion of domains to be covered in evaluation reports | Reference to inclusion of methodology reporting | Reference to inclusion of local evaluators |
|---|---|---|---|
| African Evaluation Association Guidelines (2006) | ■ | ■ | |
| American Evaluation Association Guiding Principles (2018) | | ■ | |
| Evaluation standards for Aotearoa New Zealand Evaluation Association (ANZEA) (2015) | | | |
| Active Learning Network for Accountability and Performance (ALNAP) Evaluation of Humanitarian Action Guide (2015) | ■ | ■ | |
| United Nations Evaluation Group Norms and Standards for Evaluation (2016) | ■ | ■ | ■ |
| PEPFAR[1] Evaluation Standards of practice (2017) | ■ | ■ | ■ |
| DANIDA Evaluation Guidelines (2012) | ■ | | ■ |
| NORAD Evaluation Policy (2016) | ■ | ■ | ■ |
| Sida Evaluation Manual (2007) | ■ | ■ | ■ |
| OECD/DAC Quality Standards for Development Evaluation (2010) | ■ | ■ | ■ |
| South African Department of Planning, Monitoring and Evaluation Standards for Evaluation in Government (2014) | ■ | ■ | |

Table 4.1 indicates the varying degrees of guidance provided on evaluation reporting in the documents reviewed. Most (9) of the documents include the domains that should be covered in evaluation reports – only two do not (AEA, 2018, and ANZEA, 2015). Nine of the eleven documents also include reference to methodology specifically. It can also be determined from Table 4.1 that all three Scandinavian donors under discussion have fairly detailed evaluation reporting guidelines, which clearly stipulate what should be included

[1] The US President's Emergency Plan for AIDS Relief.

in evaluation reports. and two (NORAD and Sida) have specifically noted the importance of reporting on evaluation methods. It is important to note that DANIDA, NORAD and Sida are all members of both OECD and ALNAP and should also align their reporting requirements with these bodies. Whereas NORAD and Sida mention the importance of reporting on evaluation methods, DANIDA does not. In addition, evaluation stakeholders would expect the three donors to comply with the influential OECD/DAC Standards in particular.

Table 4.1 also indicates that only seven guidelines refer to the inclusion of local evaluators in the evaluation process. Whilst the inclusion of local evaluators is not directly linked to methodology, we have included this element in the discussion in this chapter, as it is has particular relevance to the three Scandinavian donors under discussion and their adherence to the DAC Quality Standards for Development Evaluation.

It is interesting that none of the documents foreground the use of local evaluation methods except for ANZEA, which clearly requires culturally responsive evaluation: "The methodology and methods are culturally responsive and appropriate for the context, and 'fit for purpose'. Evaluation findings, judgements and conclusions are credible, trustworthy, and contextually and culturally meaningful" (ANZEA, 2015:21).

On closer examination of the eleven documents guiding the practice of evaluation reporting, it is evident that there are differing elements which are regarded as important. Table 4.2 shows the difference between three of the documents (with a focus primarily on evaluation methodology). The PEPFAR elements are shown on the left (and are used as the benchmark) and the OECD and African Evaluation Association (AfrEA) references are compared with these elements.

**Table 4.2** Methodology elements compared

| Methodology Elements (based on PEPFAR) | OECD/DAC | AfrEA |
|---|---|---|
| Overall evaluation design | | |
| Type of evaluation (process, outcome, impact, economic) | | |
| Summary of stakeholder engagement | | |
| Sampling strategy | | |
| Data collection methods are rational and aligned to evaluation questions, sources of data analytic methods, and rationale | | |

| Methodology Elements (based on PEPFAR) | OECD/DAC | AfrEA |
|---|---|---|
| Ethical considerations and assurances (e.g. non-research determination and/or IRB approval with dates; application of informed consent, if appropriate; procedures to ensure human rights protection) | | |
| Deviations and adjustments (if any) from the approved SOW/protocol | | |
| Procedures used to ensure that the data are of highest achievable quality | | |
| Data analysis plan | | |
| Limitations of the design and analytic methods | | |
| Findings and conclusions | | |

The PEPFAR evaluation reporting guidelines are very detailed and include a brief summary in the main document of the requirements, a checklist (with ten components of the evaluation report), an evaluation inventory (summary of evaluation components) and adherence checklist. The full document runs to 48 pages.

The OECD document is 24 pages long but only has four pages on evaluation reporting, although elements such as *design* (part of PEPFAR methodology elements) are contained in other sections of the document. The AfrEA document does not have a dedicated section on reporting at all but elements of evaluation reporting can be found under the *Utility, Feasibility, Precision and Quality* and *Respect and Ethics* sections. The document is 19 pages long. As the AfrEA document does not have a separate section on reporting, the four times reporting was mentioned, have been primarily used for the comparative table.

We now turn to the fourth contributor to evaluation standards – the evaluation scholars and practitioners themselves. Montrosse-Moorhead and Griffith (2017:586) present a checklist of elements they developed through a review of evaluation research articles and their own work. The full list of elements is entitled a Checklist for Evaluation-Specific Standards (CHESS), but only the categories of elements pertaining to the "investigation design and methods" is summarised below:

1. Evaluation purpose
2. Evaluation approach
3. Procedure(s) for identifying stakeholders
4. Procedure(s) for prioritising stakeholders
5. Valuing process
6. Sample

7. Sampling procedure(s)
8. Procedure(s) for establishing questions
9. Procedure(s) for establishing questions
10. General methodological orientation
11. Research design
12. Data collection instruments

When you compare the extract from the CHESS list to that of PEPFAR, it is evident that there is a significant overlap in the elements but there are also omissions. The CHESS list does not include *ethics*, *data analysis* and *limitations* (the last two covered in part in their section on *Evaluative Argument and Conclusions*). The diversity of opinion about the specific nature of what should be included in evaluation reports remains a challenge for evaluation stakeholders. When one considers the variances and the lack of detail in some of the documents, it is obvious that more work has to be done in the area of evaluation standards (and in particular reporting standards) so that evaluators, commissioners and other key stakeholders can have consensus on what counts as good practice. More importantly without clear articulation and agreement on the standards for evaluation reporting, grantees will not be able to judge the quality of the evaluations being commissioned and carried out by donors from the North. It is evident that a process which involves a cross-walk of documents from all four contributing sources (international development agencies, VOPEs, government agencies, evaluation scholars and practitioners) is urgently required.

It is also critical that key stakeholders in evaluation in Africa contribute to the task of development of standards to ensure that they reflect the contexts in which evaluation is taking place. According to Podems and Cloete (2014), standards always differ according to context and time frame. AfrEA is currently reviewing and updating their evaluation standards, which will contribute to the standardisation of evaluations and evaluation reports in Africa (SAMEA2, 2018). AfrEA seeks to develop standards that will promote high quality evaluations led by African evaluators, and intend to include evaluation theory and practice that is relevant and responsive to African contexts and needs (AfrEA, 2006). In pursuit of this objective, a discourse on the concept of a Made in Africa Evaluation (MAE) is under way with an agenda to ensure contextually relevant evaluations that include African evaluators (Basheka & Byamugisha, 2015; Chilisa, 2015). AfrEA will also need to consider the inclusion and/or engagement of all four groupings of stakeholders in the development of their standards.

[2] South African Monitoring and Evaluation Association.

Meta-evaluation, promoted by Scriven (1969) and Stufflebeam (1974), has also been used as a method to ensure good quality evaluation reports get evaluated. Both Scriven and Stufflebeam contributed to the development of meta-evaluation by developing meta-evaluation checklists to determine good quality evaluation reports (Stufflebeam & Shinkfield, 2007). This quality assurance method has been adopted in a number of organisations and institutions and should be carried out more often, particularly in Africa, where we are often the evaluands instead of the evaluators.

## Scandinavian donors and the OECD/DAC standards

This study measured the features of the evaluation reports of Denmark (DANIDA), Norway (NORAD) and Sweden (Sida), against the evaluation report domains that are outlined by the OECD/DAC and other networks and institutions. The interest in these three Scandinavian countries lay in that the three donors are known as progressive and transformative donors that provide extensive developmental assistance and are commended for their humanitarian aid generosity, with humanitarian aid values that exceed the United Nations' 0.7% gross national income standard (Selbervik & Nygaard, 2006:1; OECD, 2011:11).

Denmark (DANIDA), Norway (NORAD) and Sweden (Sida) are also part of the 32 donors that belong to the OECD network. All three of their evaluation guides and policies make reference to OECD as the governing body for their evaluations (DANIDA, 2012; NORAD, 2016; Sida, 2007; OECD, 2016). Since their inception, OECD/DAC evaluation guidelines have shaped the way most donor agencies and their clients/grantees commission or design and conduct programme evaluations.

The Quality Standards for Development Evaluation aim to improve the quality and ultimately to strengthen the contribution of evaluation to improving development outcomes. Specifically, the standards are intended to:

- improve the quality of development evaluation processes and products;
- facilitate the comparison of evaluations across countries;
- support partnerships and collaboration on joint evaluations; and
- increase development partners' use of each others' evaluation findings.

(OECD, 2010:5)

From 2005, the three donors have published evaluation reports on the development assistance they have provided for African interventions and these have been followed by OECD/DAC peer reviews which assess their development assistance aid standards against the OECD/DAC set standards (OECD, 2016). However, limited literature exists on their evaluation report features.

The review primarily used the domains outlined in the OECD/DAC's Quality Standards for Development Evaluation (2010) but also took into account the AfrEA evaluation standards, the DPME evaluation standards, the American Evaluation Association's guiding principles, and Stufflebeam's meta evaluation checklist (used for quality assurance in evaluation reports), to develop a code list which was used to review the Scandinavian donors evaluation reports for African interventions (OECD, 2016; AfrEA, 2006; DPME, 2014; AEA, 2014; Stufflebeam, 1999). The criteria used included: evaluation commissioner, evaluation information, evaluation methods, evaluation purpose, evaluation approach, programme information, and sectors. However, for the purpose of this chapter the focus will be on the evaluators involved and evaluation methodological lessons learnt from the three donors' evaluation reports. It will also focus primarily on comparing it with the OECD/DAC standards.

## Results

### *Profiling the donors' evaluation reports*

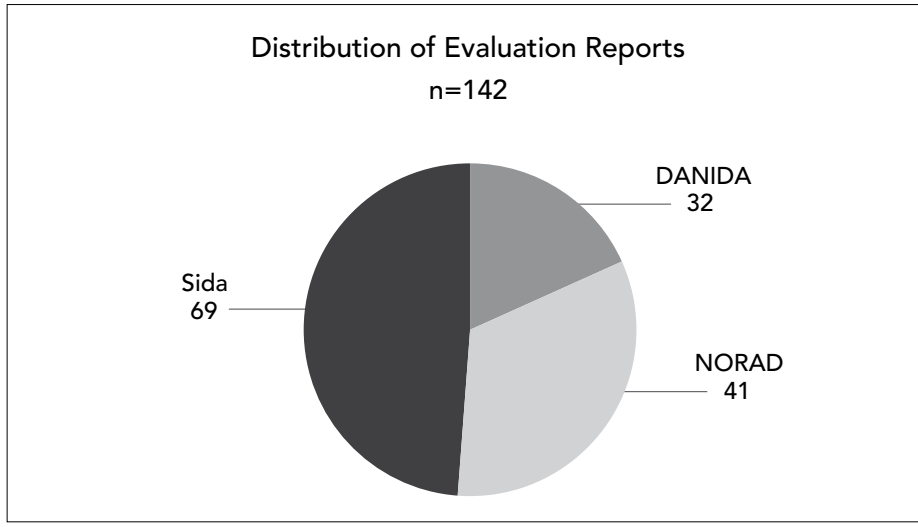



**Figure 4.2** Evaluation reports distribution

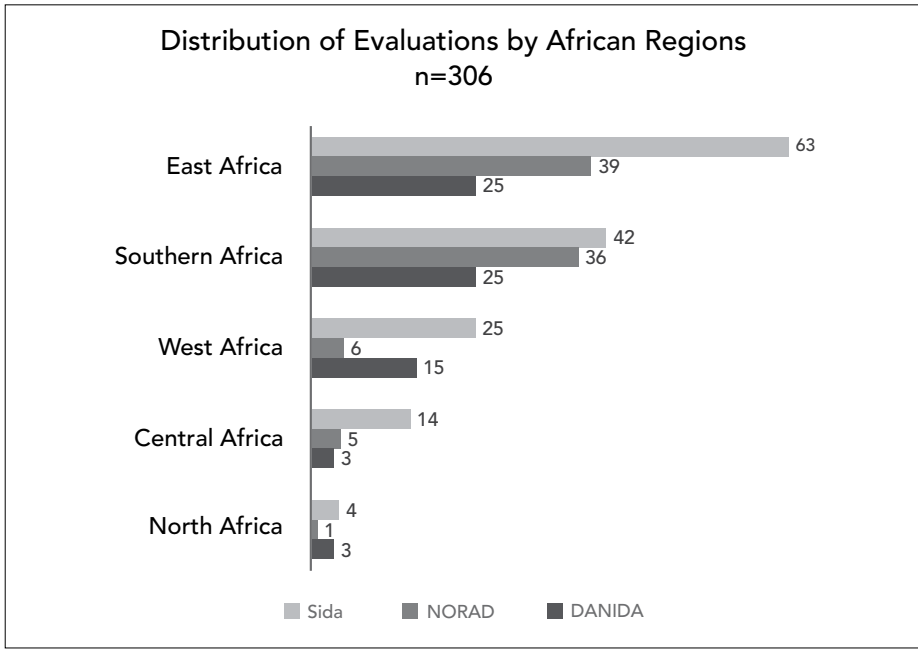



**Figure 4.3** Distribution of evaluation reports by region

Figures 4.2 and 4.3 show the spread of the evaluation reports reviewed – the highest number of reports reviewed were commissioned by Sida (69) but the number of reports overall is fairly significant in a ten-year period. The influence of both the donor-led interventions and evaluations cannot be underestimated.

Figures 4.2 and 4.3 also illustrate the number of evaluations carried out by the three donors in five African regions (the difference in numbers between the reports and evaluations is as a result of some of the reports containing multiple evaluations). In East Africa DANIDA (25), NORAD (39) and Sida (63) carried out a significant number of evaluations while in Southern Africa DANIDA (25), NORAD (36) and Sida (42) also did a large number.

What is also interesting is that most of the evaluation reports commissioned by the three donors utilised the OECD evaluation criteria. This is shown in Figure 4.4.

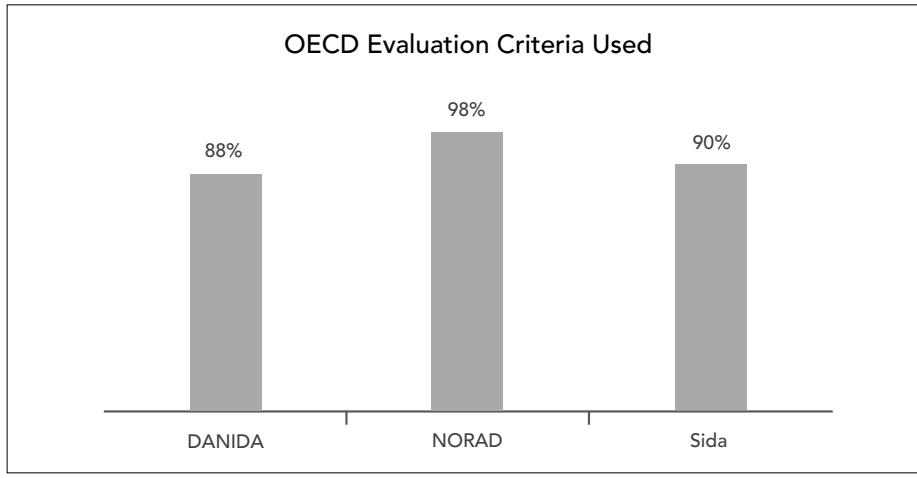



**Figure 4.4** OECD evaluation criteria used in reports

This is not surprising, as all 3 are OECD members. Let us now move on to examine the OECD reporting standards and how these were complied with by the three donors.

## The OECD/DAC reporting standards

The OECD/DAC Quality Standards for Development Evaluation (2010) contains four sections which include:

1. Overarching considerations
2. Purpose, planning and design
3. Implementation and reporting
4. Follow-up, use and learning

This review focuses primarily on the compliance of the reports with Section 3 (Implementation and Reporting), which provides recommended reporting elements for evaluation reports. Although this section has 15 elements, focus will be given to the evaluation reporting elements shown in Table 4.3.

**Table 4.3** OECD standards to be discussed in the findings

| | |
|---|---|
| 2.9 | Purpose planning and design. The purpose, scope, and evaluation questions determine the most appropriate approach and methodology for each evaluation. |
| 3.1 | Evaluation team. The members of the evaluation team possess a mix of evaluative skills and thematic knowledge. Gender balance is considered and the team includes professionals from partner countries or regions concerned. |
| 3.8 | Intervention logic. The evaluation report describes and assesses the intervention logic or theory, including underlying assumptions and factors affecting the success of the intervention. |
| 3.9 | Validity and reliability of information sources. The evaluation report explains the selection of case studies or any samples. Limitations regarding the representativeness of the samples are identified. |
| 3.10 | Explanation of the methodology used. The evaluation report describes and explains the evaluation methodology and its application. |

*Source:* OECD, 2010:10-13

The above standards stipulate the importance of reporting on designs, approaches, the composition of the evaluation team, the articulation of the intervention logic, the data collection tools, the sampling strategy, and a description of the methodology.

The discussion that follows, examines the compliance of the 142 reports with the OECD reporting standards. Figure 4.4 shows the reporting compliance against Criteria 2.9 (design), 3.8 (intervention logic, and theories and approaches), 3.9 (sampling), and elements of 3.10 (data collection tools). The other criteria are addressed further on in this section.

Figure 4.5 shows that the review of the 142 reports revealed that only *data collection* is covered in all (100%) of the reviewed DANIDA, NORAD, and Sida evaluation reports. The three donors were all missing the reporting elements of *design*, *limitations*, *theories and approaches*, *sampling*, and *intervention logic*. The least frequently reported elements were *theories and approaches*, *sampling*, and *intervention logic*.

## Evaluation reporting on methods and approaches

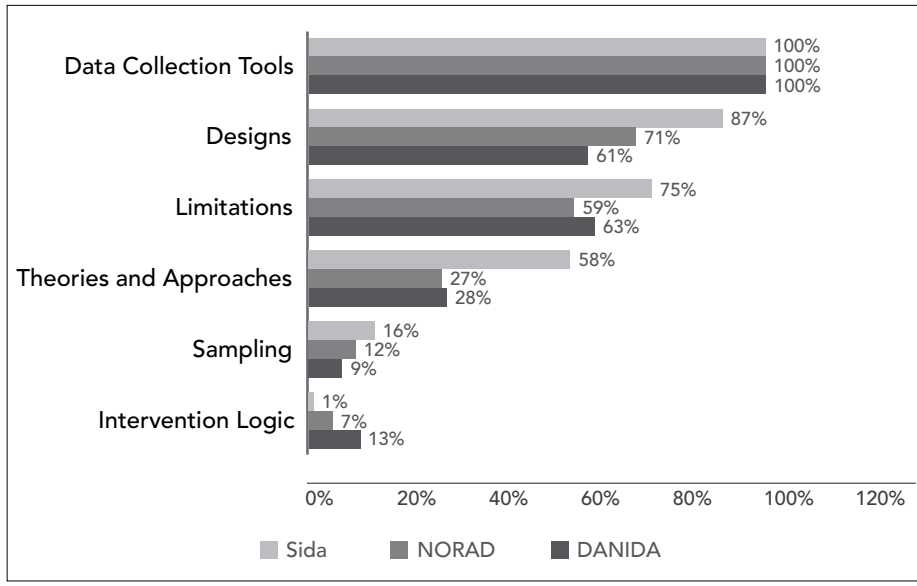



**Figure 4.5** Evaluation reporting

The findings clearly show that the reviewed reports from the DANIDA, NORAD, and Sida have gaps in adhering to their standards.

We now look more closely at OECD Criterium 3.9 and the kinds of sampling reported on by the three donors.

**Table 4.4** Sampling strategies across the donors

| Donors | Purposive | Random | Snowballing | Totals |
|---|---|---|---|---|
| DANIDA | 2 | 1 | 0 | 3 |
| NORAD | 1 | 3 | 1 | 5 |
| Sida | 4 | 0 | 0 | 4 |
| Totals | 7 | 4 | 1 | 12 |

Table 4.4 shows that in only a few cases were the actually sampling strategies named and the most common type of sampling was purposive. Again, the reporting on sampling strategies is very low. We now turn to an unpacking of the data collection methods used (OECD Criterium 3.10).

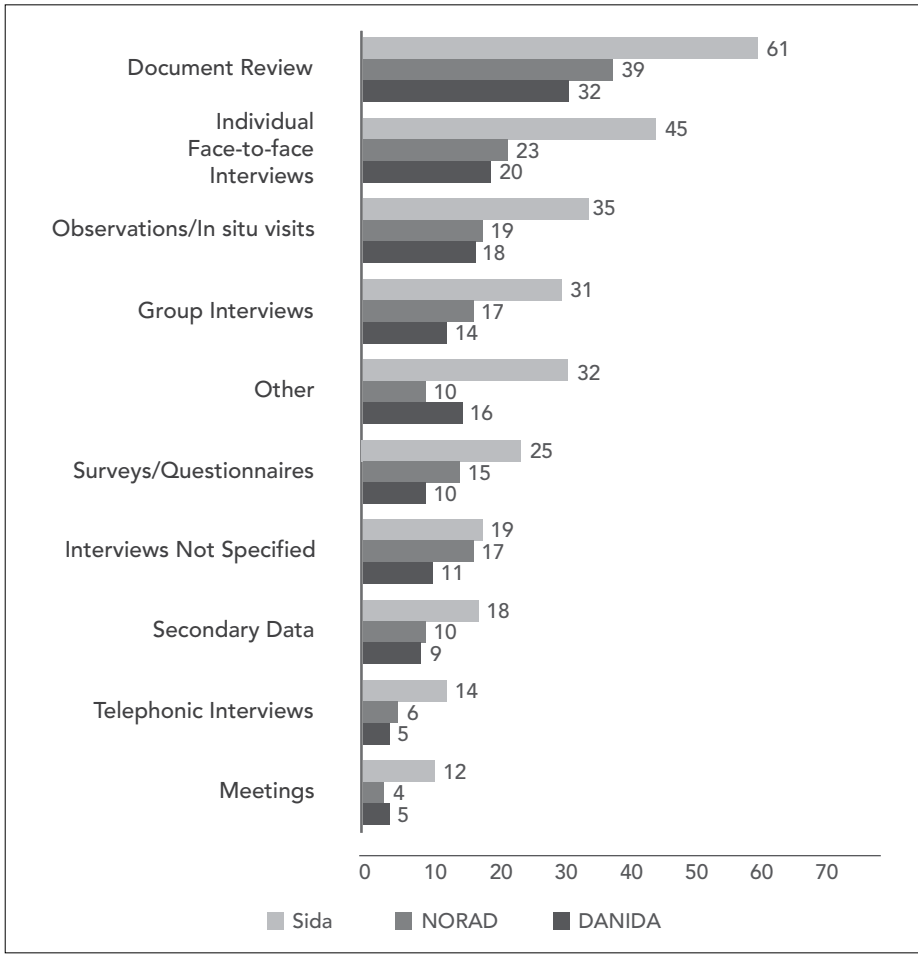



**Figure 4.6** Data collection methods used

Figure 4.6 shows that the most common data source in the donor evaluations is documents followed by individual interviews.

The findings in this section generally illustrate that Sida complies to a greater extent with the standards than the other two donors, but often at a very low level of compliance. The three donors, while complying with OECD evaluation criteria, do not apply the evaluation reporting standards to any great degree.

## Evaluation team (OECD Criterium 3.1)

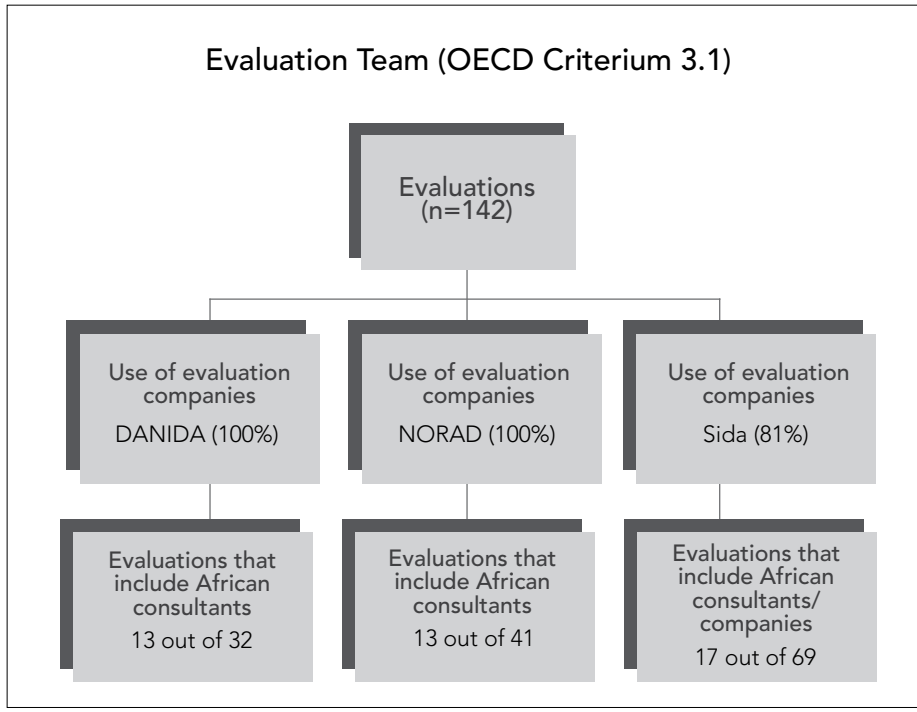



**Figure 4.7** Evaluators

Figure 4.7 shows findings on the evaluators' details by donor. From the 142 reports, DANIDA and NORAD both used evaluation companies for all the reviewed reports, whilst Sida used evaluation companies for 81% of its reports and for the other 19% used independent consultants. In addition, each of the three donors had a ration of less than 20 African consultants or African consultancy companies for their reports. The findings coincide with findings elsewhere that there is a problem of visibility of the profiles of African evaluators (Mouton, Rabie, De Coning & Cloete, 2014).

**Table 4.5** Geographical location of evaluation consultancy companies

| | DANIDA | NORAD | Sida | Totals |
|---|---|---|---|---|
| Belgium | 1 | 4 | 0 | **5** |
| Canada | 4 | 0 | 2 | **6** |
| Finland | 0 | 6 | 0 | **6** |
| Germany | 2 | 2 | 0 | **4** |
| Netherlands | 3 | 0 | 1 | **4** |
| Norway | 0 | 14 | 2 | **16** |
| Sweden | 4 | 5 | 42 | **51** |
| UK | 16 | 15 | 3 | **34** |
| USA | 1 | 4 | 0 | **5** |
| Denmark | 17 | 7 | 3 | **27** |
| Africa | 0 | 1 | 6 | **7** |
| Totals | **48** | **58** | **59** | |

Table 4.5 shows findings on the *geographic locations of the consultancy companies* used for the reports reviewed. It is important to note that some reports were conducted by more than one consultancy company per evaluation, hence the total number of evaluation consultancy companies exceeds the 142 reports reviewed.

Most of the reports (51) were written by consultancy companies in Sweden followed by United Kingdom (34) for the three donors. In addition, only seven evaluations companies were primarily African and six of these were commissioned by Sida. The findings further show that, whilst only seven evaluations included African consultancy companies, most of the evaluation reports were conducted by evaluation companies in Scandinavia (Denmark, Norway, Sweden and Finland), and the rest were from developed countries (in Europe, America and Canada).

These results coincide with Phillips' (2018) findings on four major donors (DFID[3], USAID, the EU and GIZ[4]) that indicated that evaluation contracts are usually given to companies in their home countries or internationally and then these companies often sub-contract local expertise to enable them to understand the local environment. The use of global North evaluators for African interventions raises concerns about why and how the majority of evaluation reports for African interventions are conducted by global North evaluators.

[3] Department for International Development.
[4] *Deutsche Gesellschaft für Internationale Zusammenarbeit.*

## Summary of findings

1. *Are the reporting of evaluation methods standardised across the donors?*

There is huge variance between the documents guiding evaluators in their practice around evaluation reports. It is a matter of urgency for AfrEA to involve the key stakeholders in the development of standards and guidelines (four groupings) in order to develop asset of standards which can be used for both African evaluators and in particular evaluators from the North to ensure quality evaluation reports from which we can learn.

2. *How do the methodology reporting trends, patterns and gaps differ across the three donors?*

The methodology and approaches reporting patterns of DANIDA, NORAD, and Sida show that the three do not fully comply with the OECD/DAC Quality Standards for Development Evaluation. Whilst Sida shows a trend of complying the most, they still only comply to a limited extent. Because of this lack of compliance the reports across the donors are not standardised and in fact cannot be judged in terms of their credibility and validity. This means Africans, and in fact all key evaluation stakeholders, cannot learn about how to improve interventions. In addition, it calls into question the quality of these evaluations. The key gaps around the logic of interventions and sampling need to be addressed urgently.

3. *Who are the evaluators carrying out these evaluations in Africa?*

The reports are mostly conducted by Scandinavian consultancy companies with limited inclusion of African consultants which raised the issue of capacity development of African evaluators and culturally competent evaluation. Although the guidelines and policies of the three donors stipulate inclusion of 'local expertise', it is clear that this clause does not result in many local evaluation consultancies participating in the evaluations. None of the lead evaluators are African.

The field of evaluation involves power dynamics and it indicates how 'credibility' and 'validity' are not free from power and epistemology.

## Conclusion

Donors remain the biggest drivers in development aid evaluations. From the literature reviewed and results from the reports reviewed, the research concluded that although evaluation guidelines and standards exist amongst VOPEs, multilateral and bilateral donors there are serious gaps in evaluation reporting for the three donors studied. These findings call into question the 'credibility, validity and transparency' of the evaluation reports based on the principle that evaluations are judged by their methods.

Bewyl (2006:27) argues that as there are different perspectives on what constitutes high-quality evaluation and a range of risks to the potential influence of evaluation if we do not agree on standards for high-quality evaluation practice. He argues for "evaluation standards that win the support of a wide variety of national societies and organisations, many different democratic thinking traditions, and a wide range of evaluation cultures". He claims (2006:23) that currently evaluation standards do not adequately address the issue of values and that "evaluation in the context of democratic governance should make transparent how it handles value claims".

It is evident to the authors of this chapter, that developing contextually responsive standards is an urgent matter – not only for Africa – but also internationally. In addition, there is the questions of compliance or application of standards: how do we ensure evaluators use the standards agreed upon, and who checks this? We support Bewyl's position that the values which underpin evaluation standards (and particularly those relating to reporting) need to be made transparent and disseminated to all evaluation stakeholders to ensure that high quality evaluations and high quality reports are delivered and utilised in support of democratic policy making and the improvement of social interventions. This chapter serves as a call to action to all African evaluation stakeholders to begin a process of developing standards so we can judge the quality of evaluation practice and reporting of those evaluators commissioned by international development agencies and the quality of evaluation practice and reporting we ourselves carry out.

## References


AEA (American Evaluation Association). 2018. *Guiding principles for evaluators*. https://www.eval.org/p/cm/ld/fid=51 [Accessed 20 February 2019].

AfrEA (African Evaluation Association). 2006. *African Evaluation Guidelines: Standards and Norms*. [Online]. http://www.ader-evaluare.ro/docs/African%20Evaluation%20Association.pdf [Accessed 20 June 2018].

Alkin, M.C. 2004. *Evaluation roots: Tracing theorists' views and influences*. London: SAGE.

ANZEA (Aotearoa New Zealand Evaluation Association). 2015. *Evaluation standards for Aotearoa New Zealand*. http://www.anzea.org.nz/evaluation/evaluation-standards/ [Accessed 20 February 2019].

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

Basheka, B.C. & Byamugisha, A. 2015. The state of monitoring and evaluation (M&E) as a discipline in Africa. *African Journal of Public Affairs*, 8(3):75-95. https://doi.org/10.4102/apsdpr.v3i2.80

Berhane, E. 2015. Health evaluations in Africa – A review of the health strand held at the 7th Biennial Conference of the African Evaluation Association. *African Evaluation Journal,* 3(2):1-7. https://doi.org/10.4102/aej.v3i2.170

Berlage, L. & Stokke, O. (eds). 1992. *Evaluating development assistance: Approaches and methods* (volume 14). London: Psychology Press.

Bewyl, W. 2006. The role of evaluation in democracy: Can it be strengthened by evaluations standards? A European perspective. *Journal of MultiDisciplinary Evaluation*, 6(3):10-29.

Buchanan-Smith, M.; Cosgrave, J. & Warner, A. 2016. ALNAP: Evaluation of humanitarian action guide. Unpublished evaluation report. London: ALNAP.

Chilisa, B. 2015. *A synthesis paper on the Made in Africa Evaluation concept*. Gaborone: University of Botswana.

DPME (Department of Planning, Monitoring and Evaluation). 2014. *Standards for evaluation in government*. Pretoria: Government Printing Works.

Joint Committee on Standards of Educational Evaluation. 1994. *The program evaluation standards: How to assess evaluations of educational programs*. Thousand Oaks, CA: SAGE.

Montrosse-Moorhead, B. & Griffith, J.C. 2017. Towards the development of reporting standards of evaluations. *American Journal of Evaluation*, 38(4):577-602. https://doi.org/10.1177/1098214017699275

Morris, M. 2007. *Evaluation ethics for best practices*. New York: Guilford Press.

Mouton, C.; Rabie, B.; Cloete, F. & De Coning, C. 2014. Historical development and practice of evaluation. In: F. Cloete, B. Rabie & C. de Coning (eds). *Evaluation Management in South Africa and Africa*. Stellenbosch: African Sun Media. 28-78. https://doi.org/10.18820/9781920689513

OECD (Organisation for Economic Cooperation and Development). 2010. *DAC guidelines and reference series: Quality standards for development evaluation*. [Online]. http://www.sida.se/English/how-we-work/evaluations1/ [Accessed 12 January 2018].

OECD (Organisation for Economic Cooperation and Development). 2011. *Denmark Development Assistance Committee (DAC) Peer review*. [Online]. http://www.oecd.org/derec/denmark/35140874.pdf [Accessed 3 December 2017].

OECD (Organisation for Economic Cooperation and Development). 2012. *Poverty reduction and pro-poor growth: The role of empowerment*. [Online]. https://www.oecd-ilibrary.org/development/poverty-reduction-and-pro-poor-growth_9789264168350-en [Accessed 1 March 2019]. https://doi.org/10.1787/9789264168350-en

OECD (Organisation for Economic Cooperation and Development). 2013. *Evaluating development activities: 12 lessons from the OECD/DAC*. [Online]. http://www.oecd.org/dac/peer-reviews/12%20Less%20eval%20web%20pdf.pdf [Accessed 13 January 2018].

OECD (Organisation for Economic Cooperation and Development). 2016. *Development co-operation report 2016*. [Online]. http://www.keepeek.com/Digital-Asset-Management/oecd/development/development-co-operation-report-2016_dcr-2016-en#page240 [Accessed 13 January 2018].

PEPFAR (The US President's Emergency Plan for AIDS Relief). 2016. *Evaluation standards of practice*. https://www.pepfar.gov/documents/organization/276886.pdf [Accessed 20 February 2019].

Phillips, S. 2018. Diagnostic on the supply and demand of evaluators in South Africa: Diagnostic Report. Johannesburg. Unpublished report by Twende Mbele. www.twendembele.org/wp-content/uploads/2019/04/TWENDE-DS.pdf

Picciotto, R. 2003. International trends and development evaluation: The need for ideas. *American Journal of Evaluation*, 24(2):227-234. https://doi.org/10.1016/S1098-2140(03)00021-3

Podems, D. & Cloete, F. 2014. Evaluation professionalization and capacity-building. In: F. Cloete, B. Rabie & C. de Coning (eds), *Evaluation Management in South Africa and Africa*. Stellenbosch: African Sun Media. 28-78. https://doi.org/10.18820/9781920689513

Rebien, C.C. 1997. Development assistance evaluation and the foundations of program evaluation. *Evaluation Review*, 21(4):438-460. https://doi.org/10.1177/0193841X9702100402

Rossi, P.H.; Lipsey, M.W. & Freeman, H.E. 2004. *Evaluation: A systematic approach*. Los Angeles: SAGE.

Scriven, M. 1969. An introduction to meta-evaluation. *Educational Product Report*, v2:36-38.

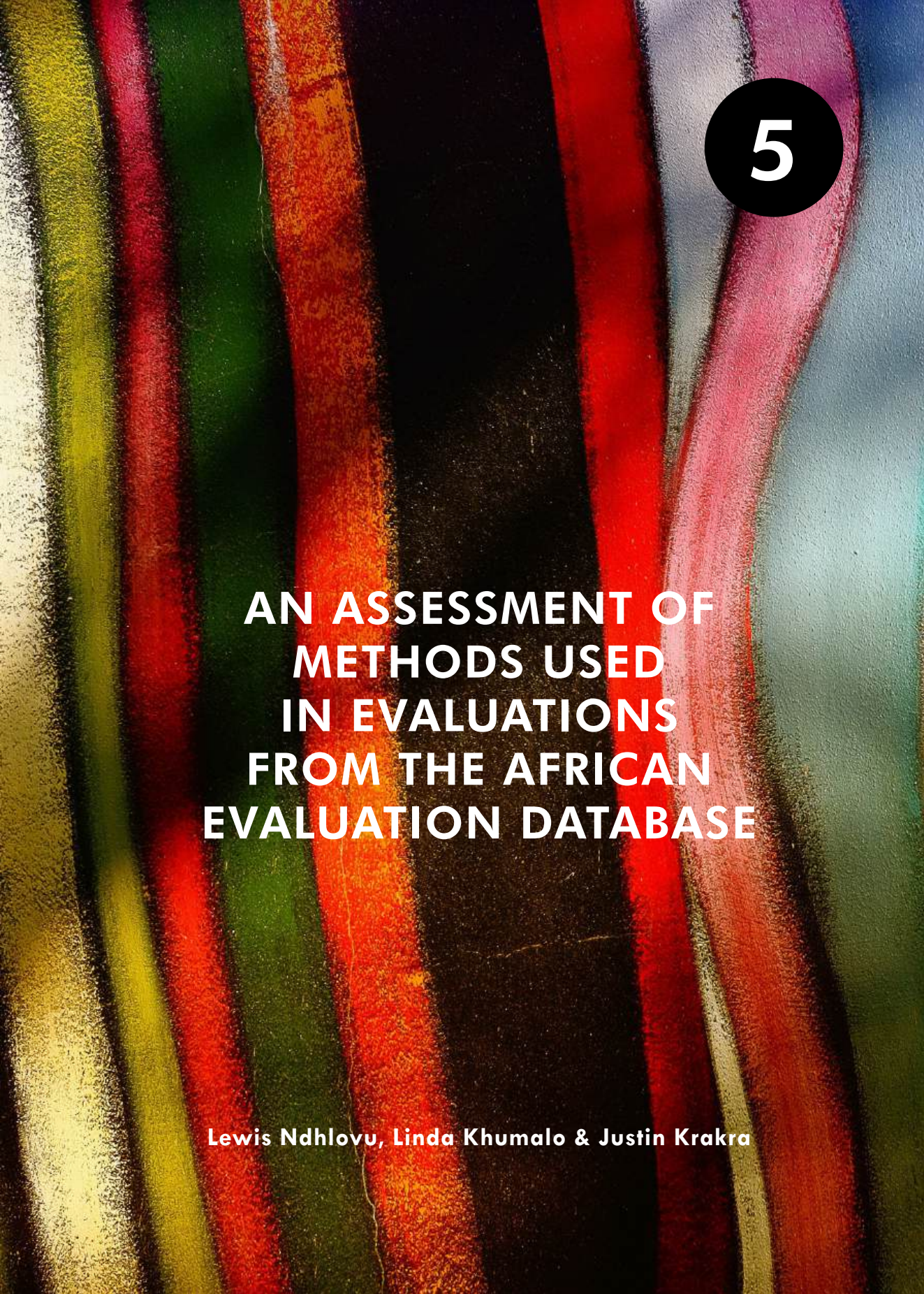
South African Monitoring and Evaluation Board. 2018. Update of AfrEA's Evaluation standards development. Board Minutes of Meeting, Johannesburg, South Africa.

Stufflebeam, D.L. & Coryn, C.L. 2014. *Evaluation theory, models, and applications* (volume 50). Hoboken: John Wiley & Sons.

Tirivanhu, P.; Robertson, H.; Waller, C. & Chirau, T. 2018. Assessing evaluation education in African tertiary education institutions: Opportunities and reflections. *South African Journal of Higher Education*, 32(4):229-244. https://doi.org/10.20853/32-4-2527

UNEG (United Nations Evaluation Group). 2016. *Norms and Standards for Evaluation* http://www.unevaluation.org/document/detail/1914 [Accessed 20 February 2019].

Weiss, C. 1988. If program decisions hinged only on information: A response to Patton. *American Journal of Evaluation*, 9(3):15-28. https://doi.org/10.1177/109821408800900302



5

# AN ASSESSMENT OF METHODS USED IN EVALUATIONS FROM THE AFRICAN EVALUATION DATABASE

Lewis Ndhlovu, Linda Khumalo & Justin Krakra

## Introduction

The importance of evaluations in determining the effect of policies and programmes is highly appreciated by policy and decision makers in both government and development sectors in Africa. For evaluations to be utilised by programme and policy makers, they have to be credible in their design, methods and recommendations. As evaluations are undertaken to inform decisions, learn, clarify options, reduce uncertainties and provide information about programmes and policies (Patton, 1986), it is crucial to ensure that there is credibility, validity, and usability from the information obtained. In this regard, the selection of appropriate methods in evaluations enables evaluation findings to address ways in which programmes and policies can be better implemented.

It has been noticed that while evaluations have been conducted, the utilisation or uptake of their findings has been very low, partly because they lack credibility. Utilisation requires competence of the evaluator at different levels of the evaluation process, from the beginning to the end. An important component of credibility in use of evaluations is that appropriate designs and methods are used. Currently, there are no universally acceptable and absolute standards for assessing evaluation methods (Patton, 2008), especially for those conducted in Africa. Evaluation associations have attempted to establish evaluation standards, although these still need to be localised in line with the Made in Africa agenda of the African Evaluation Association (AfrEA).

This chapter assesses the methods that were used in evaluation reports from the African Evaluations Database (AfrED) dated between 2005 and 2015, with a particular focus on the evaluations conducted during the years 2005, 2010 and 2015. The assessment is based on a secondary data analysis of evaluation reports from AfrED focusing on the methodological design of the studies, the data-gathering techniques, analysis, and reporting. It provides a contextual understanding of the determinants of methods choices focusing, for example, on the role of commissioners of evaluations in selection of methods, the country of origin of evaluators, the relationship between evaluation types and methods, scope of evaluation, and selection of methods. The chapter also analyses trends in the application of qualitative, quantitative, and mixed evaluation methodologies. Based on the analysis of the methods in the evaluation reports, the chapter observes that there is relatively limited use of quantitative methods, even in cases where impact evaluations are conducted This can be attributed to the limited institutional capacity for impact studies.

The objectives of the study are (i) to analyse the methods that were used in the selected evaluation reports on the database; (ii) to examine the determinants of selection of these evaluation methods; and (iii) to assess the extent to which the methods selected are guided by theories, for example, Theories of Change and programme logics. The assessment of evaluation methods in this chapter is further guided by the following research questions:

1. Is the choice of evaluation methods defined by criteria such as types of evaluations, donors/commissioners of the evaluations, publication years of the evaluations, the institution(s) conducting the evaluations, or the scope of coverage of the evaluations?
2. Are the selected evaluation reports guided by theories, for example, Theories of Change, and also the programming?
3. Is there any indication of using innovative/contextually relevant methods in African evaluations?

This chapter contributes to the discourse on the need for contextually appropriate evaluation methods in Africa by critiquing the determinants which in some cases have overlooked the supply side of evaluations by only focusing on demand. There has been an increase in the number of M&E training (courses/modules) across the continent aiming to fill the supply side of evaluations, but little is known about their effectiveness (Tarsilla, 2014; Morkel & Ramasobana, 2017). This study will outline the status and gaps in evaluation capacity development and practice in Africa. The chapter begins with a review of the literature, a description of the methods used and then presents and discusses the findings.

## Literature review

### *Evaluation capacity*

Philips (2018) recommends a thorough assessment of skills gaps which goes beyond the typical generalisation about the shortage of skills/lack of capacity in order to better inform the training/capacity building offerings in M&E. This goes back to efforts to define minimum competency requirements in the field of evaluation which are essential to defining skills gaps and being able to address them. More so, although several instances of self-taught evaluators exist, hence the way we understand supply varies according to how evaluation practitioners are understood. There is still no clear agreement as to what an evaluation consultant should offer. The challenge with having no agreed on framework for evaluation competencies in the continent creates "confusion"

as to what it means for an evaluator to be adequately skilled, and in the case of evaluation methods, what is sufficient for evaluators to be considered as having know-how/skills in evaluation methods. An international debate around evaluation competencies is ongoing and continues to be provocative and is yet to reach a conclusion (Podems, 2014; King & Podems, 2014).

Philips (2018) cautions against assigning the challenge with quality of evaluations entirely to the supply side as there are a range of other possible causes of quality problems in evaluations such as weaknesses in procurement process and management of the service providers. Moreover, the nature of supply also depends on the sector in which the evaluation is carried out as some sectors have more researchers and a better specialist skills base than others. Hence evaluations can be more established in some sectors than others.

Unpacking the landscape of methods used in evaluations has implications on how we apprehend capacity in evaluations and how this contributes to taking the supply and demand dialogue going forward.

### *The use of evaluation methods*

In conducting programme evaluations, evaluators rely on existing social research methods and methodologies (Clarke & Dawson, 1999). Different methods may produce different results when applied to the same evaluation problem, hence it is important for evaluators to be cognisant of their role. Patton (1990) points out that no paradigm is superior to another and therefore encourages evaluators to be situationally responsive and methodologically flexible.

Clarke and Davidson (1999) further note that inappropriate methods can produce misleading findings which may influence decision-making; this could result in the recommendation for the continuation of an ineffective programme or intervention. Patton (1982) states that it is essential for evaluators to make informed choices on methods which includes having an appreciation of the philosophical issues surrounding research methods and social science methodologies.

Qualitative methods are favoured in obtaining descriptions of perceptions and experiences of individuals and groups in particular programmes to aid with understanding the internal dynamics of programmes. On the other hand, as quantitative methods have the advantages of objective quantitative measurement, they merit the ability to establish whether there have been

cause-effect relationships, hence their primary use in impact evaluations (Clarke & Dawson, 1999). Similarly, Rossi and Freeman (1993) equate establishing impact to establishing causality in programmes and support the use of quantitative methods in arriving at causal explanations. The cause and effect relationship is fundamental in defining any programme's theory of change.

While quantitative methods may make it possible to assess the impact of programmes, they face criticism for offering little insights into social processes that help to explain the changes observed. Consequently, this has contributed to the rise in the use of a combination of quantitative and qualitative methods (mixed methods) evaluation designs (Clarke & Dawson, 1999). The importance of evaluators having skills that enable them to draw upon a wide range of methods (and, where appropriate, successfully mix methods) in a single research design is critical (Clarke & Dawson, 1999). Methods triangulation increases the researchers' confidence in the research findings compared to when a single method is used. This strengthens the reliability and validity of the findings. Although triangulation of mixed methods may be ideal, it is notable that they tend to be more resource intensive.

### *Selection of evaluation methods*

There is general consensus amongst scholars that evaluation methods need to be selected primarily according to the purpose of the evaluation (Clarke & Dawson, 1999; Forss, Vedung, Kruse, Mwaiselage & Nilsdotter, 2008; Bamberger, Rugh & Mabry, 2012). The approach most suitable for a specific evaluation study is determined by the specific evaluation questions and objectives adopted for the study.

Moreover, the balance in the choice of evaluation methods is also influenced by the nature of the programme being evaluated and the political or decision-making context (Rabie & Burger, 2014, as referenced in Cloete, Rabie & De Coning, 2014). Selection of evaluation designs involves choosing the best possible design, taking into account practicality and feasibility (Shadish, Cook & Leviton, 1991, as referenced in Cloete et al., 2014). According to Patton (2002), choice of design reflects interplay between resources, capabilities, purposes, possibilities, creativity, and personal judgements by the people involved. Similarly, Clarke and Dawson (1999) assert that the choice of methods is strongly influenced by the situation and context in which the evaluation is conducted. Patton (1987) supports this by stating that no rigid rules apply to

data collection and selection of methods in evaluations as these decisions are dependent on specific situations and particular policy making contexts. The purposes and circumstances of evaluations determine the choices of methods. For example, one should recognise that time and resource constraints that evaluations typically face may result in difficult trade-offs (Shadish et al., 1991). In some situations, the evaluator may find that pragmatic concerns outweigh theoretical concerns in method selection. These are the issues that the chapter seeks to address.

It is noteworthy that some clients often prescribe the choice of data collection methods in the terms of reference. This, in combination with the challenges of limited time and resources, often rules out the use of more complex designs and methods (Forss et al., 2008).

Bamberger et al. (2012) identify factors that should be considered in the selection of evaluation designs as follows:

- **The purpose of the evaluation**, given that different purposes require different methodologies and different levels of methodological rigour.
- **The complexity of the evaluand**. For instance, most conventional impact evaluation methodologies can be adapted with difficulty to evaluate complicated programmes, but are not applied to evaluate complex interventions.
- **The context within which the evaluation will be implemented**. This includes history, economic, political, socio-economic, and natural environmental issues, organisations and agents involved in the project, as well as the needs and interests of the different stakeholders. Moreover, political and organisational factors can often influence the range of evaluation designs which may potentially constrain the evaluation.
- **The level /scale of the intervention**. An evaluation could involve one or a few local communities, or could be located at a programme or sector level, sub-nationally, nationally, or at a multi-country or regional scale. The geographic scale may not be related to the complexity of the evaluation.
- **The size of the evaluation budget**. Small evaluation budgets often limit the sample size and precision of statistical estimates. Furthermore, budget constraints can also limit planning time, instrument development and testing, data collection and analysis, as well as the level of professional expertise.
- **The commissioning of the evaluation**. This could be at the start of the intervention (baseline), mid-term, and end of the intervention or after the intervention. In some instances, when a baseline study is planned, administrative delays in commissioning the evaluation and political pressures to start delivering services could result in project implementation starting prior to conducting the baseline.

- **The client**. Typically the client could be a donor agency, a government planning ministry, implementing agency or civil society. The client has a major influence on the evaluation questions that are asked and not asked, on the methodology, and timing of the evaluation. The client can further determine the groups consulted and those that are not consulted during the evaluation (Bamberger et al., 2012:212-213).

It is further worth noting that evaluations are often time-consuming, expensive and requiring skills in planning, data collection, analysis, and reporting. Quantitative methods are more suitable for medium to large-scale evaluations. Qualitative methods, on the other hand, are useful for understanding what kind of issues to consider in any evaluation. Befani and O'Donnell (2016) developed a tool to assist with determining the appropriate evaluation methods to use. The tool identifies the key themes that are central to influencing the choice of the appropriateness of evaluation methods as follows: the methods used being best suited to answering key evaluation questions, the interests and preferences of the stakeholders, the extent to which evaluation methods meet these, and, lastly, ensuring that the features of the intervention match the feasibility of the methods that are being applied.

It is clear that the decision on what evaluations to conduct and data methods to use is complex and depends on a number of contextual factors. These factors have to be examined at the design stage, but data may not be available to make conclusive statements.

## Conceptual framework

Given the complexities unearthed in the preceding section, there was a need to develop some conceptual framework that can be used as a decision tool for classifying evaluation designs and their methods. The conceptual framework is set at two levels: the evaluation types and the study designs as the first level, and the data-gathering methods as the second level. This is important because the judgement of whether an approach was of a qualitative or quantitative nature involved both levels. The determination of whether the approach was quantitative, qualitative or mixed was hinged on three key issues, namely evaluation types, study designs and the methods for gathering data. This framework allows for a standard against which to make judgements of which methods could be classified into each of the three categories of data gathering methods (quantitative, qualitative and mixed). A brief literature review of evaluation types and data-collection methods provides us with a good basis for developing a framework.

## First conceptual level: Classifying evaluations

Given that evaluation is a relatively new field, there is still a lack of consensus on the typography of evaluations, occasionally resulting in confusion about the use of terms and concepts. Evaluations are described in many ways with common reference to formative, summative, process, outcome, and impact evaluations as different types of evaluations.

In examining the different types of literature and reports in the database, it was most useful to adopt a dichotomous type of classification, which made the data from the database usable. Consequently, the classification of types of evaluations slots them into two categories of *formative* and *summative* (Scriven, 1967). Formative evaluations comprise of learning and improvement of planned intervention during the process, in order to improve the process. On the other hand, summative involves accountability and judgement of the overall merit, worth, value and significance of a completed programme. Summative evaluations can inform major decisions about future programming (Bamberger et al., 2012).

While formative evaluations tend to be equated with process evaluations, we also point out that outcome evaluations can be equated to summative evaluations which are associated with outcome and impact assessments. Impact assessments seek to make a statement about the attribution of interventions to results observed. Further, formative evaluations tend to make use of qualitative data techniques to ensure that the evaluator maintains close contact with the programme. No attempt is made to generalise findings beyond the setting in which the evaluation takes place. They typically rely on qualitative methods as the findings are context specific (Clarke & Dawson, 1999).

On the other hand, as summative evaluations are primarily focused on whether particular programmes have achieved their set objectives, the data collection tends to put more emphasis on quantitative outcome measures. Moreover, summative evaluations are typically inclined towards experimental and quasi experimental research designs given their aim in determining whether or not a programme works (Clarke & Dawson, 1999). Additionally, summative evaluations seldom rely entirely or even primarily on qualitative data because of decision makers' interests in measuring standardised outcomes, having controlled comparisons and making judgements about effectiveness from large samples with statistical pre-post and follow-up results (Patton, 2002).

Therefore, it becomes clear that the types of evaluation and their designs may determine the methods that are used. However, even with different evaluation types, data gathering processes can be of different types, quantitative or qualitative.

## The second conceptual level: Methods for gathering data

The second level defining evaluations are the methods these evaluations use to gather data. This section describes how methods were classified as quantitative, qualitative or mixed. Evaluation studies can adopt qualitative or quantitative data and often include both designs which are referred to as mixed methods.

### *Quantitative methods*

Most textbooks of evaluation and social science define quantitative studies to fall into the categories of cross-sectional study and experimental designs (Bresee, 2014). Other approaches that are sometimes included in the quantitative approach are secondary data analysis, structured observation, and content analysis.

The experimental and quasi-experimental designs have tended to focus on the outcome or impact level of the elements of the results chain. This is understandable because of the logic of attribution in exploring cause and effect. It is worth noting that these experiments tend to be complex and expensive, which warrants their use only for severe problems with multiple potential solutions and a clear commitment to using the findings (Boruch, 2004, referenced in Cloete et al., 2014).

### *Qualitative methods*

Qualitative approaches and their characteristics were also identified. These approaches tend to be most relevant in situations when non-causal questions form the basis for the evaluation when contextual knowledge, perspectives, and values of the evaluand (programme being evaluated) are required. Additionally, qualitative approaches focus on the constructed nature of social programmes, the context of interventions, processes of implementation as well as assessing programme results (Cloete et al., 2016). This description formed the basis for classifying methods in the qualitative category.

The qualitative methods that are commonly used in social science research, such as participant observations, in-depth interviews, focus group discussions and participator-action research methods (archives, checklists, content analysis, Delphi techniques, narrative analysis, storytelling and observations) were identified. In the AfrED database, the most commonly used methods were the standard qualitative methods, namely in-depth interviews, focus group discussions and observations.

### *Mixed methods*

Mixed methods evaluations aim to integrate quantitative and qualitative approaches to theory, data collection, data analysis, and interpretation. The purpose is to strengthen the reliability of data, validity of the findings and recommendations, and to broaden and deepen understanding of the processes through which programme outcomes and impacts are achieved. The OECD (2007) argues for the use of a plurality of approaches to be able to benefit from the complementarities in the information that each approach provides. Combining at least two methods has been widely accepted as beneficial as it promotes triangulation which tests the consistency of findings obtained through different instruments. The complementarity of methods thus validates results. Expansion of the richness and detail of the study through exploring different features of each method is also promoted (Cloete et al., 2016).

Table 5.1 was used as a guide for identifying the types of evaluations in the repository database. While the table presents a neat classification, some challenges were experienced in the practical classification of the evaluations. In numerous reports drawn from the database, the researchers or authors of the reports neither indicated precisely what the evaluation types used were nor the designs and methods that were adopted. Therefore, the description of the characteristics constituted a framework for forming a judgement for various classifications. Hence the clarification of some of the characteristics of the design and methods used became important.

While the focus has been on the distinction between qualitative and quantitative results, many times there was cross over between methods and the distinction became blurred. For instance, qualitative results could be presented in a quantitative form as frequency distributions, and cross variable tabulations could be constructed from qualitative data. Therefore, the judgement of use of qualitative against quantitative methods had to be

based on a wide range of criteria. Table 5.1 links the different evaluation types with methods.

**Table 5.1** Evaluation types and methods

| Evaluation types | Classification | Methods |
|---|---|---|
| Formative/ Implementation | Diagnostic, design and implementation evaluations | Qualitative<br>- Ethnographic techniques (focus group discussions, in-depth interviews, participant observations, case studies) |
| Summative/ Outcome | Impact, economic and synthesis evaluations | Quantitative<br>- Statistical surveys<br>- Experimental designs (randomised control trials)<br>- Quasi experimental designs (pre-programme/ post-programme designs, historical/time series designs) |

## Methodology

The chapter is based on a systematic analysis of a selection of evaluation reports extracted from AfrED. This entailed looking at evaluation reports in the database and categorising them into qualitative, quantitative and mixed methods. Varying issues from the evaluation reports were then identified, categorised and clustered into themes from which relationships were compared across different countries in the region to establish patterns pertaining to evaluation designs, approaches and methods.

Desk reviews on evaluation methodologies were also undertaken to outline the conceptual framework on evaluation types, methods, approaches, and designs used as well as to broaden understanding of what influences the choice of methods in evaluations.

Quantitative data analysis was based on the variables in the database. Once data were inputted and edited, the characteristics of evaluations were described. This was followed by examining the characteristics and determinants of types and selection of the methods that were used. The outcome variable is the choice of method. Some of the independent characteristics existing in the database were as follows: the evaluation types, the donors commissioning

the evaluation, the year of undertaking the evaluation, the type of institution undertaking it, and the scope. This analysis was limited by the number of independent characteristics which were collected in the AfrED database.

## Sampling strategy

The data in the AfrED database is distributed over the years 2005 to 2015. For this chapter, we selected all qualifying evaluation reports for the years 2005, 2010 and 2015. The rationale for the choice was to select the first year (2005) of input, the middle (2010) and the then end (2015). The last year in the database is 2016, hence the choice of 2015. Although the authors note that the database is not exhaustive of all evaluation reports in Africa within the given period, the aforementioned sampling strategy ensures a more representative sample of the AfrED evaluations.

The data were cleaned for the three years and a total of 451 evaluation reports were identified. Reports were deemed to be qualifying if their description, analysis, and presentation were that of evaluation of programmes. From the total population of database entries in the selected years, a total of 113 reports (25%) were disqualified as they did not meet the criteria of inclusion as evaluation reports. These fell into the categories of academic journal articles, summary reports, and annual reports. Additionally, four reports were excluded because the methods used could not be determined.

## Observations and limitations of the data collection process

The research team encountered difficulty with classifying the methodologies used from a number of evaluation reports. In some instances, the evaluation reports did not have a methods section, hence it was not apparent which methods were used. While a number of evaluation reports included a methods/methodology section, there was not always sufficient detail to determine the specific designs being used or how they were used.

## Findings

Findings with respect to the thematic areas identified in the research questions are investigated based on 338 reports obtained from the AfrED database.

The distribution of the report over the years used for analysis are shown in Table 5.2 where we note an increasing number of reports available for analysis over the three years 2005, 2010 and 2015, with just over a half of them drawn from 2015. This may suggest greater availability of reports as the data were being gathered or this may mean that an increasing number of evaluations were being progressively conducted over the years. Given the limitations of the data, it is not possible to conduct a trend analysis.

**Table 5.2** Distribution of evaluations by year of publishing

| Year | No of evaluation reports | Percentage |
|---|---|---|
| 2005 | 74 | 21,89 |
| 2010 | 87 | 25,74 |
| 2015 | 177 | 52,37 |
| **Total** | **338** | **100** |

## Methods used in the evaluation reports

As noted above, one of the key objectives of the chapter is to find out the distribution and use of different types of evaluation methods. From Table 5.3, we note that the largest number of the evaluations was identified as qualitative (55,6%) while quantitative methods were used in less than 10% of evaluation studies.

**Table 5.3** Overall distribution of methods

| Method type | Frequency | Percentage |
|---|---|---|
| Mixed | 121 | 35,8 |
| Qualitative | 188 | 55,6 |
| Quantitative | 29 | 8,6 |
| **Total** | **338** | **100** |

Of the 29 evaluation reports that were identified as quantitative, the designs that were used are depicted in Table 5.4. We note that randomised control trials (RCT) and survey designs were dominant, constituting nearly two thirds of the quantitative methods. The RCTs were primarily conducted in the health and education sectors. This is expected since the two sectors have been dominant in the evaluations space in Africa given that they tend to be resource endowed.

**Table 5.4** The study design of studies for 29 quantitative studies

| Design | Percentage |
|---|---|
| Randomised control trials | 34,4 |
| Survey design | 34,4 |
| Systematic reviews | 24,0 |
| Time series | 3,4 |
| Propensity score matching | 3,4 |

## *Theme 1 Do evaluation types explain the choice of methods?*

As indicated previously in the conceptual framework section, the evaluation types were classified into two categories: formative or summative evaluations. Of the 338 evaluations, 87,6% was summative compared to a far smaller proportion that was formative. Evaluations were defined as summative if their focus was on the outcome, impact, or synthesis. Summative evaluations were mostly outcome and impact studies, while formative evaluations focused on the design, diagnostic, and implementation evaluation. Most formative studies sampled fell under the implementation evaluation category. The data thus shows that evaluations are predominantly conducted at the end of projects to assess project outcomes and impacts, as opposed to the beginning or during the process of evaluation studies (for process purposes).This could be attributed to the fact that summative evaluations were used as tools to help in decision-making about continuation of programmes or projects or renewal at the end of funding cycles.

**Table 5.5** Percent classification of methods by types of evaluations

| Type of evaluation | Mixed methods | Qualitative | Quantitative | Total |
|---|---|---|---|---|
| Formative | 54,8 | 45,2 | 0,0 | 100 (42) |
| Summative | 33,1 | 57,1 | 9,8 | 100 (296) |
| **Total** | **35,8 (n=121)** | **55,6 (n=188)** | **8,6 (n=29 )** | **100 (338)** |

The methods used in the evaluations were associated with the types of evaluations conducted (Table 5.5). There were no quantitative methods that were used for formative evaluations. Mixed methods were adopted both for formative and summative evaluations though a high percentage was used for formative evaluations. Mixed methods typically included surveys with quantitative data collection instruments complemented with in-depth interviews

and focus group discussions for the qualitative component. Therefore, it can be concluded that qualitative and mixed methods were closely associated with formative and summative types of evaluations, yet quantitative methods were used only for summative evaluations.

### *Theme 2 Do donors/commissioners of evaluations define the choice of evaluation and methods?*

Donors have had a significant influence in the development of evaluation in Africa. Issues of accountability and transparency became more evident after the 2005 Paris conference. After the conference, donors placed more stringent regulations on how developmental aid was accounted for, thus putting stricter requirements in terms of its use of M&E. A survey in Benin in 2011 and 2012 showed that principal funding of the M&E budget was provided by Denmark, Germany, Switzerland, UNDP, the EU, the WB and others. In Kenya and Uganda, there has been concern about the government's reliance on donor funding for M&E activities (Cloete et al., 2014:388).

Therefore, it is important to investigate the role that donors have played in commissioning evaluations and the influence that they have had in the choice of methods. In the AfrED database, data were recorded about who commissioned the evaluation, however, the records did not indicate how the method of evaluation was decided upon. Therefore, the donor commissioning the evaluation was used as a proxy for the influence in choice and type of methods selected for the evaluation.

Table 5.6 depicts the wide range of donors in the database and the number of evaluations that they commissioned. The table indicates that Sida, USAID, OECD, UNICEF and ALNAP were the top five commissioners respectively ranking first (17,2%), second (16,2%), third (15,4%), fourth (11,5%), and fifth (11,2%). The rest of the donors commissioned a total of 28,5% with each organisation responsible for less than 10% of the evaluations. It is notable that the database presents that only a fraction of evaluations were commissioned by governments of the region. This confirms the view that evaluations are still largely donor driven.

**Table 5.6** Valuations by donors [funders][1]

| Donor/Commissioner | Frequency | Percentage |
|---|---|---|
| Sida | 58 | 17,2 |
| USAID | 55 | 16,3 |
| OECD | 52 | 15,4 |
| UNICEF | 39 | 11,5 |
| ALNAP | 38 | 11,2 |
| NORAD | 20 | 5,9 |
| Government website | 18 | 5,3 |
| 3ie | 17 | 5,0 |
| UNEG | 10 | 3,0 |
| DANIDA | 7 | 2,1 |
| OTHER | 24 | 7,1 |
| **Total** | **338** | **100** |

Further, it was of interest to find out if research methods were associated with particular donors. Of the 338 evaluations, just more than half were qualitative, 36% classified as mixed-method types and just less than a tenth were quantitative. Of the 29 evaluations that used the quantitative methods, 52% were conducted by 3ie, followed by those that were commissioned by USAID (17%) and governments (14%). The methods used by 3ie in the evaluation reports were predominantly quantitative (Table 5.7). This is aligned to the mandate of the work of 3ie, which is focused on conducting RCTs for impact evaluations. On the other hand, evaluations commissioned by OECD, NORAD, ALNAP and Sida were largely qualitative with at least 60% of their studies lying within that percentage range. The data collection methods were predominantly in-depth interviews; field visits (observations) and focus group discussions (FGDs). There thus remain limited actors specialising in the quantitative space which is typically more resource intense.

**Table 5.7** Methods by the donor/commissioner of the evaluation

| Donor | Mixed methods (%) | Qualitative (%) | Quantitative (%) | Percentage/ Number |
|---|---|---|---|---|
| 3ie | 11,8 | 0,0 | 88,2 | 100 (17) |
| ALNAP | 23,7 | 76,3 | 0,0 | 100 (38) |
| DANIDA | 28,6 | 71,4 | 0,0 | 100 (7) |

[1] Percentages calculated on small figures and are shown for illustrative purposes.

| Donor | Mixed methods (%) | Qualitative (%) | Quantitative (%) | Percentage/ Number |
|---|---|---|---|---|
| Government | 44,4 | 33,3 | 22,2 | 100 (18) |
| NORAD | 20,0 | 80,0 | 0,0 | 100 (20) |
| OECD | 17,3 | 80,8 | 1,9 | 100 (52) |
| Sida | 36,2 | 62,1 | 1,7 | 100 (58) |
| UNEG | 50,0 | 50,0 | 0,0 | 100 (10) |
| UNICEF | 41,0 | 53,8 | 5,1 | 100 (39) |
| USAID | 58,2 | 32,7 | 9,1 | 100 (55) |
| Other | Unclassified | Unclassified | Unclassified | 100 (24) |
| **Overall** | **35,8** | **55,6** | **8,6** | **338** |

Certain donors appear to be associated with certain methods. Of the 10 donors, five (3ie, DANIDA, ALNAP, USAID and UNICEF) undertook evaluations that included quantitative methods. These designs were predominantly RCTs conducted mostly in evaluations commissioned by 3ie and to a limited extent USAID. A few survey designs were noted in evaluations commissioned by 3ie, UNICEF and USAID. Similarly, systematic reviews were only found in evaluations commissioned by 3ie.

## *Theme 3 Does the choice of methods depend on the publication year of evaluations?*

As noted earlier in Table 5.1, over half of reports were conducted in 2015 with an increasing number shown over the years 2005, 2010 and 2015.[2]

Table 5.8 indicates the number of evaluation reports for studies conducted in single countries in the years 2005, 2010 and 2015. For the three years, we note a trend in the way the three different types of methods were being applied in evaluations. Within each year, qualitative methods continued to be dominant. However, it is also noticeable that the mixed methods and quantitative methods each show a steady increase over the years. This is against the steady decrease of use of qualitative methods shown over the same period. This could be partly an increased recognition of the benefits of triangulation and its value in outcome and impact studies amongst the evaluation and research communities.

[2] Given the limitations of not having included some evaluation reports from South Africa in the database, such as the DPME reports, this may have skewed the data.

**Table 5.8** Percent distribution of methods by year of publication

| Publication year | Mixed methods (%) | Qualitative (%) | Quantitative (%) | Total % (n) |
|---|---|---|---|---|
| 2005 | 18,9 | 77,0 | 4,1 | 100 (n=74) |
| 2010 | 33,3 | 60,9 | 5,7 | 100 (n=87) |
| 2015 | 44,1 | 44,1 | 11,9 | 100 (n=177) |
| **Total** | **35,8** | **55,6** | **8,6** | **100 (n=338)** |

## *Theme 4 Does the type of institution define the evaluation and methods choice?*

It was important to investigate the institution(s) that conducted the evaluation once they were commissioned to obtain an overview of who actually conducts evaluations. The institutions were divided into four categories:

1. Local academic institutions
2. International/external academics
3. Non-academic local institutions
4. Non-academic international/external institutions

For definitional purposes, local refers to those evaluations that were conducted by organisations resident in the African continent while international/external comprises of those outside Africa. Local organisations could be from within the continent or offices of international organisations from other continents (likely Europe).

**Table 5.9** Evaluations by institutional type

| Type of institution | Frequency | Percentage |
|---|---|---|
| Local academic | 12 | 3,6 |
| International academic | 21 | 6,2 |
| Non-academic local | 74 | 21,9 |
| Non-academic international | 231 | 68,3 |
| **Total** | **338** | **100** |

Table 5.9 indicates that 75% of evaluations were conducted by international institutions (academic and non-academic) that are based outside Africa. The remainder (25%) came from within the African continent. Non-academic international institutions included individual consultants and small consulting firms.

The dominance of international institutions conducting African evaluations raises questions on the indigenisation of evaluations, which is the promotion of local evaluators within the continent. As such, these findings contribute to the ongoing debate given that local institutions still have low visibility in conducting evaluations. This could be as a result of external evaluation teams being perceived as having more experience and skills than local evaluators. This raises the need to continue to promote and develop evaluators in the continent.

**Table 5.10** Type of evaluation by institution undertaking the evaluation

| Type of institution | Formative (%) | Summative (%) | Total % (n) |
|---|---|---|---|
| Local academic | 8,3 | 91,7 | 100 (12) |
| Academic external | 4,8 | 95,2 | 100 (21) |
| Local non-academic | 14,9 | 85,1 | 100 (74) |
| External non-academic | 12,4 | 87,6 | 100 (231) |

For each of the category of type of institutions, the largest component of the evaluations conducted was summative (Table 5.10). However, the table indicates that local non-academic institutions had the highest proportion of formative evaluations. Summative evaluations which were predominantly outcome and impact studies and tend to require more resources were conducted more by external non-academic institutions. External non-academic institutions dominated both formative and summative evaluations. The fact that the external institutions dominate both types of evaluations shows that external institutions are in the forefront of conducting evaluations much more than local institutions.

An interesting pattern is depicted in Table 5.11. External academic institutions had the quantitative studies as their largest portfolio of evaluations. The local academic institutions have also mixed methods as their largest portfolio. Non-academic institutions both local and external had almost an equal proportion of their evaluations using qualitative methods.

**Table 5.11** Methods used by type of institution undertaking the evaluation[3]

| Type of institution | Mixed methods (%) | Qualitative (%) | Quantitative (%) | Total % (n) |
|---|---|---|---|---|
| Local academic | 66,7 | 25,0 | 8,3 | 100 (12) |
| Academic external | 28,6 | 23,8 | 47,6 | 100 (21) |
| Local non-academic | 39,2 | 58,1 | 2,7 | 100 (74) |
| External non-academic | 33,3 | 59,7 | 6,9 | 100 (231) |

## *Theme 5 Does the scope of the evaluation define the range of methods?*

Further, the scope of evaluations was defined in terms of scope based on the study for which the evaluation was conducted. If the evaluation was conducted in one country, then it was classified as a single-country evaluation. However, if the evaluation was based on a study that was conducted in more than one country, then it was classified as a multi-country study. It was expected that multiple-country studies would be larger in scope than those conducted in only one country, so would be the design and methods. Table 5.12 shows the distribution of evaluations based on the number of countries of coverage. The table shows that just over two-thirds of evaluations were single-country studies.

**Table 5.12** Distribution of evaluations by country

| Number of countries | Frequency | Percentage |
|---|---|---|
| Single-country studies | 228 | 67,45 |
| Multiple-countries studies (2 plus) | 110 | 32,54 |
| **Total** | **338** | **100** |

The single-country evaluations were conducted in countries in East, West and Southern Africa. It is noted from Table 5.13 that the single countries were mostly drawn from East Africa (Ethiopia, Kenya, Uganda and Tanzania), making up 57% of the single-country evaluation studies. West Africa and Southern Africa make up 12,7% and 30,2% respectively.

[3] Percentages calculated on small figures and are merely shown for illustrative purposes.

**Table 5.13** Evaluations conducted by single countries

| Country | Frequency | Percentage |
|---|---|---|
| Namibia | 2 | 0,88 |
| Botswana | 4 | 1,75 |
| Nigeria | 7 | 3,07 |
| Zimbabwe | 12 | 5,26 |
| Rwanda | 14 | 6,14 |
| South Africa | 21 | 9,21 |
| Ghana | 22 | 9,65 |
| Tanzania | 25 | 10,96 |
| Uganda | 25 | 10,96 |
| Kenya | 29 | 12,72 |
| Zambia | 30 | 13,16 |
| Ethiopia | 37 | 16,23 |
| **Total** | **228** | **100** |

Of the 228 evaluations with data on the countries where they were conducted, the largest number of evaluations were conducted in Ethiopia (n=37), Zambia (n=30), Kenya (n=28), Uganda (n=25), Uganda (n=25) and South Africa (21) (see Table 5.14[4]). The use of methods for the countries does not show a discernible pattern. However, countries that conducted evaluations using quantitative methods primarily included Kenya, Uganda, South Africa and Ghana. In most single-country evaluations, evaluators used qualitative as well as mixed methods.

**Table 5.14** Percent distribution of evaluations conducted by single countries[5]

| Country | Mixed methods (%) | Qualitative (%) | Quantitative (%) | Total % (n) |
|---|---|---|---|---|
| Botswana | 100,0 | 0,0 | 0,0 | 100 (n=4) |
| Nigeria | 40,5 | 56,8 | 2,7 | 100 (n=37) |
| Zimbabwe | 50,0 | 50,0 | 0,0 | 100 (n=12) |
| Rwanda | 28,6 | 64,3 | 7,1 | 100 (n=14) |
| South Africa | 33,3 | 52,4 | 14,3 | 100 (n=21) |

[4] Given the limitations of not having included some evaluation reports from South Africa in the database, such as the DPME reports, this may have skewed the data and could have an effect on the way the data are currently presented and interpreted.

[5] Percentages computed on small figures and are merely shown for illustrative purposes.

| Country | Mixed methods (%) | Qualitative (%) | Quantitative (%) | Total % (n) |
|---|---|---|---|---|
| Ghana | 27,3 | 63,6 | 9,1 | 100 (n=22) |
| Tanzania | 40,0 | 60,0 | 0,0 | 100 (n=25) |
| Uganda | 28,0 | 56,0 | 16,0 | 100 (n=25) |
| Kenya | 28,6 | 53,5 | 17,9 | 100 (n=28) |
| Zambia | 30,0 | 63,3 | 6,7 | 100 (n=30) |
| Ethiopia | 40,5 | 56,8 | 2,7 | 100 (n=37) |

The methods of data collection did not show any variation based on whether the evaluations were conducted in one or multiple countries. This is illustrated in Table 5.15. The use of data-gathering methods against the scope of the study did not suggest variability. The percentage use of each of the methods for single-country or multiple-country based evaluations was nearly equal for each of the categories. Therefore, the scope of the study did not appear to define the method that was selected for evaluation.

**Table 5.15** Percent distribution of methods against whether they were conducted in single or multiple countries

| Country | Mixed methods (%) | Qualitative (%) | Quantitative (%) | Total (%, n) |
|---|---|---|---|---|
| Single country | 34,9 | 56,3 | 8,7 | 100 (n=229) |
| Multiple country | 37,6 | 54,1 | 8,3 | 100 (n=110) |
| **Total** | **35,8** | **55,6** | **8,6** | **100 (n=338)** |

## Discussion

This chapter has sought to bring an empirical foundation to questions of methods used in evaluations in the region. This can respond to a range of discussions about the evaluation landscape, including regionally appropriate methods, evaluator capacity-building needs and gaps, and more. The main focus has been on the types of evaluations, the data-collection methods and their determinants. Types of evaluations and methods were defined in a framework that helped to guide definitions and analysis. Further, it was of interest to address the question of human resources in terms of who plays each role in the evaluation process.

While the chapter attempted to establish the determinants of methods choices in evaluations, it was not always possible to infer what leads to methods

selection, more so the implications on evaluation capacity. The findings, however, point towards the capacity and skills in conducting evaluations through indicating the prevalence of specific methods such as quantitative methods for impact evaluations only by specific institutions. Thus this remains an erratic expertise. Impact studies are often expensive and there is still limited capacity to conduct them amongst local evaluators. This can be inferred by the limited local evaluators conducting evaluations despite the constant supply of M&E training courses/skills building initiatives on the continent (Morkel & Ramasobana, 2017). Perhaps this calls for further examining evaluator skills to ensure more targeted approaches towards building skills and building bridges between supply and demand of evaluation.

The data thus provided a starting point for future research to uncover the implications on the methods that African evaluators use in the ongoing evaluator competency debate, i.e. how do methods used/selected support the state of evaluation capacity in Africa?

Formative and summative evaluations were conducted in the region and the methods typically could be classified as qualitative, quantitative and mixed. The majority of the evaluations over three years were summative, clearly demonstrating a focus on whether programmes achieved their long-terms goals. This suggests that the commissioners of evaluations were keenly interested in results of interventions. Given the overwhelming number of the evaluations that were commissioned by donors, it is expected that they would be interested in the effect of developmental interventions that they would have funded. This is especially the case after the 2005 Paris Declaration of the Second High Level Forum on Aid Effectiveness (2005) and subsequent declarations (for example, the 2008 Accra Agenda for Action ) which recognised that aid should aim to produce better impact results. The Paris Declaration was endorsed and formulated around five central pillars of ownership, alignment, harmonisation, managing for results, and mutual accountability.

The evaluation methods were investigated in terms of their determinants, namely the types of evaluations, those who commissioned or conducted them, and the year when the evaluation was conducted. Given the status of the reports, it was difficult to be definitive about how decisions were made in defining the evaluation designs and methods. However, it was noted that data-collection methods were associated with the types of evaluations. At the same time, both formative and summative evaluations included mixed methods.

Further, it appears that the use of mixed methods was gaining momentum between 2005 and 2015, thus casting aside the debate of qualitative versus quantitative methods. As Patton (2015:297) puts it, "the utilisation-focused evaluator works with intended users to include any and all data that will help shed light on evaluation questions, given constraints of resources and time…. The paradigm of choices recognises that different methods are appropriate for different situations and purposes". Therefore, the complementarity of methods seems to be gaining more favour with time.

A topical issue for evaluations conducted in Africa is whether they seek to define and consequently integrate indigenous evaluation methods and practices. The data and methods section of most evaluations tended to make the typical statements that data was gathered using qualitative and quantitative methods which could not be deciphered further. Details of how the communities (as key stakeholders) were brought into participatory engagement to build trust and respect, how the local cultural values, history, and context were understood were well documented. This situation should be viewed against the backdrop that the 75% of evaluations in the database were conducted by institutions from outside the continent. This could be reflecting the fact that most of the evaluations were donor driven. At the same time, this position lends itself to questioning whose interests were being served in the evaluations, given the articulation of the 2005 Paris conference principles about empowerment and involvement. This may potentially also suggest weak stakeholder involvement in the design, planning, and utilisation of the evaluation results. Understanding these dynamics calls for further research as it was not given primacy in this chapter.

A glaring weakness for many of the programmes assessed through the database was that the documentation of the designs and methods were not explicit and lacked articulation of the theory of change for the programme interventions. The researchers noted that reports did not explicitly indicate the theory of change (TOC) for the interventions; that is, how they were expected to work. A programme's TOC is important as this would improve the programme design and consequent monitoring and evaluation. From the perspective of the methods, it would allow the evaluators to assess if the methods were relevant or not. In most of the evaluation reports, these processes were implicit rather than explicit. These gaps meant that many of the evaluations could have a misalignment between the evaluation methods and the evaluation design. There were only a negligible number of evaluations that identified a TOC.

There are a number of other factors that were deemed to influence the choice of evaluation methods. While quantitative methods could be associated with certain institutions, it would appear that most institutions tended to favour qualitative methods or an approach that constituted a mix of qualitative methods. However, a trend over the three years towards an increased use of mixed methods and quantitative methods was noted. Certainly, big donor organisations were associated with quantitative evaluations. For instance, those who used quantitative methods were multilateral international donors (3ie, USAID, UNDP, World Bank and Sida).

Most evaluations were commissioned by foreign donor agencies and the largest proportion of evaluations was executed by international organisations from outside the African continent. That three-quarter of evaluations were conducted by either international academic institutions or international non-academic organisations raise fundamental questions that have dominated the field. Do the evaluators understand the cultural context of the countries where the interventions are applied? Are there no local skills available to undertake the evaluations? It cannot be concluded whether the use of foreign expertise to conduct evaluations is a result of lack of local skills in the overall evaluation field or whether it is a mere case of donors who preferred to use institutions originating from the same country as the donor funds. The move to give value and relevance to Made in Africa evaluations or indigenisation of evaluations in the African continent will have to focus their lens on who is conducting evaluations in Africa and the ability to interpret the results contextually and correctly.

The evaluations did not show any innovation in their use of indigenous knowledge systems. Further, this position seems to be in conflict with one of the central pillars of the Paris Declaration, that of ownership. It could well be that the choice of evaluators from largely donor countries is an issue in the politics of international aid.

Government participation in the process of evaluations was identified to be under represented. The lack of involvement by governments in undertaking evaluations is consistent with the landscape of evaluations in Africa and needs to be addressed. Currently, African states are showing an increasing demand for evaluations, but the funding for the evaluations continues to be largely funded by donors.

## Conclusion

Whilst this chapter has focused on evaluation types and methods, it also highlights the current debates in the development of the evaluation practice in Africa and the evaluation praxis through an overview of methods used in African evaluations as well as what influences the choice of specific methods. It highlights deficiencies in the role of different stakeholders in participating in evaluations, their commissioning, and funding.

The large range of evaluations being conducted in the development sector in the African continent provides a unique opportunity to learn and seek ways to improve evaluation practice. The conclusions that are drawn are compromised by the limitations of the data. However, there is scope for improvement in a number of ways. The evaluations reported in the database did not show any new creative methods beyond the standard approaches that are indicated in Western-based research textbooks. Creativity in the development of indigenous approaches and methods in evaluations was lacking and should therefore be encouraged. Evaluators should go beyond the simple models of qualitative, quantitative, and mixed methods to explore indigenous methods and practices. On the other hand, it was encouraging to note the increased use of mixed methods in evaluation work. Furthermore, mechanisms should be sought for collaboration between local and international evaluators for mutual benefits. Finally, the AfrED data base should be continually improved to include more details of local evaluations to contribute to understanding the evaluation ecosystem in Africa.

## References


Balbach, E.D. 1999. Using case studies to do program evaluation. Working paper. Sacramento, CA: California Department of Health Services. 139-155.

Bamberger, M.; Rugh, J. & Mabry, L. 2011. *Real world evaluations. Working under budget, time, data, and political constraints*. Los Angeles: SAGE.

Barbara, B. & O'Donnell, M. 2016. *Choosing appropriate evaluation methods: A tool for assessment and selection*. London: Bond.

Bresee, L.C. 2014. An introduction to developing surveys for pharmacy practice research. *The Canadian Journal of Hospital Pharmacy*, 67(4):286. https://doi.org/10.4212/cjhp.v67i4.1373

Bryman, A. 1992. *Quantity and quality in social research*. New York: Routledge.

Calidoni-Lundberg, F. 2006. *Evaluation: Definitions, methods and models. An ITPS framework*. Östersund: Swedish Institute for Growth Policy Studies.

Clarke, A. & Dawson, R. 1999. *Evaluation research: An introduction to principles, methods and practice*. Los Angeles: SAGE.

Cloete, F.; Rabie, B. & De Coning, C. (eds). 2014. *Evaluation management in South Africa and Africa*. Stellenbosch: African Sun Media. https://doi.org/10.18820/9781920689513

Cooper, D.R. & Schindler, P.S. 1998. *Business research methods*. Singapore: McGraw Hill.

Corduneanu-Huci, C.; Dorsch, M.T. & Maarek, P. 2017. Learning to constrain: Political competition and randomized controlled trials in development (No. 2017-24). Working paper. Versailles: Théorie Economique, Modélisation et Applications.

DPME (Department of Planning, Monitoring and Evaluation). 2011. *National evaluation policy framework*. Pretoria: Government Printing Works.

Fisher, A.A.; Laing, J.E.; Stoeckel, J.E. & Townsend, J.W. 1998. *Handbook for family planning operations research design*. New York: Population Council.

Forss, K.; Vedung, E.; Kruse, S.E.; Mwaiselage, A. & Nilsdotter, A. 2008. Are Sida evaluations good enough?: An assessment of 34 evaluation reports. *Sida Studies in Evaluation*, 1.

Gorgens, M. & Kusek, J. 2009. Making monitoring and evaluation systems work. A capacity development toolkit. Washington, DC: The World Bank. https://doi.org/10.1596/978-0-8213-8186-1

King, J.A. & Podems, D. 2014. Introduction to professionalizing evaluation: A global perspective on evaluator competencies. *The Canadian Journal of Program Evaluation*, 28:(3):vii-xv.

Morkel, C. & Ramasobana, M. 2017. Measuring the effect of evaluation capacity building initiatives in Africa: A review. *African Evaluation Journal*, 5(1). https://doi.org/10.4102/aej.v5i1.187

Patton, M.Q. 1987. *How to use qualitative methods in evaluation* (No. 4). London: SAGE.

Patton, M.Q. 1990. *Qualitative evaluation and research methods*. London: SAGE.

Patton, M.Q. 2008. *Utilization-focused evaluation*. Fourth edition. London: SAGE.

Patton, M. 2015. *Qualitative research and evaluation methods*. Fourth edition. London: SAGE.

Philips, S. 2018. Diagnostic on the supply and demand of evaluators in South Africa. Diagnostic report. Unpublished consultancy report by Twende Mbele. http://www.twendembele.org/wp-content/uploads/2019/04/TWENDE-DS.pdf

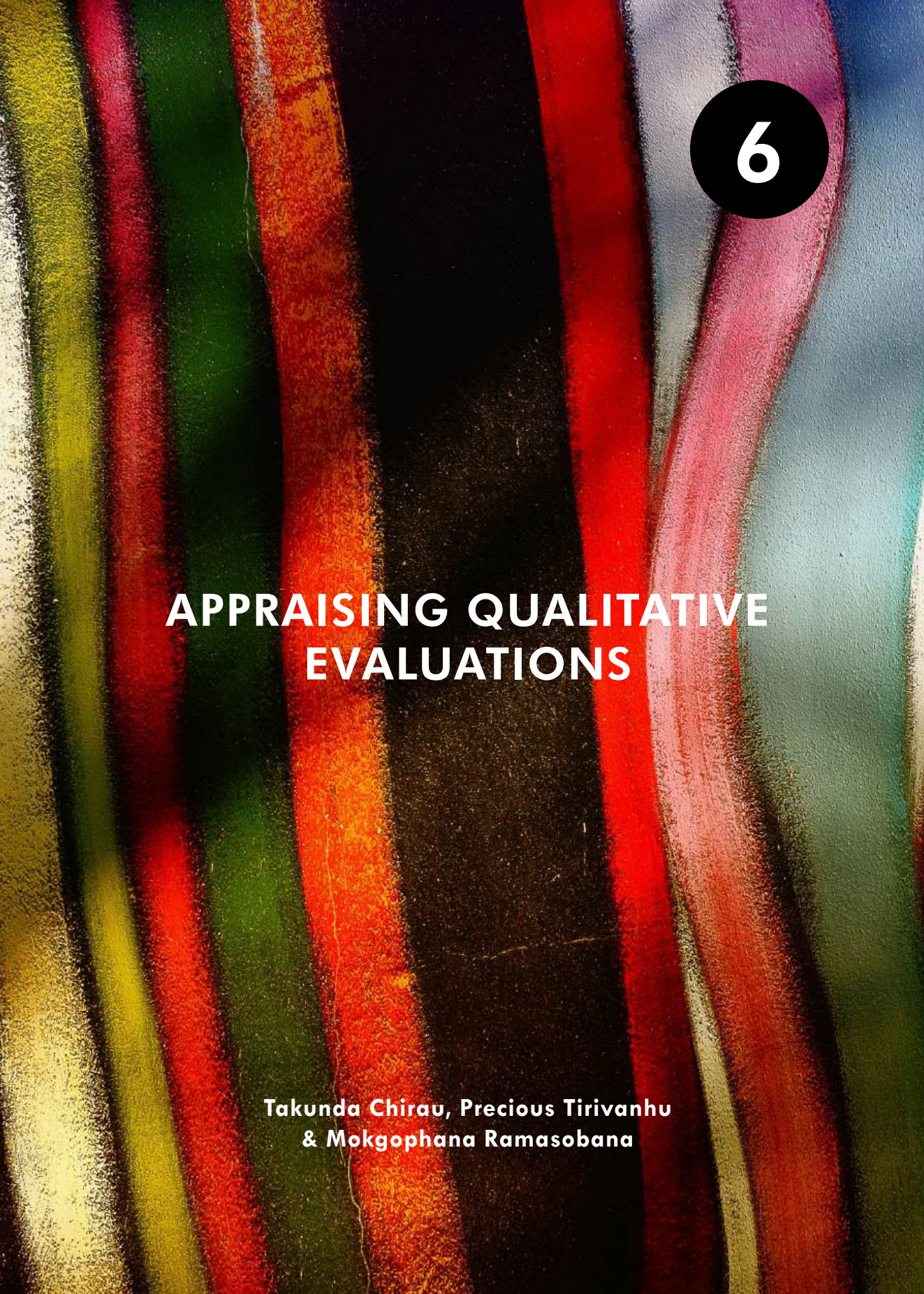
Podems, D. 2014. Evaluator competencies and professionalizing the field: Where are we now? *The Canadian Journal of Program Evaluation*, 28(3):127-136.

Roberts, A.R. & Yeager, K.R. 2014. *Evidence based practice manual. Research and outcomes measures in health and human services*. New York: Oxford University Press.

Stempel, G. & Westley, B. 1989. *Research methods in mass communication*. New Jersey: Prentice Hall.

Tarsilla, M. 2014. Evaluation capacity development in Africa: Current landscape of international partners' initiatives, lessons learned and the way forward. *African Evaluation Journal*, 2(1):13. https://doi.org/10.4102/aej.v2i1.89

USAID (United States Agency International Development). 2008. *Technical note. Conducting mixed-method evaluations.* Washington, DC: Bureau of Policy, Planning and Learning. https://www.usaid.gov/sites/default/files/documents/1870/Mixed_Methods_Evaluations_Technical_Note.pdf



6

# APPRAISING QUALITATIVE EVALUATIONS

Takunda Chirau, Precious Tirivanhu
& Mokgophana Ramasobana

## Introduction

The discourse on quality of evaluations requires more attention in light of the growing focus on what constitutes good evidence and utilisation of quality evaluation results for evidence-based policy and decision-making in Africa. Research on utilisation of evaluation results is not new and dates back to the 1960s.[1] However, in the last decade, the rise of the 'post-truth' era has led to a greater realisation of the need to promote evidence-based decision-making that promote utilisation of objective facts and knowledge (Lewandowsky, Ecker & Cook, 2017:353).[2] Within the evaluation context, such a scenario calls for high quality evaluations since the ultimate goal of evaluation is to generate utilisable results, evidence and knowledge that improves the design and implementation of policies, programmes and projects. In this regard, quality evaluation results are central in evidence-based decision-making. However, quality can be contested. There are accusations that African evaluators do not produce evaluations of sufficient quality. Evaluation reports are of poor quality which in turn impacts on good evidence and utilisation of evaluation findings.

This chapter focuses primarily on assessing the quality of qualitative evaluation reports. The existing quality standards for evaluation studies are primarily on quantitative methods, i.e. theoretical constructs of rigour and robustness informed by positivism paradigm. Qualitative evaluation studies explore subjectivity of the human experience and context which gives meaning to the experience compared to breadth offered by quantitative methods. Increasingly, quantitative methods, especially randomised control trials (RCTs), are being used in evaluations. Evaluation studies are moving towards a mixed-method approach for complementarity. While that is the case, there has not been consensus over what constitutes a good qualitative evaluation study, despite several criteria being set. In the main, the quality of qualitative evaluations literature is quite substantial globally, but there is a dearth of evaluation reports focusing on the quality of qualitative evaluations in Africa. In acknowledgement of this gap, we sought to review a sizeable number of qualitative evaluation reports, as a first step to understanding what constitutes the quality of qualitative evaluation reports.

---

1 Wiess (1967) is considered as the first publication on evaluation utilisation (Dassah & Uken, 2007).

2 The post-truth era is characterised by circumstances in which objective facts are less influential in shaping public opinion than appeals to emotion and personal beliefs. The post-truth world is argued to have emerged because of decline in social capital, growing economic inequality, declining trust in science, and a fractioned media platform.

The chapter seeks to fill this important empirical lacuna in the quality of qualitative African evaluation reports and feeds into the current literature on it. The chapter also hopes to make an analytical contribution by its recognition and appreciation of the significance of a quality framework (QF)[3] with reference to the qualitative evaluation reports in Africa. In this context, the chapter seeks to contribute to addressing meaningfully the ongoing and complex questions about what constitutes good quality for a qualitative evaluation report in Africa.

Using the QF, this chapter attempts to uncover anomalies in already published qualitative evaluation reports by different donors. The core principles of the QF are that it helps in understanding the complexities of qualitative evaluation reports and presents a set of appraisal questions pertinent to guide the action to examine the reports thereof.

The study is organised thematically. It commences by outlining debates around the issue of quality of knowledge from evaluations. Following that is the justification of this study. Thereafter follows a discussion of the Quality Framework which underpins this study. Within that discussion the study discusses the relevance of this framework to the study and what characterises the framework. The methodology section follows after, detailing the criteria of choosing the sample of the study, and development of assessment scores and quality-scoring process. The emerging results are discussed focusing on findings and analysis, ethics, design and methodological limitations, sampling, data collection, reporting, and reflexivity and neutrality.

## Quality and evaluations – a background

A review of the literature cites a number of variables that impact on evaluation utilisation including: negative and null evaluation results (Rog & Bickman, 1984:169); poor quality or irrelevant evidence (Ginsberg & Rhett, 2003); level of engagement between evaluators and intended users; capacity of users to utilise results; timeliness and relevance of information; methodological rigour and quality of presentation; validity and credibility of results (Connolly & Porter, 1980); and quality of evaluations.

[3] Quality Framework (QF) after Spencer, Ritchie, Lewis & Dillon (2003).

Evaluations should provide credible, useful evidence-based information that enables timely incorporation of its findings, recommendations and lessons into the decision-making process of organisations and stakeholders (UNFPA, 2017). Although there has been considerable research on the correlation between evaluation quality and utilisation of evaluation results (see, for example, Johnson, Greenseid, Toal, King, Lawrenz & Volkov, 2009:383; Bamberger, 2004; Bober & Barlett, 2004; Johnston, 1986; Rockwell, Dickey & Jasa, 1990; Shea, 1991; Potts, 1998; and Burry, Alkin & Ruskus, 1985), evaluation quality studies are still limited in the African context.

The global gold standard for ensuring quality evaluation has been based on the Development Assistance Committee (DAC) guidelines developed by the Organisation for Economic Cooperation and Development (OECD) in 1991 (OECD, 2010; Chinanca, 2008). The OECD/DAC criteria refer to the relevance, effectiveness, efficiency, impact, and sustainability of development interventions. However, the quality of evaluations from development aid guided by the DAC framework has been questioned and the need for continuous improvement has been highlighted by some evaluation practitioners and researchers (Ofir, 2017; Chianca, 2008; Chianca, 2007; ALNAP, 2006). Some of such evaluations are viewed as methodologically weak and the use of the DAC criteria is often mechanistic, excluding creativity in evaluation processes (ALNAP, 2006).

In addition, it is argued that the DAC criteria place limited focus on the needs of the beneficiaries and give limited focus on the equality of the implementation process and exportability of programme/project design. Scholars in the evaluation space view the DAC criteria as not being sufficient to determine the worth, significance or merit of an intervention for development but rather focuses on accountability for donor funding (Chianca, 2007:45). In addition to the DAC criteria, some developing agencies have their own complementary in-house evaluation quality-assurance systems (see, for example, UNFPA, 2017; UNEG, 2016; WFP, 2015). These highlighted quality-assurance (QA) criteria aim at ensuring quality knowledge from evaluations. The QA system varies from one method to another, i.e. qualitative and quantitative studies.

In Africanising evaluation quality, the African Evaluation Guidelines (AEGs) established a list of 30 attributes of good quality evaluations. This is presented in the form of a checklist that can be used to assess and improve the quality of evaluations. They are based on the Programme Evaluation Standards used by

the American Evaluation Association (Patel, 2013). The evaluation guidelines are clustered into four categories: utility, feasibility, propriety, and accuracy. Utility guidelines are intended to help to ensure that an evaluation serves the information needs of intended users and be owned by stakeholders; feasibility ensures that an evaluation will be realistic, prudent, diplomatic, and frugal; propriety ensures that an evaluation will be conducted legally, ethically, and with due regard for the welfare of those involved in the evaluation, as well as those affected by its results; and accuracy ensures that an evaluation will reveal and convey technically adequate information about the features that determine worth or merit of the programme being evaluated.

In addition to the AEGs, some national governments have put in place systems for ensuring high standards for evaluations. For example, the Department of Planning, Monitoring and Evaluation (DPME) in South Africa, Ministry of National Development and Planning in Zambia, Office of the Prime Minister in Uganda, Ministry of Devolution and Planning in Kenya, and Ministry of Budget and National Planning (MBNP) in Nigeria has instituted evaluation standards which every government evaluation should adhere to. Most of the evaluation standards, principles, and ethics are captured in country's national evaluation policy, and at times they are instituted as separate guidelines promoting quality evaluations across government departments. This is the case with South Africa and Zambia.

National monitoring and evaluation systems in different countries support the use of evaluations conducted through the national evaluation system by setting benchmarks of evaluation quality (DPME, 2014). Quality assessment of evaluations is conducted as a routine activity of the South African National Evaluation System (NES) (Leslie, Moodley, Goldman, Jacob, Podems, Everett & Beney, 2015). In Kenya, the County Integrated Monitoring and Evaluation System (CIMES) has provisions for quality assurance through the County Monitoring and Evaluation Committee (CoMEC). The committee provides quality assurance by verifying whether the M&E information given in the reports and the underlying data collection and analysis processes are of the needed quality and conform to the national M&E requirements (MDPMED, 2016).

An exploration of literature on evaluation-quality assessments indicates a dearth of such studies in the African context. Very few examples are cited including the Leslie et al. (2015) study that audited quality of evaluations in the South

African public sector. The assessment of UNICEF and UNAIDS evaluation using the African Evaluation Guidelines (Patel, 2002) based on utility, feasibility, propriety and accuracy. Within this context, the question becomes: What is the evaluation assessment trajectory in Africa?

## The quality assessment framework

This chapter is informed by the Quality Framework (QF) by Spencer, Ritchie, Lewis and Dillon (2003). The framework was originally applied to government-funded evaluation studies but has been more recently applied to donor-funded evaluation studies and generally on qualitative research irrespective of evaluative purpose. The framework was established with the aim to interrogate the quality of qualitative evaluation studies. The birth of the Quality Framework (QF) can be traced back to Britain.

The QF was developed as part of the research that was commissioned on behalf of the Cabinet Office in London. In developing the framework, there was a lot of consultation with secondary data on qualitative research methods relating to standards in qualitative research, review of existing frameworks for assessing quality in qualitative research, interviews with different practitioners interested in quality assessment of qualitative research and policy related evaluations, academics, research practitioners, commissioners, policy makers and funders.

## The relevance and application of the Quality Framework

The QF has been established based on certain premises, i.e. the nature of qualitative enquiry, how it can be used for evaluative purposes and how its quality can be examined. These premises are derived from extensive debates in secondary data on qualitative research and evaluation. Quality standards in qualitative research or evaluation have always been a point of contestation amongst researchers and academics. Rigour is one main point of contestation. It relates to the robustness of qualitative research/evaluation evidence, relevance and utility.

There is a dearth of Africanised frameworks for assessing quality of qualitative evaluation or research. However, this does not deter the utilisation of the QF as it fits within the context of what is being examined. The bulk of frameworks

are borrowed from the West, for example, quality framework for qualitative research (Meyrick, 2006), consolidated criteria for reporting qualitative research (COREQ) (Tong, Sainsbury & Craig, 2007) and total quality framework (TQF) for qualitative research (Roller & Lavrakas, 2015).

Of equal importance to the adopted QF for this study is the COREQ which provides a 32-item checklist specifically for interviews and focus groups (divided into three broad categories of research team and reflexivity, study design and data analysis and reporting) (Tong, Sainsbury & Craig, 2007). Whereas the QF encompasses four techniques, namely interviews, focus groups, observation, and documentary analysis (Spencer, Ritchie, Lewis & Dillon, 2003:6). The QF is not a 'cut-stone' framework. It could be modified to suit other data-collection techniques as long as there are additional appraisal questions. The frameworks identified herein, substitute and complement each other as they vary in key areas of examining or assessing quality of qualitative evaluations or research. There is, therefore, no comprehensive framework for assessing quality of both qualitative and quantitative studies.

The QF has gradually become renowned for its influence as an integrating analytical structure for coming to grips with the complexity of qualitative evaluation studies, understanding the influence it has on utilisation or usability of reports thereof. The success of the QF in winning the attention of researchers, evaluators, commissioners, and policy makers arose from the fact that it presents a fresh vision of a holistic and integrative approach with the capacity to analyse and understand the complexity of quality of qualitative evaluation studies. Hence, the QF animates this chapter and deepens the understanding of the quality of qualitative evaluation studies.

## Content of the framework

The QF is based on 18 appraisal questions (column A).[4] The appraisal questions addresses all the critical features and processes deemed necessary for qualitative enquiry. The structure of the QF is divided into assessment of findings and navigates through the research process which includes the design, sampling, data collection, analysis, and reporting. Other general features animating the QF are reflexivity, neutrality, ethics, and auditability. The QF acknowledges that it is not logical and consistent with the research process to start with 'findings'.

[4] See the Quality Framework attached as an annexure to the chapter.

However, the findings were given the first priority to allow an investigation into the research process, i.e. quality of data collected, logic of analytic process (Spencer et al., 2003).

Column B presents the quality indicators of the framework. The indicators are important to providing in-depth information necessary to examine whether or not the quality is assured. Many of these indicators apply to specific data-collection techniques, like focus groups, observation and documentary analysis. Column C is intended to be a notes space on the study being appraised. Although the QF is important to this chapter, there are some limitations to it. There are duplications of the 18 criteria, particularly with respect to the quality indicators. As such, some of the indicators are collapsed to avert any duplication.

## Methodological approach to the study

### *Sampling strategy*

This study utilised a five-stage stratified random-sampling procedure. Such a process was necessitated by the inherent biases in the AfrED database.[5] The stratification process partitioned the population into disjoint sub-populations which were more homogeneous than the original population. Proportional sampling within the sub populations allowed a more representative sample that was finally utilised by the chapter. The five-stage stratification process is illustrated in Figure 6.1.

The first-stage process focused on extraction of qualitative reports from the AfrED database. A total of 162 reports were extracted forming the study population (n=162). The subsequent stages involved stratification of reports by funding agencies, country where they were commissioned, sector, and type of evaluation. The chosen donors operating from different sectors were Sida (12) and USAID (8). Proportional random samples were then drawn from the final set of strata (evaluation type) guided by Cochran's sample size formulae for categorical variables (Bartlett, Kotrlik & Higgins, 2001). A final sample size of n=20 was utilised in the study. The final sample size was seen as the most corresponding to the QF that is adopted as analytical framework of this chapter. Due to data saturation, there was no point in reviewing other evaluation reports.

[5] The database is based on countries, funding agencies for evaluations, and sectoral focus groups.

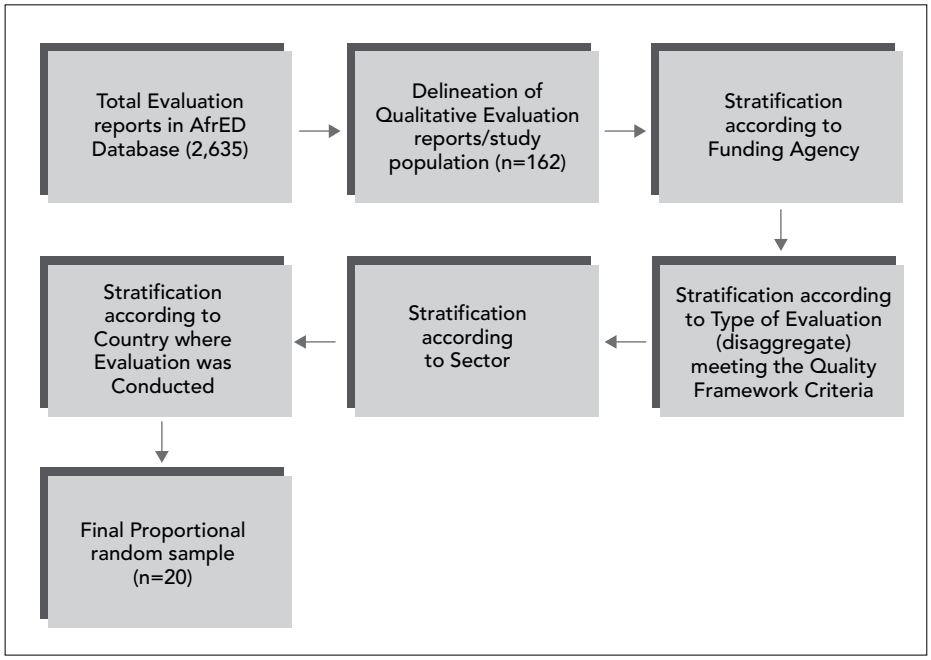



**Figure 6.1** Multi-stage stratified sampling process

## *Development of assessment scores and the quality scoring process*

A five-point ranking scale was developed for assessing each of the 18 criteria outlined in the QF (1=Poor; 2=Fair; 3=Good; 4=Very good; and 5=Excellent). In qualitative research, measures with five-point scales have been shown to create variance that is necessary for examining the relationships amongst items and scales and create adequate coefficient alpha (internal consistency) reliability estimates (Hinkin, Tracey & Enz, 1997). To ensure objectivity and validity of the scoring system, the authors thoroughly discussed the quality-assessment criteria. Investigator triangulation (Flick, Von Kardorff & Steinke, 2004:178) was conducted through peer re-evaluation of scores by each of the authors. A four-point ranking scale was utilised for assessing the total scores of the evaluation reports (18-36=Below Standard; 37-54=Fair Quality; 55-74=Medium Standard, and 75-90=Good Standard). The QF has 18 appraisal questions, and a five-point scale was used to assess each of the 18 criteria outlined in the QF (see Annexure). Each report was, therefore, scored against a total of 90 points.

## *Analysis of findings*

The QF highlights the theme of findings as its starting point to analysing the quality of qualitative evaluation studies. The framework raises some fundamental and critical questions around the credibility of findings. In particular it appraises how knowledge/understanding has been extended by research and how the evaluation addresses its evaluation questions. The data collected from the evaluation reports sampled for this study reveals that the process of analysing findings was well articulated. Findings of the analysed reports 'seem' credible as they answer the evaluation questions. Of the 20 qualitative evaluation studies examined, 19 (with a mean score of 4,4) show that the findings are trustworthy.

To a larger extent, there is evidence of how the evaluators arrived at conclusions; this is explicitly captured in almost all the evaluation reports sampled. With respect to this, 19 evaluation reports were rated as good and excellent on findings and analysis. One example of the evaluation is the Kenya Peace Initiative report evaluation findings that are thematised in the five proposed evaluation questions. The structure of the Kenya Peace Initiative reports allows the audience to have a deeper understanding of the findings for each of the proposed five evaluation questions. The conclusion is consistent with the proposed evaluation questions; there is a coherent logic.

A question of how the knowledge/understanding has been extended by the evaluation lacks depth in a number of evaluation reports examined. De Vos, Strydom, Fouché and Delport (2011:402) argue that the findings must be verified against the literature by embedding it in larger theoretical perspectives or paradigms. Data gathered indicates that more than ten evaluation reports have findings which are not well corroborated with the existing body of knowledge or literature. It remains unclear whether the evaluators were more interested in grounded theory, grounded in the data collected in the field, rather than taken from the literature review (Leedy & Ormrod, 2005:140). However, evaluation reports, for example, the end of project evaluation of the Echo-funded GoK/UNICEF Response to the Kenyan Drought Emergency (2005) and the Performance Evaluation of Reading for Ethiopia's Achievement Developed Institutional Improvement (AIR, 2015), are findings discussed in conjunction with the existing knowledge or literature review. It was noticeable within the analysed reports that the findings and analysis reinforce what is already known and to some extent elicit fresh knowledge.

Recording interviews is crucial to ensure storage of rich and original data which the interviewer cannot write down without stopping the flow of conversation and breaking the rapport with the respondent (Bowbrick, 2015:1). Audio-tape-recorders are a useful tool for storing data. The data gathered reveal that the evaluation reports did not document recording of data to preserve the accuracy of information collected. Furthermore, the reports are silent on giving the reasons why there was no recording of the interviews or focus group discussions, yet qualitative methods were used and huge amounts of data were collected. There is a possibility that during data-gathering processes, valuable information could have been lost. Hence the majority of reports do not use verbatim transcripts to support their arguments; this might have arisen from the lack of recording. However, other evaluation reports made use of recorders, for example, the Performance Evaluation of Reading for Ethiopia's Achievement Developed Institutional Improvement (AIR, 2015) and Sida evaluation on organisations in the field of HIV/AIDS and Human Rights (2007). The aforementioned evaluation reports made use of direct quotations, thereby increasing trustworthiness of their findings. Stuckey (2014) argues that verbatim transcripts are necessary to establish reliability, dependability, and trustworthiness of a study. Furthermore, qualitative research is epistemologically grounded in social constructivist, symbolic interactionist or other interpretive conceptual frameworks. The richness of the data can be noticeable through inserting most relevant or appropriate direct quotations. To reiterate, the READ II evaluation done in Ethiopia asked the question, "*How do parties involved rate their satisfaction with the government-to-government*". Social constructivism indicates that knowledge is constructed through human interaction. The quotation below from READ II is important in that regard:

> The satisfaction of the trainers and trainee was conspicuously high as we observed during the supervision and management support visits at the two rounds of training. All trainees were actively engaged in and enjoying the training. They were sharing their experiences and learning from each other as well. (AIR, 2015:12)

The above quotation demonstrates that there was high satisfaction. Presenting verbatim transcripts provides what Max Weber called *verstehen*, i.e. the perspective of a respondent, thereby providing the most fertile array of data.

## *Ethics*

Ethics is another component of the QF. Ethics is a set of moral principles which is suggested by an individual or group, is subsequently widely accepted, and

offers rules and behavioural expectations about the most correct conduct towards subjects and respondents (De Vos, Strydom, Fouché & Delport, 2011:114). Therefore, ethics is a core aspect of evaluation practice and brings unique moral dilemmas to the fore. De Vos et al. (2011:113) argue that ethical issues are pervasive and complex, since data should never be obtained at the expense of the human beings. This makes it an obligation to document how ethical issues were manoeuvred. The data gathered from the evaluation reports reviewed indicate that reports dealt with human subjects. A majority of these human subjects are 'vulnerable groups'. According to Burns and Grove (2003) vulnerability increases the need for justification for the use of such subjects. Data gathered for this study indicate that the evaluation reports are silent on the aspect of ethics. A search of the word 'ethic(s)' revealed that only two of the sampled reports outlined an approach to ethics. When evaluators undertake evaluations, they have to adhere to at least three value systems, namely community (respecting human rights), evaluation (a practice culture based on the ethic of doing no harm to subjects, and science (respecting the rules of scientific enquiry).

There is no evidence pointing to these value systems in the reviewed reports. Of the 20 sampled qualitative evaluation reports, 18 of them were rated on 1 (below standard), while two reports were rated on 5 (excellent). The two reports are Performance Evaluation of Reading for Ethiopia's Achievement Developed Institutional Improvement (AIR, 2015) by USAID, and USAID/ Uganda Strengthening Decentralization for Sustainability (SDS) Program (Kibombo et al., 2016). Ethical considerations such as informed consent, confidentiality and anonymity were adhered to. The respondents read and signed the informed consent form. In some instances, consent was approved through raising of hands. Armiger (1997) argues that informed consent is the major ethical issue in conducting research and evaluation as "it means that a person knowingly, voluntarily and intelligently, and in a clear and manifest way, gives his or her consent". What is quite disturbing is that some of the evaluations dealt with sensitive matters which would evoke emotional feelings of the past and present. Some evaluation reports dealt with gender-based violence and HIV/AIDS with a possibility that negative past behaviour might have been recalled. Often, when respondents provide rich and detailed data, special care should be taken that work is done in an ethical manner (Silverman, 2000:201). To sum up, the findings clearly indicate that ethical considerations are not sufficiently understood in evaluations. We believe

that Made in Africa evaluators must be cognisant of ethical considerations in evaluation to inform their work in a more systematised, professional manner. Furthermore, this must be institutionalised to protect vulnerable groups.

### *Design and methodological limitations*

Design is the other component of QF and raises fundamental and critical questions of "how defensible the research design is". Other questions supporting the appraisal question are concerned with how the overall research strategy was designed to meet aims of the study, discussion of the rationale of the study design, and discussion of limitations of research design and their implications on findings. It is important to highlight the fact that the design of an evaluation influences the quality of the findings. Burns and Grove (2003:195) defines a research design as "a blueprint for conducting a study with maximum control over factors that may interfere with the validity of the findings". Data gathered indicate that all the evaluation reports provided their evaluation/research design. However, three evaluations reports were rated fair, twelve evaluation reports were rated good, two evaluations were rated good and three reports were rated excellent. The reports which did not thoroughly provide reasons why they have chosen the design to answer their evaluation questions were rated between fair and good.

The Feed the Future Ethiopia External Mid-Term Performance Evaluation Report by USAID (2015), Institutional Cooperation between Furniture Technology Centre, South Africa, and Tibro Training Centre by Sida (2007) and Kampala City Council's A Project for Promoting Ecological Sanitation in Kampala, Uganda, by Sida (2008) are the reports which did not thoroughly outline their evaluation design. Outlining the evaluation design reduces inaccuracy and helps to get maximum efficiency and reliability. With respect to the evaluation reports that were rated low on this component, there were no logistical procedures of how the research design would be executed. Kumar (2005:195) rightfully argues that "one characteristic of a good study design is that it explains the details with such clarity that if someone else wanted to follow the proposed procedure, he or she would be able to do exactly as the researcher had done". This is not the case with a handful of evaluation reports examined in this study.

The findings from the evaluation reports also indicate that the methodological limitations are not thoroughly addressed. A majority of the evaluation reports did not address how the limitations could affect the internal and external validity

of the results. This is the case with the UNICEF Summative Evaluation of the Echo-funded GoK/UNICEF Response to the Kenya Drought Emergency (2005) and USAID End-Line Evaluation of Peace Initiative Kenya (PIK) by UNICEF (2016). A few of the evaluation reports acknowledged how limitations could have affected the interpretation of evaluation findings. The Final Performance Evaluation of Local Government Decentralisation Programme in Ghana by UNICEF (2015) and Evaluation of Swedish Health Sector Programme Support in Uganda 2000–2010 by Sida (2011) highlighted their methodological limitations. This justifies the conclusions arrived at. The UNICEF report was intended to measure impact of the programme. Hence, to avert co-intervention bias, the evaluation team had to thoroughly conduct a review of the environment to understand programmatic overlaps as well as potential synergies. The Sida report, in turn, acknowledged the unavailability of most reliable health data, i.e. maternal mortality rate (MMR), the neonatal mortality rate (NMR), the infant mortality rate (IMR), and other core health indicators. Puhan, Akl, Bryant, Xie, Apolone and Ter Riet (2012:2) argue that "authors should acknowledge recognised limitations and discuss their likely implications on the interpretation of the findings doing so, they reduce the probability that readers will misjudge the validity and impact of their study". This concurs with De Vos et al. (2011:111) who argue that even the most carefully planned studies ought to list the limitations of the study. We speculate that problems are not completely eliminated in a study, therefore there is a need for documenting them and their implication on the findings.

### *Sampling*

Another component of the QF is sampling. This component answers the fundamental questions of how defensible the sample design/target selection of documents and cases, and sample composition is. With respect to sampling, qualitative research or evaluations use purposive non-random sampling. Patton (2015:264 and Bartlett, Kotrlik & Higgins, 2001) argues that the logic and strength of purposeful sampling lies in selecting information-rich cases for in-depth study. The rich information cases are those from which a researcher can learn in-depth issues of central importance to the enquiry. The data collected from the sampled evaluation reports indicate that the Sida, UNICEF and USAID (not representative of all evaluation reports examined) qualitative evaluation studies reflects a brief description of their sampling strategy. However, there is no in-depth justification or defence of the chosen data sample and the process

of selection thereof. Out of the selected evaluation reports, fourteen of them were rated between poor, fair and good.

It is clear from reviewing the reports that some failed to account for the missing coverage in their achieved sample and the implications thereof. Walsh and Downe (2006) argue that a rationale for justification of the selection of participants in a study is necessary. Similarly, Sandelowski (1995:179) argues that it is important to provide a reason for the number of sample participants; too few may risk adequate depth and breadth, but too many may produce a superficial or unwieldy volume of data. Furthermore, the findings indicate that six reports thoroughly justified their sampling. The Performance Evaluation of Reading for Ethiopia's Achievement Developed Institutional Improvement (AIR, 2015) by USAID highlighted the sampling strategy and how the actual numbers were achieved using purposive and random sampling. Without the justification of the sampling, it raises a fundamental question of how they arrived at the final sample of respondents (the targeted vs the achieved) despite that almost all the reports annexed the list of respondents. Therefore, the notion of representativity is questionable to the majority of the analysed evaluation reports.

### *Data collection*

Data collection is another component of the QF. It raises some critical question of how well the data collection was carried out, and discussion of how fieldwork methods or settings may have influenced data collection. The findings of the review of the qualitative evaluation reports indicate that the majority of the evaluation reports outlined the various data-collection techniques or methods. However, the reports failed to adequately address a thorough discussion of why the chosen methods for data collection were employed. Griffith (2017:595) argues that no evaluation would be complete without describing the procedures used. There is need for evaluators and researchers to furnish audiences with a thorough data-collection description and not assume that a particular technique will be emphatic enough to provide relevant information. In qualitative research, it is important to determine the extensiveness of data-collection techniques, processes, breadth, depth, and scope (Sabol, 2001). How well the data collection was carried out is not well documented in the evaluation reports. This has implications on replicability, if ever an evaluation of that nature is conducted again.

Furthermore, data-collection techniques inform the kind of evidence that will be collected to make appropriate judgements of the policy or programme. Without such documentation, it can be difficult to ascertain how a conclusion was reached. As a result of the failure to discuss the techniques, fourteen reports were rated as poor, fair and good while six were rated as very good and excellent. While, this is the examined trend across the evaluation reports, some evaluation reports outlined their methods of data collection and provided a justification. For example, the Final Performance Evaluation of Local Government Decentralisation Programme in Ghana by UNICEF (2015) indicated that the rational for triangulating the several methodologies was to strengthen the validity of the findings, minimise bias and answer the evaluation questions explicitly.

Furthermore, there is no description of who collected the data, procedure for collecting the data and how fieldwork may have influenced the data collected. Veal (2011) argues that "a thorough reporting of the process and the results of qualitative data collection and analysis is the key to justifying and assuring that trustworthiness exists in the study".

### *Reporting*

The reporting component of the QF deals with the questions of how clear the conceptual framework links between analytic commentary and presentations of data, and how clear and coherent the reporting is. The quality of reporting in qualitative evaluation reports has impact on application of qualitative results in decision-making (O'Brien, Harris, Beckman, Reed & Cook, 2014; Tong, Sainsbury & Craig, 2007). Analysis of qualitative reports indicated a discernible link between the conceptual, methodological, and analytical framework in most of the analysed reports. There explains to some degree why some conclusions were arrived at. Examples include the Peace Initiative Kenya (PIK) end-line performance evaluation (2016) which provides excellent linkages between the conceptual, methodological, and analytical components in reporting. However, some reports such as the Sida-funded evaluation of organisations in the field of HIV/AIDS and Human Rights (2007) lacks flow of the conceptual and analytical framework because there are no explicit evaluation questions to guide discussions, findings, and recommendations. Additionally, explicit research questions enhance justification of the research design (Farrugia, Petrisor, Farrokhyar & Bhandari, 2010:279) and allow assessment of the validity of the evaluation results.

This is important in increasing the possible use of the evaluation findings for lessons and improvement of future programmes and projects. There is poor display of data integrity in some reports as a result of failing to demonstrate the link with specific aspects of the data with illustrations. It is important to allow the audience to validate the results and improve validity (White, Woodfieled & Ritchie, 2003). Most of the reports show poor coherence, i.e. presentation of results in a way that guides the reader. This involves displaying diversity of results and the inclusion of non-dominant phenomena as opposed to largely focusing on recurring themes in conclusions and recommendations.

### *Reflexivity and neutrality*

Reflexivity defines the awareness of the influence the researcher and evaluator have on the people or topic being studied, while simultaneously recognising how the evaluation experience is affecting the evaluator or researcher (Gilgun, 2008). An analysis of the selected qualitative reports indicated that most of the evaluations that were analysed were conducted by evaluators from outside of Africa, mostly Europe. Given such a scenario, it becomes apparent that contextual issues need attention with regards to the impact of epistemological differences on the generation of relevant Africa knowledge. The issue of using external evaluators and its implications on the reliability and validity of evidence or data continues to receive attention by African evaluators (see, for example, Chilisa & Malunga, 2014). Given these kinds of contestations, the need for integrating reflexivity in evaluations becomes apparent and involves the need by evaluators to explore their assumptions, emotional reactions, and cultural positioning, through specified actions. Although it may be argued that reflexivity may be practised but not made explicit, documenting may improve credibility, i.e. internal validity through demonstrating that a true picture of the phenomenon under scrutiny is being presented (Shenton, 2004).

Although some of the USAID evaluation reports are mainly anchored by internal evaluation practitioners the failure to explicitly outline reflexivity leaves no learning on the local evaluation practitioners. An example is the Outcome Evaluation of Reading for Ethiopia's Achievement Developed Institutional Improvement, which was prepared by the American Institutes for Research (AIR, 2015). Probst (2015:42) identifies some of the benefits of reflexivity to include: accountability, trustworthiness, richness, clarity, ethics, support, and personal growth. Additionally, it strengthens the integrity of the evaluation process, the quality of the knowledge generated, the ethical treatment of those being studied, and the evaluators' personal growth.

Making the reflexivity strategy explicit can play a key role in what Guba and Lincoln term "progressive subjectivity", or the monitoring of the researcher's own developing constructions, which helps in establishing the credibility of evaluators (Guba & Lincoln, 1989).

### *Discussion*

This study sought to assess the quality of qualitative evaluation studies using the Quality Framework as an analytical framework for the examined evaluation reports. The preceding sections provided findings and discussion in relation to the Quality Framework. De Vos et al. (2011:428) argue that qualitative "might be viewed with more scepticism than a quantitative report". To that end, the reviewed reports lack detailed documentation of the account of credible procedures to arrive at the conclusions. The preceding discussion has outlined some of the means of certifying credibility. Morris (2008) argues that systematic enquiry of evaluation emphasises the methodological dimension of evaluation; it is significant to explain and communicate methods and approaches in a sufficient and clearer manner to the extent that others can understand and critique the work. It is important that the "logic of conducting evaluation or even research is outlined" in a clearer way. Neuman (1994:474) argues that it is key "to provide readers with enough evidence so that they believe the recounted events and accept the interpretations as plausible". There is, therefore, no commonly agreed way of writing a qualitative evaluation report. Of course, there are acceptable guidelines for an evaluation report from one donor to another. A number of authors concur with the assertion that the methodology and the conceptual frameworks are equally important. The selection, particularly of methodology and detailing thereof, influences first the nature of the report and second the audience in accepting findings as epistemic. This study is by no means dismissing the analysed reports as not credible, but it raises critical and fundamental questions regarding credibility and trustworthiness. However, other scholars may differ with the standpoint of this study. Good quality evaluation reports are based on the notion of consistency, completeness and transparent reporting, while lack thereof threatens the credibility of evaluations (Weiss, 1998).

To reiterate, the evaluation reports were rated in terms of quality. A three-point ranking scale was utilised for assessing the total scores of the evaluation reports (18-36 = Below Standard; 37-54 = Fair Quality; 55-74 = Medium Standard, and 75-90 = Good Standard). Using the highlighted scale, the findings indicate

that there was no qualitative evaluation report rated between the scores of 18-36. One evaluation report was rated between the scores of 37-54, sixteen evaluation reports were rated between the scores of 55-74, and three evaluations are rated between the scores of 75-90 (see Figure 6.2).

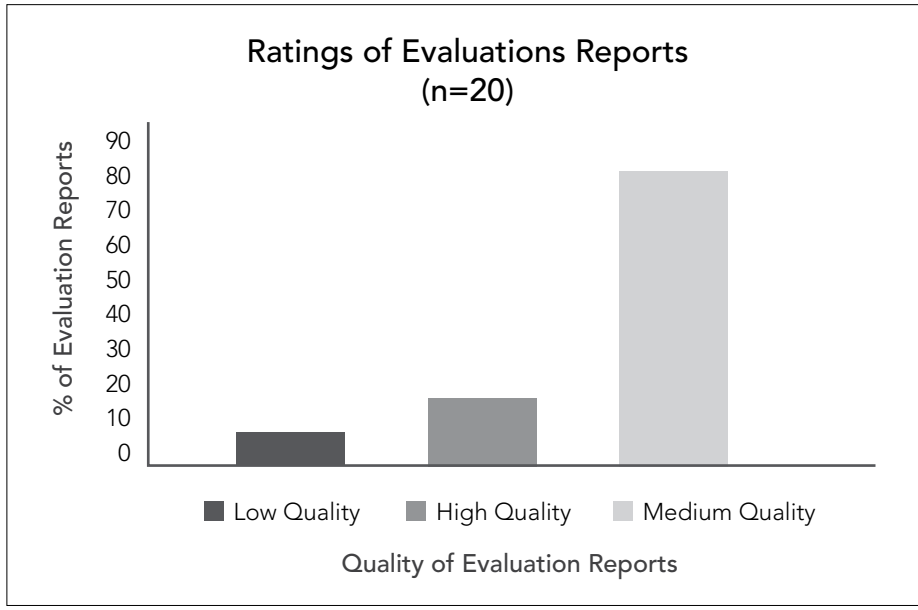



**Figure 6.2** Quality rating of evaluation reports

Figure 6.2 shows that the majority (80%) of the qualitative evaluation reports are of medium quality, 15% of reports are of high quality, and 5% are of low quality, based on the QF utilised for this study. The results show that there is lack of consensus on the logic of crafting a quality qualitative evaluation report. There are varying conceptualisations on what appraisal questions should be answered in crafting an evaluation report.

## Conclusion

The chapter set out to do an analysis of the quality of qualitative evaluation studies in the AfrED database. It is clear from the findings that the quality of the qualitative evaluation studies is diverse, even with the same donor or funder. The logic of evaluation within the analysed qualitative evaluation studies indicates that there is no consistency in the way the evaluation reports are crafted. To that end, the evaluation reports funded by the same funder or donor are diverse in the way they are written. Based on the quality framework adopted for this

study, we found that reports from the same donor scored different scores for instance on justification of data collection. This reflects the lack of consistency in the way quality is maintained from a donor/funder perspective. The analysed reports also indicate that there are some processes which are not highlighted in the evaluation report, for example, ethics. Ethics is a core aspect of evaluation practice. It is important to flag this in the report for the audience to understand that ethical considerations were adhered to by the evaluation team. As a result, numerous reports were ranked of low quality on this point.

The empirical findings reflect that there is a general tendency that current evaluation studies do not place sufficient weight (in terms of empirical evidence) and use it in conjunction with the existing knowledge. As a result, the reports do not cover sufficient depth. This is manifested, for example, in the majority of lowly ranked evaluations that do not fulfil reporting guidelines prescribed for complete and clear evaluation report.

With respect to utilisation, it is difficult to speculate if the reports were utilised (or not) given the nature of this study. Utilisation of the evaluation reports therefore calls for further assessment from the perspective of the people who were intended to use them. Therefore, concluding that the evaluation reports analysed herein are usable or not will be a misjudgement and subjective. The stance of this study remains that there is correlation between evaluation quality and utilisation of evaluation results. The burgeoning demand from policy makers and decision-makers for evaluation evidence to influence policy and decision-making in an ever-changing world, obliges evaluators to acknowledge the value of qualitative evaluations as they provide in-depth and contextual understanding.

From our perspective, it seems that the quality framework should be infused in peer-reviewing of the qualitative evaluation reports and not added onto them when and if deemed necessary. From the onset, the framework argues that for a meaningful report to be produced it has to fulfil all the eighteen appraisal questions. It therefore becomes necessary to understand how credibility of qualitative reports should be maintained to ensure that the audience is convinced of the knowledge produced to the extent that it can inform policy- and decision-making. The analysis of the study did just this by looking at how various components of the qualitative evaluation reports should be crafted.

We have found the framework to be necessary in providing qualitative scientific rigour. However, we raise some critical questions. These do not hinder the use of the framework, but call for significant reworking of the framework.

Despite the critical questions raised about the QF, the framework was of great theoretical value in terms of understanding and analysing the qualitative evaluation reports. We wish to stress that the notion of 'quality of qualitative evaluations' needs to be unpacked to identify differences in reporting of evaluation findings amongst donors/funders, an endeavour which regrettably was not addressed in this study. For that reason, the broad claims made in this study, though valid at the level of abstraction at which they are pitched, need to be qualified through acknowledgement of different reporting systems.

## References


AIR (American Institutes for Research). 2015. *Performance Evaluation of Reading for Ethiopia's Achievement Developed Institutional Improvement (READ II).* Washington, DC.

ALNAP (Active Learning Network for Accountability and Performance in Humanitarian Action). 2006. *Evaluating humanitarian action using the OECD-DAC criteria: An ALNAP guide for humanitarian agencies.* London: Overseas Development Institute.

Armiger, B. 1997. Ethics in nursing research: Profile, principles, perspective. *Nursing Research*, 26(5):330-333. https://doi.org/10.1097/00006199-197709000-00003

Bamberger, M. 2004. Influential evaluations: Evaluations that improved performance and impacts of development programs. Operations Evaluation Department, the World Bank. Washington, DC: The World Bank.

Bartlett, J.E.; Kotrlik, J.W. & Higgins, C.C. 2001. Organizational research: Determining appropriate sample size in survey research. *Information Technology, Learning, and Performance Journal*, 19(1):43.

Bober, C.E. & Bartlett, K.R. 2004. The utilization of training program evaluation in corporate universities. *Human Resource Development Quarterly*, 15(4):363-383. https://doi.org/10.1002/hrdq.1111

Bowbrick, P. 2015. The use of voice recorders in economic research. *Journal of Agricultural Economics,* 26(2):n.p. https://doi.org/10.1111/j.1477-9552.1975.tb01466.x

Burns, N. & Grove, S.K. 2003. *The practice of nursing research: Conduct, critique and utilization.* Toronto: W.B. Saunders.

Burry, J.; Alkin, M.C. & Ruskus, J. 1985. Organizing evaluations for use as a management tool. *Studies in Educational Evaluation*, 11(2):131-157. https://doi.org/10.1016/0191-491X(85)90020-3

Chianca, T.C. 2007. International development evaluation: An analysis and policy proposals. PhD thesis, Western Michigan University, Kalamazoo.

Chianca, T. 2008. The OECD/DAC criteria for international development evaluations: An assessment and ideas for improvement. *Journal of MultiDisciplinary Evaluation*, 5(9):41-51.

Chilisa, B. & Malunga, C. 2014. Made in Africa Evaluation: Uncovering African roots in evaluation theory and practice. In: F. Cloete, B. Rabie, & C. de Coning (eds). *Evaluation management in South Africa and Africa.* Stellenbosch: African Sun Media. https://doi.org/10.18820/9781920689513

Connolly, T. & Porter, A.L. 1980. A user-focused model for the utilization of evaluation. *Evaluation and Program Planning*, 3(2):131-140. https://doi.org/10.1016/0149-7189(80)90061-0

Dassah, M.O. & Uken, E.A. 2007. Towards more effective utilisation of evaluations. *Journal of Public Administration*, 42(2):119-135.

De Vos, A.A.; Strydom, H.; Fouché, C.B. & Delport, C.S.L. 2011. *Research at Grassroots for Social Sciences and Human Service Professions.* Fourth edition. Pretoria: Van Schaik.

DPME (Department of Planning, Monitoring and Evaluation). 2014. *Standards for evaluation in government.* Pretoria: Government Printing Works.

Farrugia, P.; Petrisor, B.A.; Farrokhyar, F. & Bhandari, M. 2010. Research questions, hypotheses and objectives. *Canadian Journal of Surgery*, 53(4):278-281.

Flick, U.; Von Kardorff, E. & Steinke, I. 2004. *A companion to qualitative research.* London: SAGE.

Gilgun, J.F. 2008. Lived experience, reflexivity, and research on perpetrators of interpersonal violence. *Qualitative Social Work*, 7(2):181-197. https://doi.org/10.1177/1473325008089629

Ginsberg, A. & Rhett, N. 2003. Building a better body of evidence: New opportunities to strengthen evaluation utilization. *American Journal of Evaluation*, 24(4):489-498. https://doi.org/10.1016/j.ameval.2003.10.005

Guba, E.G. & Lincoln, Y.S. 1989. *Fourth generation evaluation.* Newbury Park: SAGE.

Hinkin, T.R.; Tracey, J.B. & Enz, C.A. 1997. Scale construction: Developing reliable and valid measurement instruments. *Journal of Hospitality & Tourism Research*, 21(1):100-120. https://doi.org/10.1177/109634809702100108

Johnson, K.; Greenseid, L.O.; Toal, S.A.; King, J.A.; Lawrenz, F. & Volkov, B. 2009. Research on evaluation use: A review of the empirical literature from 1986 to 2005. *American Journal of Evaluation*, 30(3):377-410. https://doi.org/10.1177/1098214009341660

Johnston, W.P., Jr. 1986. A study of the acceptance of management performance evaluation recommendations by federal agencies: Lessons from GAO reports issued in FY 1983. Dissertation Abstracts International: Section A: The Humanities and Social Sciences, 48, 2157 (UMI 8725055).

Kibombo, R.; Owarwo, V.; Nuamanya, A.; Jessani, N.; Otikal, K. & Mubangizi, P. 2016. Evaluation: USAID/Uganda Strengthening Decentralization for Sustainability (SDS) Program. Prepared for United States Agency for International Development. http://ghpro.dexisonline.com/sites/default/files/136%20SDS_Eval_Final_Format%207.18.16_508.compressed.pdf

Kumar, R. 2005. *Research Methodology – A Step-by-Step Guide for Beginners.* Second edition. Singapore: Pearson Education.

Leedy, P.D. & Ormrod, J.E. 2005. *Practical Research: Planning and Design.* Seventh edition. New Jersey: Merrill Prentice Hall.

Leslie, M.; Moodley, N.; Goldman, I.; Jacob, C.; Podems, D.; Everett, M. & Beney, T. 2015. Developing evaluation standards and assessing evaluation quality. *African Evaluation Journal* 3(1), Art. #112, 13 pages. https://doi.org/10.4102/aej.v3i1.112

Lewandowsky, S.; Ecker, U.K.H. & Cook, J. 2017. Beyond misinformation: Understanding and coping with the"Post-Truth" era. *Journal of Applied Research in Memory and Cognition*, 6(4):353-369. https://doi.org/10.1016/j.jarmac.2017.07.008

MDPMED (Ministry of Devolution and Planning, Monitoring and Evaluation Department). 2016. *Guidelines for the development of county integrated monitoring and evaluation systems.* Nairobi: Government of the Republic of Kenya.

Meyrick, J. 2006. What is Good Qualitative Research? A First Step towards a Comprehensive Approach to Judging Rigour/Quality. *Journal of Health Psychology*, 11(5):799-808. https://doi.org/10.1177/1359105306066643

Morris, M. 2008. *Evaluation ethics for best practices.* London: Guilford Press.

Neuman, W.L. 1994. *Social research methods: Qualitative and quantitative approaches.* London: Allyn and Bacon.

O'Brien, B.C.; Harris, I.B.; Beckman, T.J.; Reed, D.A. & Cook, D.A. 2014. Standards for reporting qualitative Research: A Synthesis of Recommendations. *Academic Medicine*, 89(9):1245-1250. https://doi.org/10.1097/ACM.0000000000000388

OECD (Organisation for Economic Cooperation and Development). 2010. *DAC guidelines and reference series quality standards for development evaluation.* Paris: OECD Secretariat.

Ofir, Z. 2017. *Updating the DAC evaluation Criteria, Part 2. Why change the status quo?* http://zendaofir.com/updating-the-dac-evaluation-criteria-part-2/ [Accessed 12 January 2018].

Patel, M. 2002. A meta-evaluation, or quality assessment, of the evaluations in this issue, based on the African Evaluation Guidelines: 2002. *Evaluation and Program Planning*, 25(4):329-332. https://doi.org/10.1016/S0149-7189(02)00043-5

Patel, M. 2013. African evaluation guidelines. *African Evaluation Journal* 1(1), Art. #51. https://doi.org/10.4102/aej.v1i1.51

Patton, M.Q. 2015. *Qualitative Research and Evaluation Methods: Integrating Theory and Practice*. Fourth edition. Thousand Oaks, CA: SAGE.

Potts, S.A.K. 1998. Impact of mixed method designs on knowledge gain, credibility, and utility of program evaluation findings. Dissertation Abstracts International: Section A: The Humanities and Social Sciences, 59, 1942 (UMI 9837695).

Probst, B. 2015. The eye regards itself: Benefits and challenges of reflexivity in qualitative social work research. *Social Work Research*, 39(1):37-48. https://doi.org/10.1093/swr/svu028

Puhan, M.A.; Akl, E.A.; Bryant, D.; Xie F.; Apolone, G. & Ter Riet, G. 2012. Discussing study limitations in reports of biomedical studies – the need for more transparency. *Health and Quality of Life Outcomes,* 10:23. https://doi.org/10.1186/1477-7525-10-23

Rockwell, S.K.; Dickey, E.C. & Jasa, P.J. 1990. The personal factor in evaluation use: A case study of a steering committee's use of a conservation tillage survey. *Evaluation and Program Planning*, 13:389-394. https://doi.org/10.1016/0149-7189(90)90024-Q

Rog, D.J. & Bickman, L. 1984. The feedback research approach to evaluation: A method to increase utility. *Evaluation and Program Planning*, 7(2):169-175. https://doi.org/10.1016/0149-7189(84)90042-9

Roller, M.R. & Lavrakas, P.J. 2015. *Applied Qualitative Research Design: A Total Quality Framework Approach*. New York, NY: Guilford Press.

Sandelowski, M. 1995. Sample size in qualitative research. *Research in Nursing & Health*, 18(2):179-183. https://doi.org/10.1002/nur.4770180211

Shea, M.P. 1991. Program evaluation utilization in Canada and its relationship to evaluation process, evaluator and decision context variables. University of Windsor (Canada). Dissertation Abstracts International: Section B: The Sciences & Engineering, 53, 597.

Shenton, A.K. 2004. Strategies for ensuring trustworthiness in qualitative research projects. *Education for Information*, 22(2):63-75. https://doi.org/10.3233/EFI-2004-22201

Silverman, D. 2000. *Doing qualitative research: A practical handbook*. London: SAGE.

Sobal, J. 2001. Sample extensiveness in qualitative nutrition education research. *Journal of Nutrition Education*, 33(4): 184-192. https://doi.org/10.1016/S1499-4046(06)60030-4

Spencer, L.; Ritchie, J.; Lewis, J. & Dillon, L. 2003. *Quality in qualitative evaluation: A framework for assessing research evidence*. London: National Centre for Social Research, Government Chief Social Researcher's Office.

Stuckey, H.L. 2014. The first step in data analysis: Transcribing and managing qualitative research data. *Journal of Social Health and Diabetes*, 2(1):6-8. https://doi.org/10.4103/2321-0656.109828

Tong, A.; Sainsbury, P. & Craig, J. 2007. Consolidated criteria for reporting qualitative research (COREQ): A 32-item checklist for interviews and focus groups. *International Journal for Quality in Health Care*, 19(6):349-357. https://doi.org/10.1093/intqhc/mzm042

UNEG (United Nations Evaluation Group). 2016. *Norms and standards for evaluation*. New York: UNEG.

UNFPA (United Nations Population Fund). 2017. *Tools and guidance: Dimensions of evaluation quality at UNFPA*. New York: UNFPA Evaluation Office.

Veal, A.J. 2011. *Research methods for leisure and tourism*. Fourth edition. London: Pearson Education.

Walsh, D. & Downe, S. 2006. Appraising the quality of qualitative research. *Midwifery*, 22(2):108-119. https://doi.org/10.1016/j.midw.2005.05.004

Weiss, C.H. 1998. *Evaluation: Methods for Studying Programs and Policies*. Second edition. Upper Saddle River, NJ: Prentice Hall.

WFP (World Food Programme). 2015. *Evaluation Quality Assurance System (EQAS): Guidelines for operation evaluations*. Rome: WFP Office of Evaluation.

White, C.; Woodfield, K. & Ritchie, J. 2003. Reporting and presenting qualitative data. In: J. Ritchie & J. Lewis (eds). *Qualitative Research Practice*. London: SAGE. 287-320.

# ANNEXURE

| Quality Framework | | | |
|---|---|---|---|
| | A. Appraisal questions | B. Quality indicators (possible features for consideration) | C. Notes on study being appraised |
| 1. Findings | How credible are the findings? | a. Findings/conclusions are supported by data/study evidence *(i.e. the reader can see how the researcher arrived at his/her conclusions; the 'building blocks' of analysis and interpretation are evident)*.<br>b. Findings/conclusions 'make sense/have a coherent logic'.<br>c. Findings/conclusions are resonant with other knowledge and experience *(this might include peer or member review)*.<br>d. Use of corroborating evidence to support or refine findings *(i.e. other data sources have been used to examine phenomena; other research evidence has been evaluated)*. | |
| 2. Findings | How has knowledge/ understanding been extended by the research? | a. Literature review *(where appropriate)* summarising knowledge to date/key issues raised by previous research.<br>b. Aims and design of study set in the context of existing knowledge/ understanding; identifies new areas for investigation *(for example, in relation to policy/practice/substantive theory)*.<br>c. Credible/clear discussion of how findings have contributed to knowledge and understanding *(e.g. of the policy, programme or theory being reviewed)*; might be applied to new policy developments, practice or theory.<br>d. Findings presented or conceptualised in a way that offers new insights/ alternative ways of thinking.<br>e. Discussion of limitations of evidence and what remains unknown/unclear or what further information/research is needed. | |

| Quality Framework | | | |
|---|---|---|---|
| | A. Appraisal questions | B. Quality indicators (possible features for consideration) | C. Notes on study being appraised |
| 3. Findings | How well does the evaluation address its original aims and purpose? | a. Clear statement of study aims and objectives; reasons for any changes in objectives.<br>b. Findings clearly linked to the purposes of the study – and to the initiative or policy being studied.<br>c. Summary or conclusions directed towards aims of study.<br>d. Discussion of limitations of study in meeting aims *(e.g. are there limitations because of restricted access to study settings or participants, gaps in the sample coverage, missed or unresolved areas of questioning; incomplete analysis; time constraints?).* | |
| 4. Findings | Scope for drawing wider inference – how well is this explained? | a. Discussion of what can be generalised to wider population from which sample is drawn/case selection has been made.<br>b. Detailed description of the contexts in which the study was conducted to allow applicability to other settings/contextual generalities to be assessed.<br>c. Discussion of how hypotheses/propositions/findings may relate to wider theory; consideration of rival explanations.<br>d. Evidence supplied to support claims for wider inference *(either from study or from corroborating sources).*<br>e. Discussion of limitations on drawing wider inference *(e.g. re-examination of sample and any missing constituencies: analysis of restrictions of study settings for drawing wider inference).* | |
| 5. Findings | How clear is the basis of evaluative appraisal? | a. Discussion of how assessments of effectiveness/evaluative judgements have been reached *(i.e. whose judgements are they and on what basis have they been reached?).*<br>b. Description of any formalised appraisal criteria used, when generated and how and by whom they have been applied.<br>c. Discussion of the nature and source of any divergence in evaluative appraisals<br>d. Discussion of any unintended consequences of intervention, their impact and why they arose. | |

| Quality Framework | | | |
|---|---|---|---|
| | A. Appraisal questions | B. Quality indicators (possible features for consideration) | C. Notes on study being appraised |
| 6. Design | How defensible is the research design? | a. Discussion of how overall research strategy was designed to meet aims of study<br>b. Discussion of rationale for study design.<br>c. Convincing argument for different features of research design *(e.g. reason given for different components or stages of research; purpose of particular methods or data sources, multiple methods, time frames).*<br>d. Use of different features of design/data sources evident in findings presented.<br>e. Discussion of limitations for study evidence. | |
| 7. Sample | How well defended is the sample design/ target selection of cases/ documents? | a. Discussion of how overall research strategy was designed to meet aims of study.<br>b. Discussion of rationale for study design.<br>c. Convincing argument for different features of research design *(e.g. reasons given for different components or stages of research; purpose of particular methods or data sources, multiple methods, time frames etc.).*<br>d. Use of different features of design/data sources evident in findings presented.<br>e. Discussion of limitations of research design and their implications for the study evidence. | |
| 8. Sample | Sample composition/ case inclusion – how well is the eventual coverage described? | a. Detailed profile of achieved sample/case coverage.<br>b. Maximising inclusion *(e.g. language matching or translation; specialised recruitment; organised transport for group attendance).*<br>c. Discussion of any missing coverage in achieved samples/cases and implications for study evidence *(e.g. through comparison of target and achieved samples, comparison with population etc.).*<br>d. Documentation of reasons for non-participation among sample approached/ non-inclusion of selected cases/documents.<br>e. Discussion of access and methods of approach and how these might have affected participation/coverage. | |

| Quality Framework | | | |
|---|---|---|---|
| | A. Appraisal questions | B. Quality indicators (possible features for consideration) | C. Notes on study being appraised |
| 9. Data Collection | How well was the data collection carried out? | a. Discussion of:<br>• who conducted data collection;<br>• procedures/documents used for collection/recording;<br>• checks on origin/status/authorship of documents.<br>b. Audio or video recording of interviews/discussions/conversations *(if not recorded, were justifiable reasons given?)*.<br>c. Description of conventions for taking field notes *(e.g. to identify what form of observations were required/to distinguish description from researcher commentary/analysis)*.<br>d. Discussion of how fieldwork methods or settings may have influenced data collected.<br>e. Demonstration, through portrayal and use of data, that depth, detail and richness were achieved in collection. | |
| 10. Analysis | How well has the approach to, and formulation of, the analysis been conveyed? | a. Description of form of original data *(e.g. use of verbatim transcripts, observation or interview notes, documents, etc.)*.<br>b. Clear rationale for choice of data management method/tool/package.<br>c. Evidence of how descriptive analytic categories, classes, labels etc. have been generated and used *(i.e. either through explicit discussion or portrayal in the commentary)*.<br>d. Discussion, with examples, of how any constructed analytic concepts/ typologies etc. have been devised and applied. | |
| 11. Analysis | Contexts of data sources – how well are they retained and portrayed? | a. Description of background or historical developments and social/ organisational characteristics of study sites or settings.<br>b. Participants' perspectives/observations placed in personal context *(e.g. use of case studies/vignettes/individual profiles, textual extracts annotated with details of contributors)*.<br>c. Explanation of origins/history of written documents.<br>d. Use of data management methods that preserve context *(i.e. facilitate within case description and analysis)*. | |

| Quality Framework | | | |
|---|---|---|---|
| | A. Appraisal questions | B. Quality indicators (possible features for consideration) | C. Notes on study being appraised |
| 12. Analysis | How well has diversity of perspective and content been explored? | a. Discussion of contribution of sample design/ case selection in generating diversity.<br>b. Description and illumination of diversity/multiple perspectives/alternative positions in the evidence displayed.<br>c. Evidence of attention to negative cases, outliers or exceptions.<br>d. Typologies/models of variation derived and discussed.<br>e. Examination of origins/influences on opposing or differing positions.<br>f. Identification of patterns of association/linkages with divergent positions/ groups. | |
| 13. Analysis | How well has detail, depth and complexity *(i.e. richness)* of the data been conveyed? | a. Use and exploration of contributors' terms, concepts and meanings.<br>b. Unpacking and portrayal of nuance/subtlety/intricacy within data.<br>c. Discussion of explicit and implicit explanations.<br>d. Detection of underlying factors/influences.<br>e. Identification and discussion of patterns of association/conceptual linkages within data.<br>f. Presentation of illuminating textual extracts/observations. | |
| 14. Reporting | How clear are the links between data, interpretation and conclusions – i.e. how well can the route to any conclusions be seen? | a. Clear conceptual links between analytic commentary and presentations of original data *(i.e. commentary and cited data relate; there is an analytic context to cited data, not simply repeated description).*<br>b. Discussion of how/why particular interpretation/significance is assigned to specific aspects of data – with illustrative extracts of original data.<br>c. Discussion of how explanations/ theories/conclusions were derived – and how they relate to interpretations and content of original data *(i.e. how warranted)*; whether alternative explanations explored.<br>d. Display of negative cases and how they lie outside main proposition/theory/ hypothesis etc.; or how proposition etc. revised to include them. | |

| Quality Framework | | | |
|---|---|---|---|
| | A. Appraisal questions | B. Quality indicators (possible features for consideration) | C. Notes on study being appraised |
| 15. Reporting | How clear and coherent is the reporting? | a. Demonstrates link to aims of study/research questions.<br>b. Provides a narrative/story or clearly constructed thematic account.<br>c. Has structure and signposting that usefully guide reader through the commentary.<br>d. Provides accessible information for intended target audience(s).<br>e. Key messages highlighted or summarised. | |
| 16. Neutrality & Reflexivity | How clear are the assumptions/ theoretical perspectives/ values that have shaped the form and output of the evaluation? | a. Discussion/evidence of the main assumptions/hypotheses/theoretical ideas on which the evaluation was based and how these affected the form, coverage or output of the evaluation *(the assumption here is that no research is undertaken without some underlying assumptions or theoretical ideas)*.<br>b. Discussion/evidence of the ideological perspectives/values/philosophies of research team and their impact on the methodological or substantive content of the evaluation *(again, may not be explicitly stated)*.<br>c. Evidence of openness to new/alternative ways of viewing subject/theories/ assumptions *(e.g. discussion of learning/concepts/ constructions that have emerged from the data; refinement restatement of hypotheses/theories in light of emergent findings; evidence that alternative claims have been examined)*.<br>d. Discussion of how error or bias may have arisen in design/data collection/ analysis and how addressed, if at all.<br>e. Reflections on the impact of the researcher on the research process. | |

| Quality Framework | | | |
|---|---|---|---|
| | A. Appraisal questions | B. Quality indicators (possible features for consideration) | C. Notes on study being appraised |
| 17. Ethics | What evidence is there of attention to ethical issues? | a. Evidence of thoughtfulness/sensitivity about research contexts and participants.<br>b. Documentation of how research was presented in study settings/to participants *(including, where relevant, any possible consequences of taking part).*<br>c. Documentation of consent procedures and information provided to participants.<br>d. Discussion of confidentiality of data and procedures for protecting.<br>e. Discussion of how anonymity of participants/sources was protected.<br>f. Discussion of any measures to offer information/advice/services etc. at end of study *(i.e. where participation exposed the need for these).*<br>g. Discussion of potential harm or difficulty through participation, and how avoided. | |
| 18. Auditability | How adequately has the research process been documented? | a. Discussion of strengths and weaknesses of data sources and methods.<br>b. Documentation of changes made to design and reasons; implications for study coverage.<br>c. c. Documentation and reasons for changes in sample coverage/data collection/ analytic approach; implications.<br>d. Reproduction of main study documents *(e.g. letters of approach, topic guides, observation templates, data management frameworks etc.).* | |

(*Source:* Spencer et al., 2003:9-15)

7

# INDIGENISING EVALUATION KNOWLEDGE: EXPLORING THE EPISTEMIC IDENTITY OF AFRICAN EVALUATORS

Precious Tirivanhu & Caitlin Blaser Mapitsa

## Introduction

National Monitoring and Evaluation Systems (NMESs) continue to emerge in Africa due to the growing need for governments to strengthen accountability, and a call in the Sustainable Development Goals for country-led evaluation systems (Rosenstein, 2013). These systems are built to generate knowledge on the implementation and results of public sector strategies, policies and programmes in relation to national goals, such as economic growth or poverty reduction (Engela & Ajam, 2010). In essence, these systems provide norms for the judgement of outcomes and impacts of government programming. However, there is growing concern that the donor-driven roots of the M&E sector are creating systems that are far from transformative, but rather more deeply entrenching neo-colonial development epistemologies and pathways.

This chapter begins by exploring the emergence of NMES within the trajectory of the knowledge indigenisation discourse in Africa. It then looks at recent data from a survey on evaluation professionals in the region, to try and bring some empirical grounding to questions of indigeneity of the sector. Finally, it considers the relationship between the epistemological identity of evaluation practitioners (professional identity, country of residence, evaluation field, specialised training and years in the evaluation sector) and their relationship to evaluation approaches and methodological choices. It uses these to make a case for stronger engagement between the fields of monitoring and evaluation and indigenous knowledge systems to ensure contextually appropriate development pathways are being charted.

The establishment of the Department of Planning, Monitoring and Evaluation (DPME) in South Africa in 2009, for example, pioneered monitoring and evaluation within the public sector, with the governments of Benin, Uganda, Ghana and Kenya, amongst others, rapidly taking steps to develop national evaluation policies and systems. The mandate of DPME is to facilitate inter alia the development of plans or delivery agreements for the strategic cross-cutting priorities or outcomes of government; monitor the performance of individual national and provincial government departments and municipalities; and promote good planning and M&E practices in government (DPME, 2014). Over the last decade, Uganda, Benin, Ghana, Côte d'Ivoire and Niger, amongst others, have made tremendous strides in entrenching NMESs. This development has happened in parallel with increasing professionalisation of the evaluation sector in the region. The African Evaluation Association (AfrEA) has seen its

membership of voluntary organisations for professional evaluators (VOPEs) grow and, in 2017, the Department of Higher Education and Training (DHET) accredited the *African Evaluation Journal* for the first time, demonstrating that the sector has entrenched itself in higher education institutions as well.

Both the supply and demand sides of the evaluation profession are struggling for capacity to grow at a rate commensurate with the expansion of the systems and gaps are now being acknowledged (Abrahams, 2015). This chapter focuses on the supply side. Within this context, a number of challenges have emerged, including the need for contextually relevant and utilisable evaluation results, and the need to increase the pool of competent African evaluators to ease the reliance on evaluators from the North. The knowledge-generation agenda through NMESs is happening within a trajectory of a growing discourse on the decolonisation of knowledge and indigenisation of evaluation in Africa (Ndlovu-Gatsheni, 2018; Chilisa, Major, Gaotlhobogwe & Mokgolodi, 2016). Additionally, there is a growing body of knowledge on addressing the Western-centric character of management knowledge within public sector institutions (Girei, 2017). This is in light of the debates around the highly Eurocentric nature of the evaluation enterprise where approaches and methodologies are biased towards Western epistemologies.

According to Chilisa et al. (2016:314), it is fundamental for evaluators and researchers to debate the assumptions about the nature of reality (ontology), knowledge (epistemology) and values (axiology) that inform Afrocentric methodologies and worldviews in evaluation enquiry and practice. This is viewed as a prerequisite for indigenising evaluation knowledge. The decolonisation of knowledge or epistemic freedom discourse recognises the need for African-rooted epistemologies in addressing the contextual nature of social development. Epistemologies and knowledge systems at most African institutions, including universities and the public sector, are argued not to have considerably changed. They remain rooted in colonial, apartheid worldviews and epistemological traditions (Heleta, 2016). Indigenous knowledge, the locally-based knowledge created before colonisation, continues to exist in most parts of the world and within the global knowledge economy in an increasingly marginalised position (Connel, 2016). Thus, there is need for the deconstruction of existing methodologies and methods that (re)produce the coloniality of knowledge towards a reconstruction/or reinvention of research and evaluation practice.

As NMESs grow in Africa, there is need for a better understanding of the epistemological grounding of the M&E profession (Cloete, 2016). There is a growing recognition that evaluation processes and products may be customised to reflect culturally-based programme strength and effectiveness, and to discover knowledge relevant to institutional transformation that enhances public sector performance in Africa (Morelli & Mataira, 2010). Epistemology enhances understanding of the nature of knowledge from evaluation, its possibility, scope and general basis (Hamlyn, 1995:242). This ensures the adequacy, relevance, validity, and legitimacy of evaluation knowledge.

Evaluation professionals are grappling with the development of African-rooted evaluation approaches. This discourse is being driven by the African Evaluation Association (AfrEA) within the theme of Made in Africa Evaluation, calling attention to the need for more discussion around African evaluation approaches, tools and methods. Such approaches are envisaged to incorporate cultural knowledge and values that guide understandings of and responses to developmental needs, priorities and African aspirations. The seminal works for this approach have been informed by African relational paradigms[1] as a basis for culturally relevant evaluations based on African worldviews (Chilisa, Major, Gaotlhobogwe & Mokgolodi, 2016; Chilisa & Malunga, 2012). However, this discussion has lacked both a broader theoretical linkage to the construction of knowledge, and a specific empirical foundation of the professional evaluation sector. With regards to improving the pool of African experts, there are growing efforts on evaluation capacity building (ECB) in order to improve competencies of African evaluators. This professionalisation agenda aims at enhancing the required competencies, standards and guidelines that inform professional evaluation practice. However, there is not yet consensus across the sector about what these competencies look like, nor what the most appropriate mechanism would be for strengthening them.

Within this context, addressing these evaluation supply side constraints invariably encompasses unpacking and understanding issues of epistemological identity and linkages with evaluation approaches, methods, paradigms and beliefs in knowledge. The notion of Made in Africa has obvious epistemological connotations since it aims at inculcating African-rooted epistemologies to evaluators. On the other hand, development of contextually relevant ECB

[1] These philosophies, world views and paradigms share relational ways of perceiving reality and of being, ways of knowing and value systems (Chilisa et al., 2016:317).

systems requires an appreciation of the epistemological identity of African evaluators, since the process requires shifts in relationships, power and identity (Walters, 2007). As outline in the preceding sections, it is evident that there is widespread assertion that African evaluators are utilising Western approaches and methodologies. In addition, the role of identity in the broader knowledge decolonisation discourse is gaining momentum. In fact, decolonising knowledge should be African-people centred (Chilisa et al., 2016:316; Chilisa, 2012).

Within the broader framework of the Made in Africa Evaluation approach there is need for in-depth understanding on methodological choices by African evaluators. However, issues of identity in relation to methodological choices and approaches still need empirical research. Methodological choices reflect different epistemologies, ontologies and axiologies. Questions around factors that significantly affect methodological choices by African evaluators need to be explored to appreciate the epistemic constructs of African evaluators. This chapter asserts that unpacking and understanding of existing epistemological, ontological an axiological beliefs of African evaluators are important steps in indigenising knowledge. This chapter, therefore, addresses three research questions:

1. What does empirical evidence say about the frequent assertion that most evaluators undertaking evaluations in Africa are from the West?
2. What are the perceptions on the Made in Africa Evaluation approaches?
3. What is the association between various attributes of identity and choice of methodological approaches by African evaluators?

This chapter draws from a collaborative survey conducted in 2016 by the Centre for Learning on Evaluation and Results for Anglophone Africa (CLEAR-AA) and the Centre for Research on Evaluation, Science and Technology (CREST). The evaluation profession is saddled with a history stemming from the international donor space and remains strongly linked to the neo-colonial development paradigms. While some have indicated that a lack of transformation in the sector remains a limitation of the field, these assertions have had a weak empirical basis (Cloete, 2016). By drawing on new empirical data, this chapter explores how the politics of identity are located in the evaluation profession.

## An overview of foundational issues

### *The growth of public sector M&E systems in Africa*

There is no doubt that public sector M&E systems in Africa have seen unprecedented growth over the last decade (Cloete, 2014). However, this growth has been characterised by three main issues. The first is that it has been uneven, with governments who are prepared to confront failures of policy or implementation at the forefront, and other countries conspicuously silent. The second is that this growth in systems and mandate has not always been matched by an equivalent growth in capacity to implement, so there has been unevenness in implementation. Finally, there are limitations in expanding both the supply of professional evaluators and the public sector demand for evaluation due to a mismatch between supply and demand in the region (Goldman, Mathe, Jacob, Hercules, Amisi, Buthelezi, Narsee, Ntakumba & Sadan, 2015). This mismatch is still being understood and demands stronger empirical foundations to be interrogated effectively.

On the demand side, there is a large global literature on evaluation use that focuses on understanding the role of the commissioning organisation, from management to changing programme design (Bourgeois & Nare, 2015; Patton, 2008; Sanderson, 2002). While an explicit and deliberate focus on use by programme managers is one of the features which sets evaluation apart from research, it also prioritises the viewpoints of programme managers.

Development evaluation and other approaches offer a counterpoint which mandate an emphasis on the "public good", or the beneficiary result from the evaluation process. However, neither approach, which aligns the evaluation process to a stakeholder interest, sufficiently acknowledges the diversity in perspectives that different role players bring. One is that they are unlikely to be homogenous or static. Another is that the relational aspects are absent. A third, is that they fail to link these group interests to epistemological approaches or worldviews they may reflect. This makes it difficult to interrogate. Discourse from indigenous knowledge systems may be able to contribute to this debate.

One of the primary distinguishing points between evaluation and research is that evaluation fundamentally entails a value judgement, by the determination of a programme's alignment or appropriateness to a particular values system as fundamental to the field (Babbie & Mouton, 2010). The evaluation literature often comes with an assumption that programmes have a set of stakeholders,

that range from donors to programme managers, agencies and beneficiaries, and that these stakeholders follow a similar ordered epistemological orientation. With the growth of African public sector M&E systems, it is clear that beneficiaries, implementing agencies and management are all dealing with common issues of neo-colonialism. This linear spectrum does not reflect the experiences and reality of class and cosmopolitanism in the region. The challenge thus lays in the development of a shared epistemological identity that enhances shared beliefs in evaluation knowledge.

## *The relationship between epistemology and identity*

Epistemology (the theory of knowledge and justification) provides the philosophical grounding for determining the nature, adequacy and legitimacy of knowledge (BonJour, 2010; Pritchard, 2010; Audi, 2003; Crotty, 1998). It defines the knowledge claims or beliefs (how we know what we know), the basic concepts applied in expression of such beliefs and status to be ascribed to the understandings of research results (Dobson, 2002). Nobles (2006) describes epistemology as a science involving the study of: the nature of reality; how truth is defined; the relationship between the knower, knowing, and the known; what can be known; and what should/could be done with the known. In scientific research, the relationship between the knowing scholar and his or her subject is moderated by epistemology (McDougal, 2014). This study adopts a typology by Audi (2003) that defines knowledge as comprising beliefs that are accepted to be true, and are conclusively justified at any given time, i.e. conclusively justified true beliefs.

Identity can geneally be defined as whom or what someone is (the view of self), the various meanings people can attach to themselves, or the meanings attributed by others (Beijaard, 1995). Identities are the shared social meanings that persons attribute to themselves (Burke & Reitzes, 1991:242). According to Stets and Burke (2000), a key characteristic of the self is that it is reflexive, can take itself as an object and can categorise, classify, or name itself in particular ways in relation to other social categories or classifications. This self-categorisation forms identity. According to Hogg and Abrahams (1998), a social identity is a person's knowledge that they belong to a social category or group, i.e. a set of individuals with a common social identification or same social category. The main processes involved in social identity formation are self-categorisation, social comparison, intergroup relations and self-enhancement motivation (Stets & Burke, 2000:225).

The nexus between epistemology and identity is at the core of this chapter. In exploring this nexus, the chapter borrows theoretical constructs of epistemological identity theory (EIT). According to Demerath (2006), EIT explains how individuals enhance their knowledge of self and the world by creating and maintaining identities. Within the context of this chapter, therefore, commitment by evaluators to various identity variables is assumed to be fundamental in creating knowledge rooted in African epistemic foundations. The African-rooted epistemology should be grounded in the history, philosophy, and culture of classical Africa while also embracing the new challenges and transformations that characterise Africa (Hall, 2010). Identities are always in transition as a result of continuous learning, and EIT assumes that this transitional process is for epistemological reasons. EIT explains the epistemological value of identity as a means of organising knowledge and experience (Demerath, 2006:498). It is assumed that an identity creates patterns and perceptions that create beliefs or 'order' by evaluators and motivation to maintain those beliefs.

This chapter assumes variables for epistemological identity to include professional identity, country of birth and residence, years of experience in evaluation, levels of education and specialised training in evaluation. These play a pivotal role in shaping evaluation epistemologies. Methodological and evaluation practice-based decision-making and particular knowledge-based practices are deeply intertwined with an individual's meaning-making process and internal value system. Thus, this chapter explores evaluation epistemology issues as a function of the individual evaluator's perceptions of himself/herself and of the evaluation profession.

## *Conceptual framing – exploring epistemological identity*

The theoretical debates outlined above inform the conceptual framework for this chapter. As described in Figure 7.1, this chapter assumes that knowledge from NMESs is directly linked to the epistemology of the public sector and other players within the NMES. On the other hand, the epistemology of the public sector is affected by the identity of evaluation managers and practitioners. We assume that epistemological identity is fundamental in generating knowledge and results from NMESs that are relevant to the African context, enhancing the utilisation of such knowledge by policy makers. The study explores the identity of evaluation managers and practitioners using data from the African Evaluation Database (AfrED). Although the data may not be exhaustive it will contribute to

painting a picture for a much broader discourse. This chapter explores identity related variables and assesses impacts of those variables on epistemic identity and perceived impacts on knowledge generated from the NMES. This will not only test epistemological identity theory in one particular case, but also help us begin to differentiate between the individual and organisational drivers of epistemology. As we aim to influence NMES practice, understanding this difference will provide crucial insight into what interventions are likely to be the most effective.

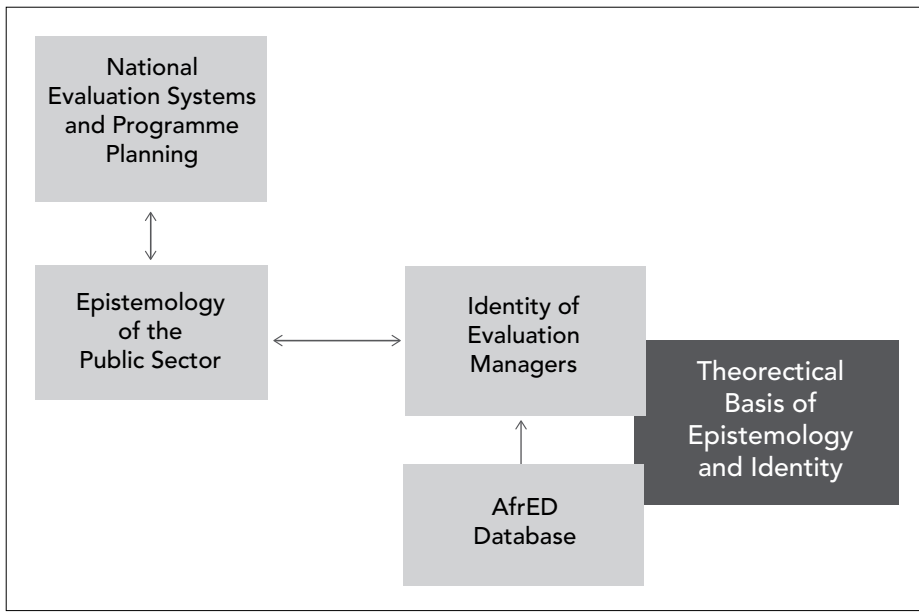



**Figure 7.1** Mapping of conceptual framework

## Methodology

This chapter draws the study population from the African Evaluation Database (AfrED) developed by CLEAR-AA and CREST in 2017. The database builds a reference point for evaluation practice and scholarship in Africa. The first iteration of this database in January 2017 included over 9000 academic journal articles and reports on M&E in Africa in 20 African countries: South Africa, Zimbabwe, Ghana, Ethiopia, Nigeria, Kenya, Tanzania, Uganda, Zambia, Lesotho, Namibia, Swaziland, Botswana, Malawi, Democratic Republic of Congo, Rwanda, Sudan, Togo, Liberia and Mozambique. The names and available contact information of all evaluators and researchers were drawn from the database.

During the period 10 to 29 November 2016, e-mails were sent to 3032 persons identified as individuals involved with evaluations in Africa. Eligible candidates for the survey were obtained from three sources:

- The documents listed in the African evaluation database were examined and all e-mail details of authors recorded.
- E-mail details were obtained from the South African, Ethiopian, Ghanaian African Evaluation Associations.
- E-mails were obtained from CREST's own internal database (previous students of evaluation and a list of various individuals making enquiries).

Based on these sources, 3032 individuals were sent an invitation to participate in the survey. Of these, 426 e-mails returned undelivered and nine individuals indicated that they did not want to participate in the survey. It is, therefore, assumed that around 2597 individuals received the invitation. Each e-mail contained a description of the survey, some information regarding CLEAR and CREST and a request to participate in the survey. In total, 549 individuals indicated their willingness to complete the survey. These 549 individuals were then sent a subsequent e-mail with a link to participate in the survey. A reminder was sent after one week to complete the survey and 421 completed the survey.

On 11 November, the remaining 2051 who had not indicated their willingness to participate were e-mailed a link to the survey, with 139 completing the survey. There were also four individuals who completed the survey through a general link provided. By 4 January (the date on which the survey was closed), a total of 564 individuals had completed the survey (22% response rate). The sample profile indicated that 37% were from South Africa, 55% were male, the majority was between the ages of 30 and 45 years and 75% were in possession of a master's or doctoral degree.

## Results

This chapter initiates a conversation on the epistemological orientation of African evaluators. The hope is that, through continued action research in the evaluation space, these discussions will be a starting point for using M&E as a tool for epistemological changes in the public sector. Emergent results so far both reinforce assertions that governments are largely blind to their epistemological orientations, but also call into question the assertions that the evaluation community is the primary constraint in the growth of appropriate

national evaluation systems. It also reinforces the diversity of both individual beliefs and institutional practice across the region, highlighting the challenges with assuming common regional trends. It has found certain correlations between evaluator identity and practice, strengthening the argument that this area of enquiry remains important in understanding the ongoing formation of national evaluation systems.

## Embedding African evaluation systems in public management

The growth of M&E in the region is an acknowledgement of the potential of M&E to improve public sector performance. However, it largely does not engage with the possibilities that targets could be set and progress defined differently. Engaging with thinkers on indigenous knowledge systems offers the possibility to provide an alternative method of interrogating public sector priorities and processes. While these views and perspectives should be integrated into planning processes, if evaluation processes and methods draw on indigenous knowledge, these recommendations will increasingly shape target-setting and programme design. This has been acknowledged by the professional evaluator community, through the Made in Africa discussion that has emerged from the African Evaluation Association. However, there has not been an immediate uptake by public sector evaluation systems, in part due to the difficulty bureaucracies face in adopting new ways of working. Results from the survey indicate that 64% of the respondents do not believe that there is a solid Made in Africa body of research. Although considerable efforts are being made through various fora, such as the AfrED database and other voluntary organisations of professional evaluators (VOPEs), there is need for more efforts to move from philosophical debates around what constitutes Made in Africa towards developing and empirically testing African-rooted evaluation approaches and methodologies that take cognisance of African context, culture, history and beliefs.

## Professional identity of evaluators

Within the evaluation community, it is virtually taken as a truism that M&E is inextricable from the donor community. As a result, its processes of conceptualisation, priority setting and management are Northern-driven, and both a cause and consequence of this is that evaluation practitioners are largely

not African. Cloete (2016) emphasises this assertion, saying: "Evaluations in Africa are still largely commissioned by non-African stakeholders who mostly comprise international donor or development agencies that run or fund development programmes on the continent."

While this perception is widely felt by a number of stakeholders and supported by ample anecdotal evidence, it does not seem particularly well supported by the AfrED survey results. Of a population of 3032 evaluators, researchers and evaluation managers, whose contact details were found in the AfrED database, 564 completed the survey. Of those who responded, 64 were born in a country outside of Africa, composed of 17 North Americans, three Australians, 10 respondents from the UK and 26 respondents from the rest of Europe, seven respondents from Asia, and one from South America. Fewer respondents were resident outside of African countries. While this percentage is slightly over 10%, and this is not an inconsequential percentage, it is lower, for example, than the percentage of non-African academic staff at the University of Witwatersrand in South Africa (University of the Witwatersrand, 2015). Due to the cosmopolitan nature of Africa's educated elite and the globalisation of highly skilled professionals, this percentage does not seem to immediately flag a disproportionate problem of the composition of the sector.

It is also important to point out that evaluation was only the primary professional identity of 43% of the survey respondents. While this is not surprising, given the nascent state of the evaluation profession and the tendency of evaluation professionals to have disciplinary backgrounds in other fields, it is still important for an analysis of individual identity. There is a possibility that either identity, or training as an evaluator, comes with a willingness to influence the alignment or orientation of evaluation systems in a way that is different from the way someone identifying as a civil servant, or technical specialist might engage. This could be a matter for further investigation.

Furthermore, of those whose primary professional identity in the M&E space was managing and commissioning evaluations in the public sector, all respondents were of African origin. Furthermore, of all respondents to the survey only 7% hailed from international development agencies. It is possible that further interrogation of the evaluations on the AfrED database will yield that these agencies are commissioning a disproportionate percentage of the evaluations, but this is not immediately apparent. Even if it were to be the case, this would

still have limited bearing on issues specific to planning and target setting within the public sector.

An additional driver of identity in the evaluation space that needs further exploration is that of gender. While evaluation is often a feminised profession, with roots in anthropological research or social work, this is not necessarily currently the case in Africa. Of the AfrED survey respondents, 55% are male, but the majority of respondents are based in South Africa. Of the South African respondents, 63% are female, indicating that the gender balance in other countries is skewed towards male evaluators. This is important both to contextualise evaluation as a profession in Africa and also to understand national specificities and drivers of identity-related dynamics.

## The association between professional identity, approaches and methods

it is often asserted in discussions of the African evaluation landscape that there is a relationship between identity and evaluation approaches. This section explores the extent to which this is true by considering the professional identity of evaluators and its link to evaluation approaches and methods. The associations between professional identity and evaluation approaches as well as professional identity and evaluation methods were explored using the $X^2$ Test for independence at the 95% confidence interval. The results indicate significant associations between professional identity (identity-related variable) as the explanatory variable and methods and approaches as response variables (epistemology-related variables), making a case for further exploration of the implications professional identity of evaluation practitioners has on the evaluation landscape. The variables are further disaggregated to show the various sub-categories of professional identity and the significantly associated sub-categories of evaluation approaches and methods ($P>0.05$). Results on evaluation approaches indicated that individuals who practised evaluation in any capacity significantly utilised participatory evaluation approaches. Researchers significantly utilised naturalistic evaluations and randomised control trials (RCTs) while managers utilised empowerment and participatory evaluations.

The results on use of evaluation methods show that evaluators at any capacity significantly preferred individual face-to-face interviews, secondary data analysis and stakeholder meetings/consultations. Researcher preferred individual

face-to-face interviews, secondary data analysis, stakeholder consultation and surveys. Those involved in managing evaluations significantly utilised focus group discussions (FGDs), interviews, observations, secondary data analysis, stakeholder consultations, surveys and testing. Trainers were biased towards the same research methods.[2] Those who commission evaluations preferred FDGs, interviews, observations, secondary data analysis and stakeholder consultations. University faculty members and instructors significantly utilised FGDs, interviews, observations, secondary data analysis, stakeholder consultations and surveys. Students preferred interviews, observations, secondary analysis and stakeholder consultations. Finally, retired experts significantly preferred interviews, secondary data analysis and stakeholder consultations.

These results highlight the importance of professional identity in the selection and utilisation of evaluation methods and approaches. There is a general bias towards the use of qualitative evaluation approaches and methods by African evaluators. Although similar studies are scarce in the evaluation discourse, research from other field such as teaching (Hong, 2010; Alvos & Aylwin, 2007) have established that 'self' is crucial in constructing individual decision- making and the mode of interaction with the environment in a given context. A striking observation is the limited utilisation of RCTs and naturalistic evaluations which are only significantly utilised by researchers. They are normally viewed as highly technical and complicated and associated with the medical profession. According to Togerson (2002), they are the ultimate expression of evaluative research. Such assertions likely lead to biases towards the utilisation of RCTs by certain professional identities, for example, Rowe (2016:8) argues that RCTs is a flawed approach in education since it portrays a wrong understanding of the nature of learning through incommensurability of beliefs around reality and knowledge.

Results from the $X^2$ Test of independence showed no significant association between the use of RCTs and level of education and specialised training in evaluation. This is contradictory to commonly held perceptions linking the use of RCTs to specialised training in quantitative evaluation methodologies. Such a scenario needs cognisance when promoting quantitative based Made in Africa approaches as it should be noted that education and specialised training

[2] Test methodologies include functional and non-functional testing to validate the AUT. Examples of testing methodologies are unit testing, integration testing, system testing, performance testing.

do not necessarily affect the choice in use of these approaches perceived as a domain for quantitative oriented evaluators. However there were significant associations ($p<0.05$) between the use of RCTs and areas of focus for evaluation related work. These areas are economic development ($X^2 2=6.136$; $p=0.047$), education ($X^2 2=7.249$; $p=0.027$), environment ($X^2 2=14.148$; $p=0.001$), health ($X^2 2=8.588$; $p=0.014$), sanitation ($X^2 2=11.155$; $p=0.003$) and youth development ($X^2 2=9.542$; $p=0.008$). This indicates an expansion of the use of RCTs into other areas outside the health sector.

There were significant associations between years of experience in evaluation with a number of evaluations methods, including FGDs ($X^2 58=213.149$; $p=0.000$), secondary data analysis ($X^2 58=99.958$; $p=0.001$) and surveys ($X^2 54=88.320$; $p=0.002$). Significant associations were observed between years of experience in evaluation and the use of utilisation focused evaluation ($X^2 54=77.343$ $p=0.020$). Significant associations were also observed between the highest level of education and use of evaluability assessments ($X^2 10=47.396$; $p=0.000$) and Programme monitoring ($X^2 10=21.989$; $p=0.015$). Specialised training in evaluation was significantly associated with the type of evaluations conducted: clarificatory evaluation ($X^2 4=13.490$; $p=0.009$); economic valuation ($X^2 4=18.432$; $p=0.001$); impact evaluation $X^2 4=13.353$; $p=0.010$); process evaluation ($X^2 4=25.468$; $p=0.000$) and programme monitoring ($X^2 4=11.504$; $p=0.021$).

## Country of birth, country of residence, methods and approaches

Another discourse found in discussions of the African evaluation landscape, is that evaluation practitioners often come from abroad and use methods and approaches that may not be contextually relevant. While this is obviously a complex and nuanced discussion, the contribution of this chapter is simply to see if there is an empirical basis for a correlation between country of birth, country of residence, and evaluation methods and approaches.

Results indicated no significant association between the country of residence and the use of various methods. However, a significant association was observed between the country of birth and the use of the testing method ($X^2 90=113.506$; $p=0.048$). Considering evaluation approaches, significant association was observed between country of birth and use of RCTs ($X^2 82=133.497$; $p=0.000$) and realist evaluation ($X^2 82=111.286$; $p=0.017$). Significant associations were

also observed between country of residence and the use of RCTs ($X^2$76=128.729; p=0.000); realistic evaluation ($X^2$80=106.239; p=0.027) and research ($X^2$2=97.601; p=0.048). Country of origin effects are still poorly understood in most disciplines (Verlegh & Steenkamp, 1999) and are scarce in the field of evaluation with regards to approaches and methodological choices is a grey area. Country of origin effect has received attention in the field of consumer behaviour, specifically in the area of product evaluation (Verlegh, 2001; Peterson & Jalibert, 1995).

Lessons from such studies suggest that the country of origin may have implications on methodical choices by evaluations in three main domains, namely cognitive, affective and normative. The cognitive dimension asserts that country of origin provides an epistemic cue for approaches and methodological choices while the affective element assumes that the country of origin has symbolic and emotional value to evaluators. The normative dimension asserts that evaluators hold social and personal norms related to country of origin. These factors are enshrined in the socio-economic, social political and epistemological beliefs in the countries of origin. This does suggest that further exploration into evaluators roots and their methodological choices is merited, given that the evidence in the AfrED database does confirm a link between evaluators' country of origin and their selection of evaluation methods and approaches.

## Epistemology, identity and implications for national evaluation systems

The hypothesis of this chapter is that the identity of evaluation practitioners would play a key role in the epistemology of NMES in the region and, as a result, will heavily influence public-sector orientation and practice. While the survey results are not conclusive, analysis at this point indicates that this is not categorically the case, but may be relevant in specific situations. It then becomes important to interrogate the relationship between institutions, identity and epistemology. If individual identity is not driving the epistemology of institutions, how are these constructed and how can this be changed?

The AfrED survey offers a few hints about this. One interesting finding is a convergence between professional qualifications as an evaluator and professional careers in evaluation. Of those who received an MA degree in evaluation, 43% consider evaluation their primary professional identity. This correlation between higher-education offerings in evaluation and professional

identity in evaluation indicate that there are important roles institutions for higher education can offer in crafting professional identities, and possibly creating some of the institutional mechanisms for reinforcing them, such as professional associations, journals, etc. This is bolstered by the level of experience reported by respondents to the survey: nearly 70% of evaluators and evaluation commissioners have ten years or less experience in the field demonstrating that evaluation remains a largely emergent profession in the region.

An additional important finding is when asked whether there was a Made in Africa body of research on evaluation that they could identify, 64% did not believe this was the case. While this could possibly be linked to the level of awareness of the approach, these results indicate that there is relatively little interest, ownership, or engagement by evaluation practitioners or commissioners in African epistemological approaches. At the same time, this is contradictory to the experience of the African Evaluation Association annual conference in 2016, where the Made in Africa stream received the most paper submissions and hosted several of the best-attended sessions at the conference. This dilemma needs further investigation.

## Conclusion

This chapter explores the epistemological identity of African evaluators within the context of growing focus on NMESs and indigenisation of knowledge within Africa. It is evident that the emergent profession of evaluation in Africa is developing in tandem with public-sector monitoring and evaluation systems. Each is driving the other's growth and both are being catalysed by a range of internal and external factors. NMESs are first and foremost a reflection of the existing epistemology of state institutions, through the way planning processes are designed and targets are set and measured. However, they are also opportunities to pivot this belief system and realign an institution to more appropriate knowledge systems. It was observed that NMESs are evolving with a trajectory of indigenisation of knowledge from a Eurocentric focus towards African-rooted knowledge systems.

This chapter utilises data from a survey conducted using contacts from the AfrED database. Three research questions were addressed: (a) Are most evaluators undertaking evaluations in Africa from the West?; (b) What are the

perceptions on the Made in Africa Evaluation approaches?; and (c) What is the association between various attributes of identity and choice of methodological approaches by African evaluators? While the held belief in the monitoring and evaluation fraternity is that most evaluators are from the North, results from this study showed a pattern contrary to these assertions. Results indicated that a majority of the respondents (64%) did not believe in a unique, Made in Africa approach to evaluation practice. This is understandable given the nascent nature of the evaluation profession and the epistemic beliefs of most of the evaluators coming from common training background. This highlights the need for more innovative approaches to better understand a Made in Africa agenda, to allow meaningful strides in the decolonisation of evaluation knowledge discourse.

Results indicated significant associations amongst professional identity, approaches and methods. For example, results on evaluation approaches indicated that individuals who practised evaluation in any capacity significantly utilised participatory evaluation approaches. Researchers significantly utilised naturalistic evaluations and randomised control trials (RCTs) while managers utilised empowerment and participatory evaluations. Results on methods, for example, indicated that evaluators at any capacity significantly preferred individual face-to-face interviews, secondary data analysis and stakeholder meetings/consultations. Researchers preferred individual face-to-face interviews, secondary data analysis, stakeholder consultations and surveys. These results highlighted the importance of professional identity in the selection and utilisation of evaluation methods and approaches. There is a general bias towards the use of qualitative evaluation approaches and methods by African evaluators, which was confirmed by the study. These findings tend to suggest that efforts to promote the Made in Africa Evaluation agenda might need to make deliberate efforts on promoting approaches with a qualitative epistemological grounding.

Within the indigenisation of evaluation knowledge discourse, two challenges need to be addressed. The first challenge is the integration of indigenous knowledge in M&E to extend beyond a tool for professional practitioners and rather be integrated by all stakeholders in the evaluation process. The most pivotal element of this is setting targets and looking at results in a way that acknowledges contextually relevant development pathways. Another essential component is to integrate training on contextually relevant values

and management systems into burgeoning higher-education offerings in monitoring and evaluation. This should include action research about which methods, tools and approaches are most contextually relevant, and how they can be piloted and refined according to an indigenous values system. The second challenge is to continue to rigorously interrogate these shifts in practice and perspective, to better understand the extent to which this changes results. The M&E field is in a place with so much momentum that it is possible to build an empirical basis for both the need and impact that a foundation of indigenous knowledge would have on monitoring and evaluation systems and the subsequent programme outcomes.

## References


Abrahams, M.A. 2015. A review of the growth of monitoring and evaluation in South Africa: Monitoring and evaluation as a profession, an industry and a governance tool. *African Evaluation Journal*, 3(1):1-8. https://doi.org/10.4102/aej.v3i1.142

Audi, R. 2003. *Epistemology: A contemporary introduction to the theory of knowledge*. New York: Routledge.

Avalos, B. & Aylwin, P. 2007. How young teachers experience their professional work in Chile. *Teaching and Teacher Education*, 23(4):515-528. https://doi.org/10.1016/j.tate.2006.11.003

Babbie, E. & Mouton, J. 2010. *The practice of social research*. Tenth edition. Cape Town: Oxford University Press.

Beijaard, D. 1995. Teachers' prior experiences and actual perceptions of professional identity. *Teachers and Teaching: Theory and Practice*, 1(2):281-294. https://doi.org/10.1080/1354060950010209

Bledsoe, K. & Donaldson, S.I. 2015. *Culturally responsive theory-driven evaluation. Continuing the journey to reposition culture and cultural context in evaluation theory and practice*. Charlotte: Information Age Publishing.

BonJour, L. 2010. *Epistemology: Classic problems and contemporary responses*. Maryland: Rowman & Littlefield.

Bourgeois, I. & Nare, C. 2015. The 'usability' of evaluation reports: A precursor to evaluation use in government organizations. *Journal of MultiDisciplinary Evaluation,* 11(25):60-67.

Burke, P.J. & Reitzes, D.C. 1991. An identity theory approach to commitment. *Social Psychology Quarterly*, 54(3):239-251. https://doi.org/10.2307/2786653

Chilisa, B. 2012. *Indigenous research methodologies*. Los Angeles: SAGE.

Chilisa, B.; Major, T.E.; Gaotlhobogwe, M. & Mokgolodi, H. 2016. Decolonizing and indigenizing evaluation practice in Africa: Toward African relational evaluation approaches. *Canadian Journal of Program Evaluation,* 30(3):313-328. https://doi.org/10.3138/cjpe.30.3.328

Chilisa, B. & Malunga, C. 2012. Made in Africa Evaluation: Uncovering African roots in evaluation theory and practice. Bellagio Centre. Report of the African Thought Leaders Forum on Evaluation for Development: Expanding Thought Leadership in Africa. Unpublished conference background paper. 14-17.

Cloete, F. 2016. Developing an Africa-rooted programme evaluation approach. *African Journal of Public Affairs*, 9(4):55-70.

Cloete, F.; Rabie, B. & De Coning, C. (eds). 2014. *Evaluation management in South Africa and Africa*. Stellenbosch: African Sun Media. https://doi.org/10.18820/9781920689513

Connell, R. 2016. Decolonising knowledge, democratising curriculum. Paper read at University of Johannesburg discussions on Decolonisation of Knowledge, March.

Crotty, M. 1998. *The foundations of social research: Meaning and perspective in the research process*. London: SAGE.

Demerath, L. 2006. Epistemological identity theory: Reconceptualizing commitment as self-knowledge. *Sociological Spectrum*, 26(5):491-517. https://doi.org/10.1080/02732170600786208

Dobson, P. 2002. Critical realism and information systems research: Why bother with philosophy? *Information Research*, 7(2).

DPME (Department of Planning, Monitoring and Evaluation). 2014. *Strategic Plan 2015–2020*. Pretoria: Government Printing Works.

Engela, R. & Ajam, T. 2010. Implementing a government-wide monitoring and evaluation system in South Africa. *ECD Working Paper Series, No. 21*. Washington, DC: The World Bank.

Girei, E. 2017. Decolonising management knowledge: A reflexive journey as practitioner and researcher in Uganda. *Management Learning,* 48(4):453-470. https://doi.org/10.1177/1350507617697867

Goldman, I.; Mathe, J.E.; Jacob, C.; Hercules, A.; Amisi, M.; Buthelezi, T.; Narsee, H.; Ntakumba, S. & Sadan, M. 2015. Developing South Africa's national evaluation policy and system: First lessons learned. *African Evaluation Journal*, 3(1):1-9. https://doi.org/10.4102/aej.v3i1.107

Hall, P.A. 2010. African American studies: Discourses and paradigms. In: J. Davidson (ed.). *African American Studies*. Edinburgh: Edinburgh University Press. 15-35.

Hamlyn, D.W. 1995. Epistemology, history of. In: T. Honderich (ed.). *The Oxford Companion to Philosophy*. Oxford: Oxford University Press. n.p.

Heleta, S. 2016. Decolonisation of higher education: Dismantling epistemic violence and Eurocentrism in South Africa. *Transformation in Higher Education,* 1(1):1-8. https://doi.org/10.4102/the.v1i1.9

Hogg, M.A. & Abrams, D. 1988. *Social identifications: A social psychology of intergroup relations and group processes*. London: Routledge.

Hong, J.Y. 2010. Pre-service and beginning teachers' professional identity and its relation to dropping out of the profession. *Teaching and Teacher Education*, 26:1530-1543. https://doi.org/10.1016/j.tate.2010.06.003

McDougal, S. 2014. Africana studies' epistemic identity: An analysis of theory and epistemology in the discipline. *Journal of African American Studies*, 18:236-250. https://doi.org/10.1007/s12111-013-9265-2

Morelli, P.T. & Mataira, P.J. 2010. Indigenizing evaluation research: A long-awaited paradigm shift. *Journal of Indigenous Voices in Social Work*, 1(2):1-12.

Nobles, W.W. 2006. *Seeking the Sakhu*. Chicago: Third World Press.

Patel, M. 2013. African evaluation guidelines. *African Evaluation Journal*, 1(1):1-5. https://doi.org/10.4102/aej.v1i1.51

Patton, M.Q. 2008. *Utilization-focused evaluation*. Saint Paul, MN: SAGE.

Peterson, R.A. & Jalibert, A.J.P. 1995. A meta-analysis of country-of-origin effects. *Journal of International Business Studies*, 26(4):883-900. https://doi.org/10.1057/palgrave.jibs.8490824

Rosenstein, B. 2013. December. Mapping the status of national evaluation policies. In: Parliamentarians Forum on Development Evaluation in South Asia and EvalPartners. Unpublished consultancy report. http://www.pfde.net/index.php/publications-resources/2014-02-28-19-05-00 [Accessed 18 September 2018].

Rowe, M. & Oltmann, C. 2016. Randomised controlled trials in educational research: Ontological and epistemological limitations. *African Journal of Health Professionals Education*, 8(1):6-8. https://doi.org/10.7196/AJHPE.2016.v8i1.683

Sanderson, I. 2002. Evaluation, policy learning and evidence-based policy making. *Public Administration*, 80(1):1-22. https://doi.org/10.1111/1467-9299.00292

Stets, J.E. & Burke, P.J. 2000. Identity theory and social identity theory. *Social Psychology Quarterly*, 63(3):224-237. https://doi.org/10.2307/2695870

Torgerson, C.J. 2002. Educational research and randomised trials. *Medical Education*, 36(11):1002-1003. https://doi.org/10.1046/j.1365-2923.2002.01335.x

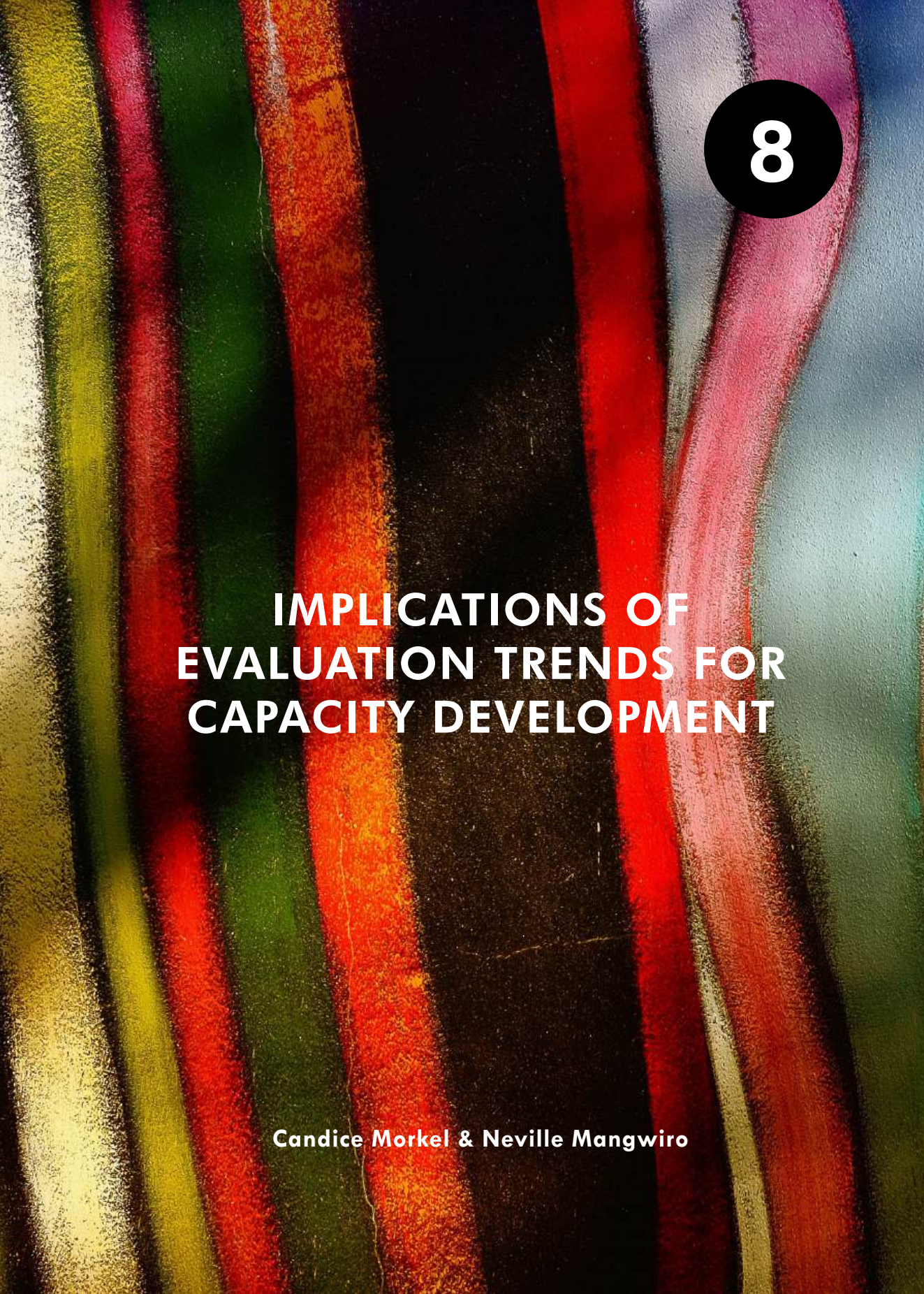
Verlegh, P.W.J. 2001. Country of origin effects on consumer product evaluations. Master's thesis, University of Wageningen.

Verlegh, P.W.J. & Steenkamp J.E.M. 1999. A review and meta-analysis of country-of-origin research. *Journal of Economic Psychology*, 20:521-546. https://doi.org/10.1016/S0167-4870(99)00023-9

Vigneswaran, D. & Quirk, J. 2015. *Mobility makes states*. Philadelphia: University of Pennsylvania Press. https://doi.org/10.9783/9780812291292

Walters, H. 2007. Capacity development, institutional change and theory of change: What do we mean and where are the linkages. Preparatory paper for the seminar "Scrutinizing Success and Failure in Development", Wageningen University, The Netherlands.

Wilson, S. 2008. *Research is ceremony: Indigenous research methods*. Manitoba: Fernwood.



8

# IMPLICATIONS OF EVALUATION TRENDS FOR CAPACITY DEVELOPMENT

Candice Morkel & Neville Mangwiro

## Introduction

Evaluation capacity development (ECD) remains one of the most sought-after yet elusive outcomes in the evaluation sector (Picciotto, 1998:39; Stockdill, Baizerman & Compton, 2002:16). Although there are extensive efforts to strengthen evaluation capacity in Africa, the precise mechanisms by which evaluation capacity development strengthens evaluation practice is largely unknown. The absence of empirical evidence which confirms the link between ECD and improvements in evaluation practice exacerbates the challenge (Preskill & Boyle, 2008; Tarsilla, 2014). Regardless of the challenges, the sector continues to recognise ECD as one of the primary solutions to the persistent weaknesses and limitations in the demand and supply of monitoring and evaluation (M&E) capacity on the continent (Wao, Onyango, Kisio, Njatah & Onyango, 2017:1; Porter & Goldman, 2009).

Research and practice have shown that there has been an absence of coherent strategies or models for ECD in Africa, resulting in fragmentation and piecemeal approaches to capacity development (Basheka, 2016:115; Tarsilla, 2014:8; CLEAR-AA, 2016c:25). The lack of harmony in short-term training or formal graduate education in M&E across the continent could, in part, be attributed to the absence of agreed-upon norms and standards for M&E education provision, and uncertainties around how to build capacity for evidence use, or capacity development in general (Stewart, 2015:555; Denney & Mallet, 2017:1; Preskill, 2014:118). This is despite the existence of guidelines for evaluator competencies provided by, for example, the International Development Evaluation Association (IDEAS), the African Evaluation Association (AfrEA) and the American Evaluation Association's (AEA) Joint Committee on Standards for Educational Evaluation.

A superficial analysis of the problems associated with current approaches to ECD will lead some to wrongly assume that the problem rests with those who are responsible for conducting evaluations – commonly termed the "supply side". However, proponents of systems-thinking include the "demand side" in their approach to ECD – which incorporates the institutional capacity and appetite for conducting, institutionalising, and using evaluations. The terms 'supply' and 'demand' in evaluation have therefore become widely applied (Porter & Feinstein, 2014:7; Merkle, 2016:2; Porter & Goldman, 2013; Mulenga & Porter, 2013:3; Development Bank of Southern Africa, African Development Bank

and World Bank, 2000:53) and the matching of the two has become common in the discourse on ECD (Smith & Morkel, 2018:41). However, assumptions that evaluation supply and demand mechanisms work in the same way as those of other products, services or commodities in capitalist markets may not hold. It is critical to understand the interplay between the drivers and contextual issues that mediate between supply and demand that can assist in developing effective solutions to ECD challenges (Blaser Mapitsa & Khumalo, 2018:4).

This chapter examines empirical results of evaluation reports from the AfrED database in order to unpack the relationship between the demand for evaluations and the capacities needed to meet that demand. The analysis further explores ways in which current M&E training and education provision can be enhanced to respond to capacity development needs. In achieving its objectives, the chapter also draws evidence from a secondary analysis of the results of a survey of evaluation practitioners' perceptions of ECD challenges in the sector. The primary question guiding the chapter is: How can a snapshot of trends in evaluation supply, demand and practice provide an indication of the kinds of skills and capabilities that may be required of evaluators in the future and serve as a guide for those tasked with ECD?

## Brief background to evaluation capacity development

Capacity development has been on the African agenda for the past six decades, yet contestations persist around the concept itself (Taut, 2007:120; Denney & Mallett, 2017:5; Kwang-Ho, 2009:201; Kotvojs & Hurworth, 2013:5; Compton & Baizerman, 2007:118). ECD in particular has been on the evaluation agenda for at least two decades, despite not yielding the expected results (Kotvojs, 2017:14). Many continue to highlight the weaknesses in evaluation 'supply', which is generally defined as the individuals who must produce evaluations (Wao, Kisio, Njatha & Onyango, 2017:1; Merkle, 2016:4; Holvoet, 2012:9; UNICEF, 2008:121; Podems, 2015:10). Findings from a 2005 study involving evaluation practitioners in South Africa, for example, illustrated the widely held opinion that both individual evaluation capacities and the quality of evaluation training were weak (Abrahams, 2015:4). Almost ten years later, a survey conducted on evaluation capacity in the South African government yielded similar results (Podems, Goldman & Jacob, 2014:76). Despite more than three decades of capacity building in Africa, and an increasing intensity in building evaluation capacity in particular, research has shown that the problems of capacity continue to

persist (Wao, Onyango, Kisio, Njatha & Onyango, 2017:1; Basheka, 2016:95). It is plausible that inadequate diagnoses of capacity needs may result in ever-increasing ECD interventions, without the concomitant increase in evaluation capacity (Morkel & Ramasobana, 2017:7; Tarsilla, 2014:5). The scant use of diagnostic information to inform ECD interventions may also have a role to play in this (Kotvojs & Hurworth, 2013:5; Kotvojs, 2017:13). Although there is insufficient evidence to point to the exact reason for the mismatch, there is evidence that many ECD programmes do not implement systemic approaches that are known to be more effective, but instead continue to rely on short-term training programmes (Tarsilla, 2014:5).

The African Capacity Building Foundation, in its Africa Capacity Report 2019, defines capacity as the "ability of people, organizations, and society as a whole to manage their affairs successfully" (ACBF, 2019:1). Capacity development, therefore, is broadly defined as a process by which "people, organizations, and society as a whole unleash, strengthen, create, adapt, and maintain capacity over time". The contestation around the concept of ECD has added to the challenge. Some scholars and practitioners prefer to use the term 'evaluation capacity *building*' (ECB), whilst others prefer 'evaluation capacity *development*' (ECD). Stockdill, Baizerman and Compton (2002) use the term ECB, and define it as: "a context-dependent, intentional action system of guided processes and practices for bringing about and sustaining a state of affairs in which quality programme evaluation and its appropriate uses are ordinary and ongoing practices within and/or between one or more organisations/ programs/sites". Others (Preskill & Boyle, 2008:444; Brinkerhoff & Morgan, 2010:3) argue that ECB is a systems-level intervention that incorporates both individual and institutional components. Tarsilla (2014:2), on the other hand, proffers that evaluation capacity *building* usually refers to the limited types of interventions targeting individuals, whilst evaluation capacity *development* focuses on effecting institutional and systems-wide changes. In this chapter, the term evaluation capacity development is used. It encompasses the entire spectrum of initiatives an organisation undertakes to strengthen evaluation systems, including the enabling environment. Where the intention is to refer to the building of individual human capacity in particular, the term 'training' is used explicitly.

According to Wubneh (2003:169-170), there are four dimensions to what is termed "capacity building", which would fit the definition of capacity development used in this chapter. The first dimension is the development

of human capacity, and focuses on the development of individual skills and abilities; the second is value system re-organisation; the third is institutional restructuring, whilst the fourth is the review of organisational development practices. Though Wubneh (2003:169-170) focuses on the development sector in general, the typology easily translates to the evaluation sector, as the aim of ECD is to change the way an individual and an organisation behaves, in order to observe positive changes in practice.

Taking guidance from these definitions, it is clear that short-term training becomes a necessary, but not sufficient, part of ECD Interventions. Such interventions need to be combined with an integrated strategy to enhance individual and organisational capability if they are to be effective at a systems level. It is argued within the evaluation community that ECD is multi-faceted and systemic (Lennie, Tacchi, Wilmore & Koirala, 2015:326; Stockdill, Baizerman & Comption, 2002; Preskill & Boyle, 2008). Research has shown that training constitutes only one component of building competence as well as confidence in evaluation (Bamberg, Perlesz, Mckenzie & Read, 2010). Volkov and King's (2007) Checklist for Building Organisational Evaluation Capacity is instructive in this respect. It outlines eight guidelines for organisational evaluation capacity building, which have little to do with training. Examples of the guidelines include: cultivating a positive, ECD-friendly internal organisational context; purposefully creating structures to facilitate the development of evaluation capacity; building and reinforcing the infrastructure to support evaluations and securing financial resources to support an evaluation programme in the organisation (Volkov & King, 2007:1-3). The definition provided by Labin, Duffy, Meyers, Wandersman & Lesesnen (2012:308) also illustrates how ECD needs to be pursued beyond the individual, with their definition focusing on "an intentional process to increase individual motivation, knowledge, and skills, and to enhance a group or organisation's ability to conduct or use evaluation". Leviton (2013:93), in a paper on evaluation capacity building, poses a reminder that "training and technical assistance are a process, just as the experience of doing an evaluation is a process". Proponents of ECD would likely agree that training is only one aspect of developing evaluation capacity, however, a much narrower view is often adopted in the planning and execution of ECD interventions. In many cases, training (typically in the form of short-term courses) becomes the lodestar of ECD. There is a need for more integrated approaches to ECD, which apply existing knowledge about the effectiveness of combining multiple strategies and tactics and incorporating them into ECD interventions.

A more long-term view must also be adopted in thinking about ECD if efforts at strengthening evaluation capacity are to be sustainable

## Looking back to move ahead: What capacity development in Africa has taught us

### *History of capacity development in Africa and the influence on ECD*

The historical roots of capacity development in Africa are well-documented (Basheka, 2016; Denney & Mallett, 2017; Development Bank of Southern Africa, African Development Bank and World Bank, 2000). The post-independence Africa of the 1960s was characterised by structural adjustment programming and the influence of Bretton Woods institutions, which shaped the interventions undertaken to ensure the continent's growth and development (Basheka, 2016). Assisting the African continent to emulate the growth trajectories of the 'West' became the focus of the international development community as the ideal model of addressing the continents' poverty and social welfare crisis (Denney & Mallett, 2017:6). The decade of the 1960s was also punctuated by the growing popularity of the modernisation theory of development, which incorrectly assumed that Western modes of materialism, economic growth and the principles of the "free market economy" would release the continent from the grip of under-development (Davis, Theron & Maphunye, 2009; Denney & Mallett, 2017). Technical assistance or "technical cooperation" therefore became synonymous with building the capacities of developing countries through "international experts" who were meant to transfer their skills to local counterparts (Demongeot, 1994:479).

Tarsilla (2014:2) argues that ECD remains donor centric and reproduces the old, self-serving pattern of strengthening local staff of international organisations rather than strengthening local, contextually relevant ownership of evaluation. Evaluation capacity development in Africa has its history in the 1980s and 1990s (Basheka & Byamugisha, 2015:79-80). A tidal wave of initiatives solidified the importance of evaluation capacity development in Africa and a host of institutions, such as the African Evaluation Association (AfrEA) and others were established in order to focus on building the capacity of evaluation in Africa. Preskill and Boyle (2008:443) suggest that the first decade of the 21st century marked the beginning of the wave of interest in and focus on evaluation capacity building, particularly in North America.

Despite the rising tide of criticism against the weaknesses in capacity development in Africa and the absence of evidence that these have any impact on building national capacity at all (Demongeot, 1994:479), capacity development initiatives (and their budgets) continue to increase (Wubneh, 2003:166; Denney & Mallett, 2017:v; Kotvojs, 2017:13). For example, the International Monetary Fund (IMF) spends about one quarter of its budget on capacity development activities, with direct spending on capacity development growing by 6% from 2015 to 2016 (IMF, 2017:1-2). Between 1991 and 2016, the African Capacity Building Foundation (ACBF) committed US$700 million to capacity development in sub-Saharan Africa alone (ACBF, 2016:5), and the global investment in capacity development now exceeds US$30 billion annually (Kotvojs, 2017:13), despite yielding less than stellar results (Kotvojs & Hurworth, 2013:5; Kotvojs, 2017:14). Programmes such as the CLEAR Initiative, the International Programme on Development Evaluation Training (IPDET), capacity building by voluntary organisations for professional evaluation (VOPEs) and universities have flooded the development community – yet the weaknesses in evaluation capacity and evidence use in Africa persist (Stockdill, Baizerman & Compton, 2002:16; Stewart, 2015:550; Denney & Mallett, 2017:v). Despite the misalignment between the proliferation of ECD interventions and strengthened evaluation practice, ECD is still seen as key to improving institutional capacities for evaluation planning and to produce high quality evaluations (Merkle, 2016) Therefore it is necessary to continue to produce evidence on "what works" in ECD so as to move beyond the impasse.

## *Challenges in the landscape of ECD efforts in Africa – what capacity, for whom?*

Whereas twenty years ago it would have been less easy to find a suitable M&E training programme on the African continent, in recent times training opportunities have rapidly expanded for anyone desiring professional development in this area. There is an ever-growing variety of ECD actors in Africa, including private training organisations, individual consultants, international funding agencies, international development institutions, voluntary organisations for professional evaluation (VOPEs) as well as institutions of higher learning. Some attempts have been made to develop databases of the various evaluation education opportunities available globally as well as in Africa. For example, in 2017, a specialised task force of the Global Goals for Sustainable Development identified 31 institutions across the world that provide capacity development programmes in evaluation, ranging from

academic institutions to multilateral organisations and civil society organisations (Agrawal, Rao & Kumra, 2017). A similar mapping exercise in 2016 charted 45 academic institutions across Anglophone Africa with programmes in M&E (CLEAR-AA, 2016a:24). However, there is no single, consolidated database or inventory of the myriad ECD opportunities and service providers in Africa or elsewhere, with little efforts at harmonisation and few attempts to measure their effectiveness.

Measuring and defining evaluation capacity is also an area of contestation. Competency frameworks, such as the Canadian Evaluation Society Competencies for Canadian Evaluation Practice and the Evaluation Competency Framework of the South African Department of Planning Monitoring and Evaluation provide an indication of what constitutes a good evaluator (Buchanan & Kuji-shikatani, 2014:34; Podems, Goldman & Jacob, 2014:81). However, these are not necessarily universally applicable and usually focus on the competency of the individual evaluator. Just over a decade ago, there was a disproportionate focus on individuals' technical skills in making M&E systems work (Blaser Mapitsa & Khumalo, 2018:2) which may have influenced the dominance of individualised technical training to this day. However, it is widely recognised that evaluators need a broad range of skills, and that evaluation capacity is not limited to the individual, but applies to organisations as well (Wao et al., 2017:1; Lucas, 2013:6; Mapitsa Blaser & Khumalo, 2018:3; Edwards, Stickney, Milat, Campbell & Thackaway, 2016:265). Although thinking has transformed from an individualistic perspective on evaluation capacity towards a more organisational one (Labin, 2014; Volkov & King, 2007; Bourgeois, Toews, Whynot & Lamarche, 2013; Lennie, Tacchi, Wilmore & Koirala, 2015), individualised training programmes, focusing on technical aspects of evaluation continue to dominate ECD interventions, particularly in Africa (Tarsilla, 2014:4).

Two additional ECD short-comings in Africa were identified by Tarsilla (2014). One of these is the 'parachuting' of evaluation expertise from outside the continent, which strangles the development of local evaluation capacities in Africa (Tarsilla, 2014:6; CLEAR-AA, 2016a:3). Tarsilla also laments that most of the continents' ECD programmes are based on content developed outside of Africa, most notably North America, Europe and Australia, with little or no contextualisation to the African context.

It is beyond the scope of this chapter to analyse ECD interventions in Africa in its entirety, as it cannot do justice to its complex and emergent nature. The systemic

nature of ECD is highlighted here so as to reinforce the point that Africa's evaluation capacity problems do not lie solely in improvements in evaluation education and training. The limitation, therefore, is that the data do not provide for an organisational or systemic perspective of ECD, highlighting the need for more empirical evidence on what needs to be done to improve evaluation capacity in Africa.

This chapter assesses the landscape of evaluation demand from selected AfrED evaluations to understand the kind of capacity required to conduct evaluations in Africa for the period between 2005 and 2015. An analysis of a survey conducted by CREST (Centre for Research on Evaluation, Science and Technology) with over 500 M&E practitioners in 2016 during the compilation of the AfrED provides a snapshot of the evaluation skills and capabilities that are available on the continent. Although the data may not be generalisable, it does provide an account of evaluation practitioners' view of ECD, and opens up the opportunity for further research on how ECD providers need to respond to the growing demand for evaluation services on the continent.

## Approaches and methods

This chapter adopts a mixed-methods approach to explore trends about the effectiveness of evaluation capacity development interventions in Africa. Both qualitative and quantitative data were used, namely (i) literature on ECD and ECB; (ii) the results of an online survey conducted in 2016 with 564 evaluators/ M&E practitioners as part of the African Evaluation Database (AfrED) project; and (iii) the repository of evaluations on the AfrED.

An online survey was conducted between 10 and 29 November 2016 by CREST to 3032 individuals via an e-mailed link. These included e-mails listed in the evaluation reports on the African evaluation database; e-mail addresses from the South African, Ethiopian, Ghanaian and African evaluation associations; and e-mail addresses from CREST's own internal database. A total of 564 individuals completed the survey, amounting to a 22% response rate. The survey contains items relevant to this study (e.g. M&E qualifications and training; evaluation as a professional identity; and primary involvement in evaluation). Quantitative data relevant to the research questions were analysed using descriptive statistics. The open-ended questions in the survey were included as part of the qualitative data and analysis. In the survey, an open-ended question was posed to participants, asking respondents what they thought is the key to large scale,

quality evaluation capacity development in Africa. A total of 407 responses were received for this question. Inductive coding was used to analyse the qualitative data and determine the main codes, categories, and themes emerging from the data.

The AfrED currently has a total of 2635 documents, with a geographical scope of twelve sub-Saharan Anglophone countries. A random sample of evaluations was retrieved from the database using systematic sampling. Every 30th evaluation report on the complete list of evaluations, per country, were selected using a random start. A total of 59 (n=59) evaluation reports were analysed using deductive coding. A set of *a priori* codes was developed, which included the following: evaluation type; evaluation design; data-collection methods; data-analysis method; evaluand country; sector; funder; government as client; lead evaluator local/international; lead evaluator consultant/private; and role of local stakeholders. The literature review on ECB and ECD provided the framework for the analysis and discussion of the survey responses as well as the evaluation reports. The subsequent section addresses the findings of this study.

## Results and discussion: Snapshots of supply and demand and trends to guide the future of ECD

### *Evaluation supply trends – findings from the survey*

#### *The need for more training in technical competencies, with practical experience and opportunities for real-world practice*

Evaluators in Africa appear to be highly qualified, yet most of the responses to the question on the key to large scale, quality ECD in Africa revolved around accessing more training. Despite the fact that more than half of the respondents were in possession of a master's degree (55%), and a smaller percentage possessed a doctorate (17%), the stated need for additional training was significantly high (123 out of 407 responses). Further exploration is needed to determine whether the content of ECD interventions are appropriately responsive to the differentiated needs of evaluation practitioners on the continent, as 76 respondents possessed a master's degree in evaluation, whilst 126 respondents possessed a diploma in evaluation studies. Respondents were actively working in evaluation practice, and either conducted (n=185), managed (n=113) or provided technical assistance in evaluations (n=91). These individuals were more likely to have a diploma in evaluation studies (between 48 and 60%)

than any other qualification. This was followed by a master's degree in evaluation (between 37% and 51%). Those who trained others in evaluation (n=99) were equally as likely to have a diploma or a master's in evaluation studies (47%).

Respondents also reflected on a need for training in technical skills for evaluators such as project management, data management as well as quantitative and qualitative methods. This negates the view that the pendulum has swung in favour of ECD interventions that are contextual, and less technical. There appears to be a continued demand for training in fundamental technical aspects of M&E. For example, one respondent stated a need for training in: "evaluation design, outcome and impact evaluation, theory of change and results based evaluation methodologies, and using evaluation as evidence for decision-making". Another respondent stated that what was required is a "model that combines training, practice and reflection over a sustained period of time". This last statement is reflective of research that has shown that experiential learning is an important but insufficiently practiced component of ECD (Leviton, 2013:93; Tarsilla, 2014:6). More significantly, African practitioners who participated in Tarsilla's 2014 study viewed the practical utility of training courses as 'minimal', and despite facilitators' expertise, the training programmes provided no opportunity to put their learning into practice. It comes as no surprise, therefore, that respondents preferred training that provided opportunities for "real-world practice".

A second cluster of statements emerged from the data specifically around the development of skills in management and analysis of data. It appears that there are concerns amongst evaluators around the quality of data, and a need for training on the production and analysis of data as well as management information systems to support these. Examples of statements from respondents in response to the question on how to remedy this include:

> Provide capacity on techniques of monitoring database management that can fit into evaluation process"; "Recruitment of competent data collectors, use of real-time data-collection devices, effective training, co-ordination and close monitoring and follow-up"; and "In my opinion the key lies in having access to quality of data which assist in arriving at evidence-based conclusions.

Significantly, the second highest cluster of responses (110 out of 407) was categorised under the broad category of "Continuous professional development of the demand and supply side" (including leadership and evaluation culture), which also links to training. The statements on the need for more

training, combined with this set of statements, represent more than half of the statements responding to the question on what Africa needs for high-quality ECD. This seems to indicate a great need for ongoing capacity building of both evaluators as well as those commissioning and managing evaluations. It is possible that the need for continuous professional development drives the popularity and unceasing demand for evaluation training in Africa. Therefore, a more nuanced view of ECD needs is required. ECD providers need to be cognisant of the needs of individuals in order to respond appropriately. Differences between sectors (particularly the state and civil society) and the practice and utilisation of M&E also matter. ECD providers therefore need to ensure that the design of ECD programmes are appropriately targeted to participants, particularly at post-graduate level. For example, whereas the fundamentals of research methodology, sampling and data-processing may be necessary for those who have not yet been trained in these areas, it could be an ill fit for participants with master's and PhD qualifications. Ongoing professional development for the latter cohort would need to provide learning opportunities on advanced, innovative and emerging issues in evaluation, rather than courses focusing on technical skills development of this nature. Evaluation theory and specialised curriculum design would be important for more advanced ECD. The importance of appropriate targeting is therefore emphasised, and the use of diagnostic exercises, pre-tests, training-needs analyses or skills audits would be useful before the design of any ECD programme.

### *Communities of practice*

The remaining categories of statements were comparatively smaller than the statements around training and professional development, yet were large enough to categorise and compare with similarly smaller categories. This also assisted in exploring the hierarchy of ECD needs amongst evaluation practitioners. Forty statements were about collaboration and partnerships. Respondents reflected on the need for professional exchanges between the global North and South, and between senior and emerging evaluators. Mentorship, skills transfer and the building of communities of practice also emerged as important actions that needed to be taken, pointing to a possible realisation amongst respondents that training on its own is not sufficient in building evaluation capacity. Considering that the survey data also showed that the vast majority of respondents in the age category 31 to 40 identified themselves as managers, trainers and commissioners of evaluations, this could be an indication of an emerging need amongst mid-career evaluation

practitioners to form part of professional communities where learning takes place. This is also the cohort that self-identified as evaluators, and considered evaluation as their primary discipline. They are mainly involved in managing, supervising or consulting on evaluations rather than carrying out all evaluation activities. Evaluation education should therefore be cognisant of the particular needs of this cohort of professionals, and steer more ECD interventions towards supporting the strengthening of evaluation management activities, and also consider more innovative means of ECD through, for example, professional exchanges and peer learning.

### *Country and place of work, professional identity, professionalisation and accreditation*

The majority of respondents (62%) self-identified as evaluators in any capacity (this included trainers, students and managers). Evaluation was considered the primary professional identity of a large proportion of the respondents (43%), whilst for 31% of respondents it was considered their secondary professional identity. However, it is significant that most self-identified evaluators were responsible for managing evaluations (39%) or commissioning evaluations (14%). The largest employer for these evaluators was the state (24%), closely followed by the NGO sector (19%). This has important implications for ECD. Primarily identifying as an evaluator implies a much more concentrated focus on evaluation as a discipline, than managing it as a secondary discipline. Similarly, the contextual differences in evaluation practice between the state, national NGOs, international organisations and other sectors also have important implications for ECD. The skills and capabilities required to commission evaluations are different from those required of an evaluator, or someone who designs management information systems to support evaluation practice. For example, a director of agricultural projects who needs to evaluate the success of a micro-finance project may require only foundational theoretical knowledge in order to manage such an evaluation, versus the director of evaluation in the same state department. This once again points to the need for extensive diagnostics in order to differentiate between the various ECD needs and focus areas, whether at an individual or organisational level.

Respondents reported moderate levels of experience in evaluation-related activities. Fewer than half of the respondents (41%) had less than five years of experience in doing evaluation-related work. On average, respondents had about eight years of experience working in evaluation-related activities.

Most respondents had about 6–10 years' experience in the various types of evaluation activities such as conducting evaluations (n=135), planning/contracting evaluations (n=87), technical assistance in evaluation (n=67), training others in evaluation (n=68) and in writing about evaluation (n=51).

Education and training institutions would also need to consider how newly established institutional arrangements for M&E in late-adopting countries would need to shape ECD. Approximately half of those between the ages of 31 and 49 (50%) were more likely to self-identify as primarily evaluators. A much smaller proportion (less than 30%) of those between the ages of 50 and 69 felt that they were primarily an evaluator. This is to be understood in relation to the dominance of South African respondents and the fact that evaluation is considered a relatively 'new' profession in the country. More professional opportunities to work in evaluation have emerged in the last decade and a half in South Africa, and therefore the results may represent this state of affairs. The need for foundational training and development may be more pronounced in these circumstances. How individuals identify their roles and functions in respect of their institutional structure has implications for how evaluation education may need to be shaped.

In terms of place of work, 38% of respondents worked primarily in South Africa as M&E practitioners in various organisations and in a range of capacities, followed by Ghana (9%). Those who worked in South Africa did mostly contractual work with national non-governmental organisations (NGOs) (78% of respondents), whilst 70% were contracted by the state (national) and 69% by international development agencies (some may have multiple clients).

The implications for ECD in the case of evaluation practitioners (who have some experience in conducting a range of evaluation activities) would be around what would constitute effective continuous professional development in strengthening evaluation practice. The challenge in designing curriculum and capacity development strategies is that there are variations amongst professionals in the type of work that they will do, and the amount of time spent on it. For example, a senior public official who is tasked with planning or contracting evaluations has different capacity-building needs from those who are involved in providing technical assistance. ECD programmes need to take these variations into account in designing capacity development interventions, which poses a challenge to the ideal of harmonising competencies and curriculum across the continent (particularly for institutions of higher learning).

The ideal curriculum and ECD programme would need to find the right balance between generic skill development and knowledge building, and specialised training subjects.

The issue around professionalisation and certification was certainly a concern to respondents. The fourth largest category of statements was around professionalisation, certification and building a pool/cohort of professionals (35 statements), whilst the fifth was around strengthening university programmes (34 statements). Respondents were also concerned about cultural representivity in building a pool of professionals. There may be a link between the desire for professionalisation and certification and the statements reflecting a demand for improved university-based evaluation education programmes. A major challenge for the evaluation sector is that few countries have been able to address the contestations that have dominated the professionalisation debate over the decades, as well as the binary positions that have been adopted around whether or not to 'professionalise' or 'credentialise'. A study by Podems (2014) indicated the highly contested and sometimes political nature of the development of competency criteria for evaluators, and has an effect on moving forward on the issue of professionalisation.

The desire for academic institutions to adopt a more definitive role in ECD amongst respondents is congruent with the rising demand amongst evaluation professionals in Africa for more university-based qualifications. It is also aligned to the increase in the number of academic institutions that are building post-graduate qualifications in M&E (as opposed to or in addition to executive short courses). Some of the statements included, for example: "Professionalising the evaluation field and expanding evaluation training institutions" and "Evaluation should be professionalised and made part of university curriculum". Building a cohort of professional evaluators who produce high-quality evaluations (and other related services) is key to addressing the perceived shortage of African evaluators on the continent.

### *Evaluation practice*

In terms of evaluation purpose, 94% stated that they frequently or occasionally used evaluations for decision-making, followed by evaluations to improve practice. Although it is commonly thought that accountability and compliance too often drive an organisation's evaluation purpose, the findings show that this is the least common purpose of evaluation for respondents (24% never used).

In terms of evaluation approaches used, respondents most commonly used participatory evaluation (88% frequently or occasionally used). The least common evaluation approach was feminist evaluation (76% never used). In terms of evaluation types, programme monitoring and process evaluation were most commonly used (93% frequently or occasionally used). The least common type of evaluation was evaluability assessment (44% never used). The most common evaluation method utilised was face-to-face interviews (98% frequently or occasionally used). In respect of the use of methods, various qualitative methods tended to dominate amongst respondents. These included individual interviews, focus groups and stakeholder interviews. Individual face-to-face interviews and document review were the methods most frequently used by most respondents with up to fifteen years of experience.

ECD interventions need to strike a balance between the depth and scope of subject matter covered in training and development programmes. Training programmes, in particular, need to weigh up whether the more pressing need is to develop a broad range of technical skills in specific methods, types, and approaches to evaluation, how much time to spend on evaluation management or whether the need is to provide skills development in only a limited number of specialist areas. The findings show that there is a need to broaden the repertoire of evaluation types, approaches and methods used in evaluation practice, which can be addressed by ECD institutions. The dominant use of interviews must also be more closely examined, as it is a common misconception that it is easier to collect and analyse qualitative data. The limited use of quantitative methods and those which rely on statistical tests may point to an absence of the technical skills that are required to perform these.

Although the second and third most frequently undertaken studies were process and impact evaluations (n=259), it is also commonly known that 'impact' is understood in different ways amongst evaluation practitioners. For example, the findings show that two quantitative approaches (surveys and testing) are less frequently employed, which are a large part of impact evaluation. It would be important that ECD providers pay attention to the need to harmonise the understanding of the concept to ensure that impact evaluations are rigorous, and that commissioners/managers of evaluation have a working understanding of conducting these.

### *Contextualisation of evaluation practice*

An important cluster of statements emerged around participatory methods, localisation and diversity (26 statements). Examples of the statements made by respondents to improve evaluation capacity building on the continent included a need for training participatory impact assessment and participatory monitoring and evaluation as well as capacity building to national evaluators on participatory and result-based monitoring and evaluation. It is important to evaluators that local stakeholders are involved throughout the evaluation process. The illusion of participation of stakeholders in evaluations is often created by the articulation of participatory data-collection methods in evaluation, such as focus groups. These, however, are not sufficient to address the challenge of ensuring that local communities (often the beneficiaries of programmes being evaluated) are sufficiently empowered to participate in meaningful ways in evaluations, beyond data collection. ECD interventions need to intentionally create awareness and capabilities that will strengthen practice in this regard. An illustrative statement from a participant reinforces this notion:

> Furthermore, there is a need for increased capacity to conduct evaluations but equally important is the ability to understand the context of the major stakeholders in the evaluation. In the African context technical knowledge of constructing evaluations will not be enough unless trust is established and relationships are managed at every stage of the evaluation. So leadership ability plays an important role.
>
> (Survey respondent, 2016)

These responses are linked to a number of recurring themes that emerged which were not repeated as frequently as the above (less than twenty statements each), but nevertheless are important issues that have resonated in the discourse on ECD. One of these was the need for training in Made in Africa Evaluation (MAE) and understanding context and culture (sixteen statements). An example of a statement by a respondent echoed the need for ECD content that is rooted in African axiology: "Development of professional competencies for evaluators, development of a body that oversees the evaluation profession and the development of accredited, affordable evaluation courses. All of these must be rooted in African philosophies." Scholars such as Chilisa, Major, Gaothlobogwe and Mokgolodi (2016) posit that there is a need for a greater focus on indigenised, culturally relevant methods and approaches to evaluation in ECD. An endogenous approach to ECD, emerging from African-centric approaches to evaluation, is missing from the current range of ECD interventions (Tarsilla, 2014:8). The absence of this in ECD interventions on

the whole, but curriculum and instructional design in particular, needs to be addressed in order for evaluators to remain relevant and effective in doing evaluation work in Africa.

### *Quality assurance*

Two sets of statements that appeared with equal frequency (21 statements each) were clustered around standardised protocols across Africa as well as funding/resources. The issue of standardisation has been on the evaluation capacity development radar for a number of years. The defining of competencies and the standardisation of evaluation practice is a contested space, and no continent has been able to fully resolve it. Greater efforts by the ECD sector are required to provide an enabling environment for evaluators in Africa to improve the professionalism and quality of evaluation practice. This includes moving past the impasse around whether or not to standardise or to focus instead on the harmonisation of evaluation quality standards. A related and recurring theme deemed important (and linked to the concern over the weaknesses in evaluation competencies on the African continent and in some way to the standardisation of evaluation practice) was clustered around requiring high quality evaluations and quality assurance (ten statements). Examples of statements from respondents included the need for "training in quality assurance"; "Better evaluation quality control"; "More dedicated commitments by evaluators to quality evaluations"; and "Strong evaluation bodies that regulate quality of evaluations and maintain a pool of continuously learning evaluators". With regards to the resource challenge, respondents offered a number of solutions to what they perceive will lead to effective evaluation capacity and practice on the continent. These include: resource pooling, making training more affordable to Africans, as well as scholarships for African evaluators.

## *Evaluation demand – findings from the evaluation database*

The review of the evaluations in the AfrED was an attempt to identify trends in the commissioning and conducting of evaluation to help provide an indication of the "demand environment" for evaluation. This provided insights into the kinds of skills that have been required to meet the demand, and what that means for ECD in Africa.

### *Donor-led evaluations*

Trends in the data emerging from the AfrED point to the practice of placing a premium on the use of external evaluators from the West as opposed to local

consultants (CLEAR-AA, 2016a:3). The data revealed that international evaluators continue to lead the production of evaluation reports on the continent (40 out of the 59 reports that were examined, were led by evaluators from outside of the country). Only one of the cases observed had a combination of local and international evaluators involved. Only five of the cases studied were led by local evaluators. In the vast majority of cases (n=28) local stakeholders were involved only in data collection, closely followed by no role (n=23). This aligns with Tarsilla's (2014) findings on the 'parachuting' of international evaluation expertise onto local communities.

What does this mean for ECD in Africa? It is still unclear why international evaluators dominate the local evaluation landscape. There is some speculation of an enduring perception that there is a lack of qualified, experienced evaluators on the continent. This may not hold water, considering that the data in this chapter revealed the high levels of experience and qualifications (general and evaluation specific) of evaluators in Africa. Another speculation is that African evaluators may not be *au fait* with the specific forms and frameworks that are used by individual donors for evaluating their interventions. There is a need, therefore, to empower African evaluators to respond to terms of reference and bids in a way that will enable them to compete on an equal footing with international evaluators, who may have (through experience) become familiar with donor requirements for evaluations. This may require bilateral organisations and donors to invest in local, country-led evaluation systems and to ensure that there is alignment between international needs for information about performance, and local needs for evidence for decision-making.

### *Evaluation types, design and methods of data collection*

In the vast majority of cases observed (n=21), impact evaluation was the type of evaluation undertaken. The type of evaluation conducted appears to be related to the commissioner of the evaluation. For example, impact evaluations appear to be mostly commissioned by donors. Impact evaluation was also the type of evaluation least undertaken by respondents to the survey, whereas impact evaluations were the most common type of evaluations conducted by donors in the sample of evaluation reports observed. Whether or not a lack of experience and skills in impact evaluation amongst African evaluators exists, must be further investigated. However, the space for international evaluators to dominate local evaluation practice is evident in the area of impact evaluation.

Further research is needed to understand whether these patterns are more generalisable and, if so, why.

A fairly even spread of other types of evaluation were undertaken such as formative evaluation, summative evaluation, process evaluation, and performance evaluation (between one and six cases spread across the various types). Document reviews (n=20), interviews (n=24) and observation or field work (n=15) remained the dominant methods of data collection amongst the evaluations conducted. In a vast majority of cases, a mix of these qualitative methods were used. It is expected, therefore, that thematic content analysis was used as the primary method of analysis in the vast majority of evaluations (n=46).

ECD interventions need to focus on the ability of evaluators to make observations about data that will produce high-quality analyses and evaluation reports that have a high utilisation value. The robustness of evaluation practice is important to commissioners and users of evidence. Evaluation training and education must begin to re-shape the content of curriculum and instructional design towards producing these kinds of results. Much of the curriculum and training in evaluation education (particularly short courses) tend to focus on the fundamentals of M&E, such as designing theories of change, logic modelling, and results-based management. This is important foundational training that is needed for all M&E practitioners, yet evaluators in particular need training that is focused on developing the technical and analytical skills revolving around the "how to" of evaluation. From the two sets of data analysed above, a case may also be made for greater variations in the types of evaluation education and ECD provided to different cohorts of M&E practitioners (for example, those who commission and manage evaluations – who may be more likely be employed in government – and those who conduct evaluations). The standardisation (or harmonisation) of ECD interventions can become a reality if stakeholders and role-players agree on the competencies that different M&E practitioners may need to have in various roles, and the pathways to professional development in each of these cases.

## Conclusion

The evidence points to the continued need for ECD. However, proper assessments and diagnostics must be prioritised to recognise the differentiated capacity building and capacity development needs of individuals and institutions

if evaluation practice is to be strengthened effectively and sustainably. The relatively young evaluator or evaluation practitioner population, both in terms of experience and age, points to a need for ECD to focus on building the profession "from the ground up" in Africa. A large proportion of evaluators have less than five years' experience, and although they are highly qualified in a range of disciplines, there is a gap in their experience in performing a range of M&E-related functions. ECD providers must respond by offering many more opportunities for experiential learning, in addition to classroom-based knowledge building and capacity development. ECD must move beyond classroom-based training and incorporate a broader range of interventions such as mentorships, create institutional structures and provide the resources that facilitate building individual capacities, and provide a supportive culture and context for individual and organisational learning.

There is a need to ensure that ECD programmes become more adept at targeting the right participants, particularly in programmes that would include professionals with a first (or higher) academic qualification. The need for foundational training and development may be more pronounced in some circumstances, whereas more advanced levels of training may be more appropriate in others. Furthermore, evaluation practitioners self-identify differently within the sector (e.g. primarily as evaluators or commissioners of evaluations, public sector or civil society). There is a need for continuous professional development (CPD) programmes that are clear about what learning outcomes are being offered, who is being targeted and which sector the programme would best suit. Practitioners and scholars can then select to participate in programmes and be confident about the results that they will achieve.

University-based training programmes are increasingly becoming important to evaluation practitioners, so academic institutions need to remain relevant in meeting the 'supply-side' desire for more robust, university-based qualifications in M&E. However, these cannot afford to ossify in theory, as respondents clearly articulated the need for opportunities in the "real-world". The inclusion of a practical component which simulates the realities in evaluation practice can no longer be considered a novelty for evaluation education.

Whether it is best to develop a broad evaluation knowledge-base or specific technical skills in conducting evaluations, and how much needs to be known

around evaluation management, is not cast in stone. ECD providers need to strike the ideal balance between foundational/generic capacity development and building specialised capabilities (and may even further disaggregate according to specific sectors). Much more effort must, therefore, be put into building diagnostic studies, training-needs analyses and skills audits as a non-negotiable part of the process of the design of ECD interventions.

The limited range of evaluation types, approaches and methods used in evaluation practice observed in the evaluations examined in this sample of the AfrED perhaps points to a need to broaden the repertoire of evaluators. This includes the limited use of quantitative methods in impact evaluation, which is particularly concerning.

The impasse around standardisation/professionalisation/certification within the sector needs to be transcended if evaluation capacity is to experience a fundamental shift in the right direction in Africa. Evaluation practitioners have clearly articulated their desire for the ECD sector to provide an enabling environment and to take the lead in the professionalisation of evaluation practice in Africa.

The localisation of evaluation practice as well as the empowerment of local evaluators and indigenous peoples (or the 'beneficiaries' of development interventions) beyond tick-box 'participation' in the global South is rapidly gaining momentum. This is taking place within a broader African debate around the decolonisation of curriculum in the academe in general. In order to remain relevant and to respond to this demand for indigenisation and cultural responsiveness, the ECD environment needs to re-calibrate its curriculum and instructional design. Related to this is the need to break free of the decades-long pattern of international evaluators dominating the local evaluation scene. African evaluators need to be empowered to develop the knowledge, experience, and skill to respond to specific donor frameworks and templates for evaluation-related activities in order to compete equally with the international evaluator community. Rather than bemoan the state of affairs, more disruptive ways of addressing the challenges are called for. For example, commissioners of evaluations (particularly donors) could, for example, require as part of evaluation contracts that local evaluators be 'inducted' into the process of an evaluation from the inception of the project as part of a skills-transfer initiative, or that a certain percentage of local evaluators be required to jointly lead the evaluation.

The role of HEIs in evaluation capacity development is crucial, in so far as they are the locus of individual learning and development, and the creation of bodies of knowledge that support the growth of a discipline. However, it is important that a collaborative posture is adopted between all actors who play a role in ECD, as improved qualifications and proficiencies only tell part of the evaluation capacity story. As more organisations raise the demand for academic qualifications in M&E, either in recruitment practices or as part of their internal capacity-building strategies, higher education institutions will need to pay more attention to consensus around the design of competency-based evaluation education programmes, the standardisation of qualifications, and curriculum design.

## References


Abrahams, M.A. 2015. A review of the growth of monitoring and evaluation in South Africa: Monitoring and evaluation as a profession, an industry and a governance tool. *African Evaluation Journal*, 3(1):1-8. https://doi.org/10.4102/aej.v3i1.142

ACBF. 2016. African Capacity Building Foundation in action: 25 Years of capacity development impact across Africa. Unpublished report. https://elibrary.acbfpact.org/acbf/collect/acbf/index/assoc/HASH1d44/8f6b094e/cd2f4197/26.dir/25%20Years%20of%20Capacity%20Development%20impact.pdf

Bamberg, J.; Perlesz, A.; McKenzie, P. & Read, S. 2010. Utilising implementation science in building research and evaluation capacity in community health. *Australian Journal of Primary Health,* 16(4):276-283. https://doi.org/10.1071/PY10027

Basheka, B. 2006. Evaluation capacity building (ECB) in Uganda. *Administratio Publica*, 24(2):95-121.

Bourgeois, I.; Toews, E.; Whynot, J. & Lamarche, M.K. 2013. Measuring organizational evaluation capacity in the Canadian Federal government. *The Canadian Journal of Program Evaluation*, 28(2):1-19.

Brinkerhoff, D. & Morgan, P.J. 2010. Capacity and capacity development: Coping with complexity. *Public Administration and Development*, 30(1):2-10. https://doi.org/10.1002/pad.559

Buchanan, H. & Kuji-shikatani, K. 2014. Evaluator competencies: The Canadian experience. *The Canadian Journal of Program Evaluation*, 28(3):29-47. https://doi.org/10.3138/cjpe.29.3.70

Chilisa, B.; Major, E.T.; Gaotlhobogwe, M. & Mokgolodi, H. 2016. Decolonizing and indigenizing evaluation practice in Africa: Toward African relational evaluation approaches. *The Canadian Journal of Program Evaluation*, 30(3): 313-328. https://doi.org/10.3138/cjpe.30.3.05

CLEAR-AA. 2016. The state, locus, focus and nature of current monitoring and evaluation education and training opportunities in Anglophone Africa. Johannesburg. Unpublished scoping report.

CLEAR-AA & Twende Mbele. 2016a. Strengthening monitoring and evaluation education and training in Anglophone Africa: A regional workshop. Accra, Ghana. Unpublished scoping report.

CLEAR-AA & Twende Mbele. 2016b. Strengthening monitoring and evaluation education and training in Anglophone Africa: A regional workshop. Nairobi, Kenya. Unpublished scoping report.

Compton, D.W. & Baizerman, M. 2007. Defining evaluation capacity building. *American Journal of Evaluation,* 28(1):118-119. https://doi.org/10.1177/1098214006298172

Demongeot, P. 1994. Market-oriented approaches to capacity building in Africa. *Public Administration and Development,* 14(5):479-488. https://doi.org/10.1002/pad.4230140504

Denney, L.; Mallett, R. & Benson, M.S. 2017. *Service delivery and state capacity: Findings from the Secure Livelihoods Research Consortium*. London: Secure Livelihoods Research Consortium.

Development Bank of Southern Africa, African Development Bank and World Bank. 2000. Monitoring and evaluation capacity. Conference discussion paper. Johannesburg, 25-29 September.

Dewatcher, S. & Holvoet, N. 2016. Facing up to (online) fashion and fads … Face-to-face contact is here to stay in M&E capacity building. Evidence from 35 national evaluation societies. *African Evaluation Journal*, 4(1):1-11. https://doi.org/10.4102/aej.v4i1.158

Edwards, B.; Stickney, B.; Milat, A.; Campbell, D. & Thackway, S. 2016. Building research and evaluation capacity in population health: The NSW Health approach. *Health Promotions Journal of Australia*, 27(3):264-267. https://doi.org/10.1071/HE16045

IMF (International Monetary Fund). 2017. 2018 Quinquennial review of the fund's capacity development strategy: Concept note. Unpublished paper. Washington, DC.

Kotvojs, F. 2017. The capacity development evaluation framework: Tested on three initiatives. *Evaluation Journal of Australasia*,14(4):13-24. https://doi.org/10.1177/1035719X1701700403

Kotvojs, F. & Hurworth, R. 2013. Valuating capacity development: What do we really want to know? *Evaluation Journal of Australasia,* 3(1):4-14. https://doi.org/10.1177/1035719X1301300102

Kwang-Ho, S. 2009. Theoretical and strategic frameworks for national evaluation capacity building (NECB). *International Area Review*, 12(3):199-218. https://doi.org/10.1177/223386590901200312

Labin, S.N. 2014. Developing common measures in evaluation capacity building: An iterative science and practice process. *American Journal of Evaluation*, 35(1):107-115. https://doi.org/10.1177/1098214013499965

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

Lennie, J.R.; Tacchi, J.; Wilmore, M. & Koirala, B. 2015. A holistic, learning-centred approach to building evaluation capacity in development organizations. *Evaluation*, 21(3):325-343. https://doi.org/10.1177/1356389015590219

Leviton, L.C. 2013. Some underexamined aspects of evaluation capacity building. *American Journal of Evaluation,* 35(1):90-94. https://doi.org/10.1177/1098214013502844

Lucas, B. 2013. *Current thinking on capacity development.* GSDRC Helpdesk Research Report No. 960. Birmingham, UK: GSDRC, University of Birmingham. http://www.gsdrc.org/docs/open/hdq960.pdf [Accessed 7 May 2017].

Mapitsa Blaser, C. & Khumalo, L. 2018. Diagnosing monitoring and evaluation capacity in Africa. *African Evaluation Journal*, 6(1):1-10. https://doi.org/10.4102/aej.v6i1.255

Merkle, C. 2016. UN women's experience with strengthening evaluation systems in Africa: Enhancing quantity, quality and use of evaluations. *African Evaluation Journal*, 4(1):1-8. https://doi.org/10.4102/aej.v4i1.127

Morkel, C. & Ramasobana, M. 2017. Measuring the effect of evaluation capacity building initiatives in Africa: A review. *African Evaluation Journal*, 5(1). https://doi.org/10.4102/aej.v5i1.187

Mulenga, O. & Porter, S. 2013. Study on the demand for and supply of evaluation in Zambia. Centre for Learning on Evaluation and Results Anglophone Africa (CLEAR-AA). Johannesburg: University of the Witwatersrand.

Podems, D. 2014. Evaluator competencies and professionalizing the field: Where are we now? *Canadian Journal of Program Evaluation*, 28(3):127-136.

Podems, D. 2015. Road map to: Feasibility study on professionalisation of evaluation in South Africa or strengthening evaluators in South Africa. Unpublished consultancy report submitted to South African Monitoring and Evaluation Association (SAMEA) and Department of Planning, Monitoring and Evaluation (DPME). https://ucarecdn.com/0094d98b-46ba-43b2-aff7-3cbb0915050a/

Podems, D.; Goldman, I. & Jacob, C. 2014. Evaluator competencies: The South African government experience. *The Canadian Journal of Program Evaluation*, 28(3):71-85.

Porter, S. & Feinstein, O. 2014. Demand for and supply of evaluation in selected sub-Saharan African countries. https://www.gov.uk/dfid-research-outputs/demand-for-and-supply-of-evaluations-in-selected-sub-saharan-african-countries

Porter, S. & Goldman, I. 2013. A growing demand for monitoring and evaluation in Africa. *African Evaluation Journal*, 11(1):1-9. https://doi.org/10.4102/aej.v1i1.25

Preskill, H. 2008. Evaluation's second act: A spotlight on learning. *American Journal of Evaluation*, 29(2):127-138. https://doi.org/10.1177/1098214008316896

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

Smith, L. & Morkel, C. 2018. Trends in supply and demand for evaluation in Africa, a view from CLEAR-AA. *Evaluation Matters*. Abidjan: African Development Bank.

Stewart, R. 2015. A theory of change for capacity building for the use of research evidence by decision makers in southern Africa. *Evidence & Policy*, 11(4):547-557. https://doi.org/10.1332/174426414X1417545274793

Stockdill, S.H.; Baizerman, M. & Compton, D.W. 2002. Toward a definition of the ECB process: A conversation with the ECB literature. *New Directions For Evaluation*, 93:7-26. https://doi.org/10.1002/ev.39

Tarsilla, M. 2014. Evaluation capacity development in Africa: Current landscape of international partners' initiatives, lessons learned and the way forward. *African Evaluation Journal*, 2(1):1-13. https://doi.org/10.4102/aej.v2i1.89

Taut, S. 2007. Defining evaluation capacity building utility considerations. *American Journal of Evaluation*, 28(1):120. https://doi.org/10.1177/1098214006298062

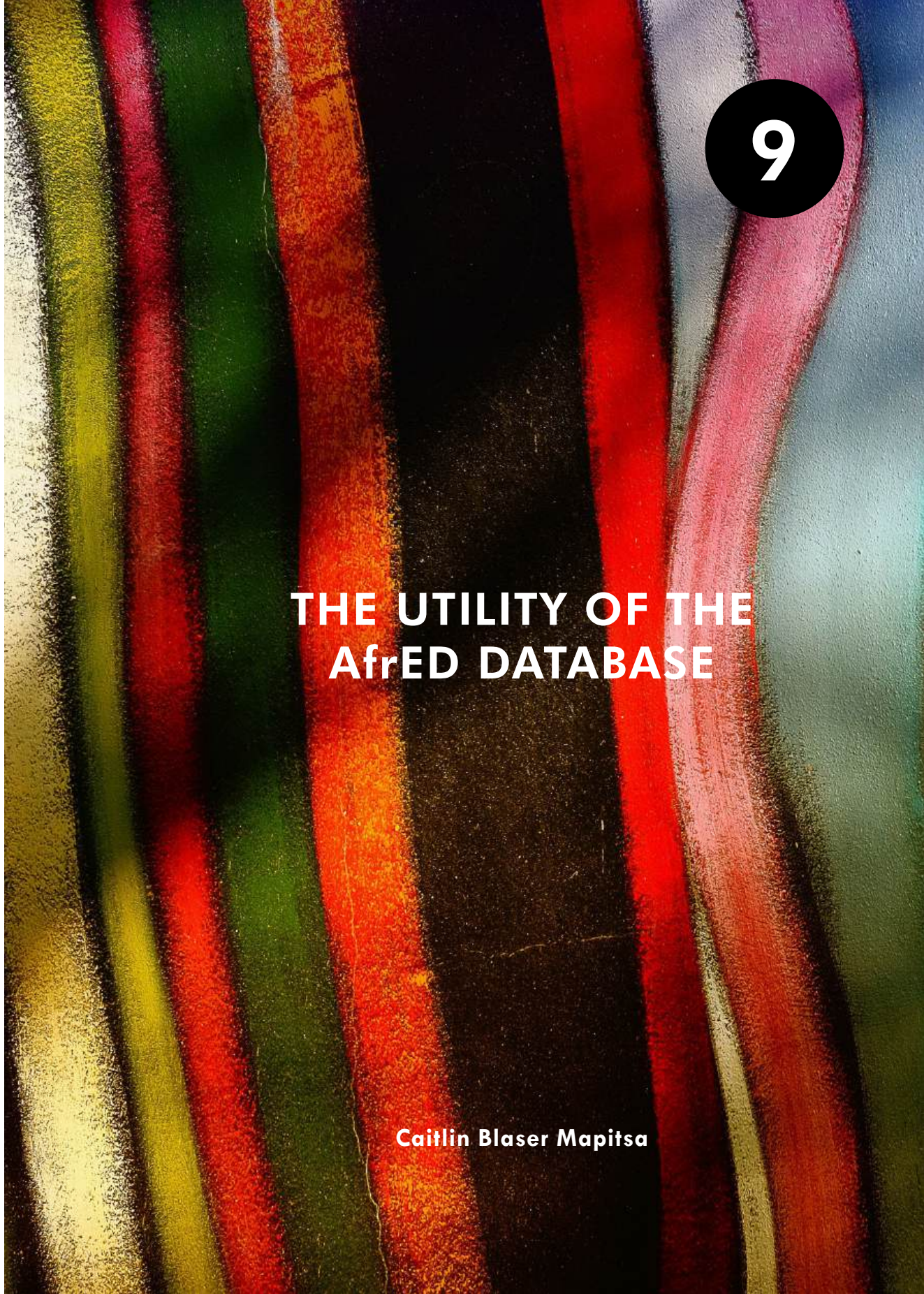
UNICEF. 2008. The role of monitoring and evaluation in evidence-based policy making bridging the gap. Evaluation Working Papers, Issue 12. Geneva: UNICEF Regional Office for Evaluation.

Volkov, B.B. & King, J.A. 2007. A checklist for building organisational evaluation capacity. http://dmeforpeace.org/sites/default/files/Volkov%20and%20King_Checklist%20for%20Building%20Organizational%20Evaluation%20Capacity.pdf [Accessed 8 February 2017].

Wao, H., Onyango, R., Kisio, E., Njatha, M. & Onyango, N.O. 2017. Strengthening capacity for monitoring and evaluation through short course training in Kenya, *African Evaluation Journal*, 5(1):1-9. https://doi.org/10.4102/aej.v5i1.192

Wilcox, Y. & King, J. 2014. A Professional Grounding and History of the Development and Formal Use of Evaluator Competencies. *The Canadian Journal of Program Evaluation,* 28(3):1-28.

Wubneh, M. 2003. Building Capacity in Africa: The impact of institutional, policy and resource factors. *African Development Review,* 15(2-3):165-198. https://doi.org/10.1111/j.1467-8268.2003.00070.x



9

# THE UTILITY OF THE AfrED DATABASE

Caitlin Blaser Mapitsa

## Introduction

The various chapters in this book have helped to better understand the trends and dynamics of evaluations on the continent. They show that while the evaluation field has been growing in Africa over the last two decades, there is a dearth of empirical research that will allow scholars to better understand the relationship between evaluation practice and the larger political economy that underpins the evaluation agenda in the region. This is in line with the literature which demonstrates that Africa's developmental context demands effective systems of evaluation to ensure good governance and effective programming (Basheka & Byamugisha, 2015; Porter & Goldman, 2013). At the same time, it means that the professional trajectory of the evaluation sector must grow in a way that is appropriate to the needs of the region (Tarsilla, 2014). Finally, this growth on both sides must be supported by rigorous evaluation, ample empirical data, and an evidence-informed approach to monitoring and evaluation systems development (Mouton, 2010).

The construction of the AfrED database, from which most chapters draw their analyses, is providing some initial steps in creating a foundation of information to analyse in the making of research and capacity building decisions. As it develops, it has the potential to propel the development of the sector in the right direction. Without this empirical basis to help us describe the sector, understanding the most appropriate methods to support its growth, and measuring the success of these methods, will remain an exercise in futility (Morkel & Ramasobana, 2017). This is a double challenge in the monitoring and evaluation field, where the ability to rigorously demonstrate results is central to the legitimacy of the profession.

## The need for empirics to underpin evaluation

Data systems are often considered the foundation of evaluation systems (Gorgens & Kusek, 2010). While the relationship between data and evaluation is debated, there is no question that access to data is weak in Africa. This weakness remains a blight on the evaluation landscape. As just one example sixty countries in the world have complete vital registration system in place, and not one of them is in Africa (Beguy, 2016). Only 32 countries, which account for 65% of the continent's population have had a census in the last decade (Devarajan, 2013). Given the developmental landscape, it may not be a surprise

that the Statistical Capacity Index average for African countries is lower than the average for low income countries. However, it is discouraging that this has remained unchanged in over a decade (World Bank, 2017). Given that the census is often the root of data for development planning, the impact of this capacity gap is huge.

The impact of weak data systems on the evaluation sector is doubled. It is easy to see that good data are required for evaluations to be rigorous, and that accessible, affordable data are required for evaluations to be implementable. At the same time, data *about* evaluations and evaluation systems are essential to understand the effectiveness of evaluation practice. Given the centrality of measuring effectiveness in evaluation discourse, this is critical for the legitimacy of the field. When data limit the rigorousness of evaluations produced, it creates a challenge to building the evaluation sector. However, when data also limit available information about the evaluation sector, it becomes that much harder to argue for the importance of evaluations in the region.

This double affect is even further amplified when considering that current global trends in the evaluation sector are promoting the use of meta-analysis and systematic review techniques, described by Vedung (2010) as the fourth wave of evaluation diffusion. The importance of strong systems of empirical data vary by wave, but it is an unfortunate confluence that the growth in African evaluation systems coincides with a global emphasis on experimental design, and heavily quantitative approaches which rely all the more on rigorous data systems.

## Challenges to building data systems for evaluation

Given the interlinkages spelled out above between data systems and evaluation systems, it follows that the constraints to robust data systems in Africa cannot be divorced from those faced by the evaluation sector.

One challenge links immediately to the availability of skills in the region to both build and manage data systems, as well as to analyse and evaluate their contents (Mouton, 2011). Not only are there skills shortages in Africa, there is no consensus on which skills are required for professional evaluators, as professionalisation of the evaluation sector remains emergent (Abrahams, 2015). In spite of these robust or unresolved debates, there is little question that there is a gap in appropriately skilled evaluation practitioners in Africa.

A second challenge is that Africa continues to experience growth in the digital divide, at a time when many of the fastest growing global industries are in digital science (Hilbert, 2016). The increase in real time, crowd sourced, and other "new-wave" data offers many promises for better understanding the transformative promises of the Sustainable Development Goals. However, there are many restrictions to African evaluators accessing, let alone creating, open data repositories. Many technological innovations in this space offer a promise of opportunities for the continent, such as increased file sharing, e-books, increased open-source academic publishing, and P2P file-sharing networks (Kitchin, 2014). However, with each innovation also come limitations. The availability of data may well require more ICT resources to access, more analytical resources to engage with, and more sophisticated legal support around intellectual property to publish on (Bezuidenhout, Leonelli, Kelly & Rappert, 2017). The rapid change in the way and scale on which data are generated holds significant promise for people and institutions with resources constraints to be able to use data for decision-making, although these innovations do not solve, but rather shift, the previous barriers.

## The contribution and limitations of the database

The AfrED has started providing an empirical base for many claims in the sector. However, this was a first iteration of the database, and even when it is more robust there are limitations to what it can contribute to the discussion.

The most immediate limitation of that database is around its construction. Many evaluation databases were not included in the initial construction. It was impossible for the database to be exhaustive, due to limitations of time and resources, and it is hoped that with new approaches, including greater collaboration with evaluation commissioners in the future, the next iteration of the database will be more exhaustive. Even without a completely exhaustive database, however, it was possible to start pulling out specific trends with robust information from certain agencies while acknowledging that there are gaps.

The AfrED database was built by the Centre for Research on Evaluation, Science and Technology (CREST), one of Africa's leading institutions for database development. The two collaborating organisations saw the need for the database to move forward the research agenda in the evaluation field. However, there was an initial challenge in defining the scope of the work, since the population of peer-reviewed articles on evaluations and evaluation reports themselves were not known.

The actual database itself is a modest first step in pulling together information from the evaluation sector. While it was possible to do a comprehensive review of academic peer-reviewed journal articles, this population was larger than expected. While this is a nice surprise, it makes the initiative larger in scale than expected, and all of the issues highlighted in the section above (ranging from resource scarcity to more tools to navigate intellectual property and a need for more information on the landscape) come all into play.

The initial idea was that a large percentage of the evaluation reports themselves could be included in the database, but the scope of work did not allow for this. Similarly, the survey of evaluation commissioners and practitioners was only able to include a small percentage of the relevant stakeholders. However, the emergent research and articulated needs have demonstrated that these two areas are critical for creating a foundation of information for future research in the sector.

While this database contributes to research on the African evaluation landscape, one data point will not solve systemic problems. The challenges facing African data systems, and those facing the evaluation sector, will require a systems-wide approach, resourcing, and a multi-stakeholder partnership to address in a more comprehensive way.

Government-led evaluations, which will be central to understanding how public sector evaluation systems in Africa are evolving, were not comprehensively included in the database. This is due to a range of challenges the public sector faces in making evaluations easily accessible by the public in a centralised database.

Even once the evaluations contained in the database are more comprehensive, there will still be limits to what empirical evidence can bring to the evaluation landscape. The diversity of views across the sector will remain and will likely result in contested interpretations of the data. If this can contribute to robust debate, the evaluation sector will no doubt be the richer because of it.

## Empirical contributions to evaluation research

The evaluation sector in Africa is evolving rapidly, and Africa's unique developmental context means that the profession and its trajectory of growth takes on its own shape. However, without a foundation of evidence that helps us

describe the sector, understanding the most appropriate methods to support its growth, and measuring their success, will remain an exercise in futility. The construction of the AfrED database has started providing this foundation; as it develops, it can help propel the development of the sector.

Some of the big-picture impressions and assumptions about the evaluation landscape are now being empirically validated, and this will help direct the evaluation research agenda moving forward. Here are a few highlights of some big questions and emerging answers.

### *Assertion 1 Evaluators for African projects are flown in from the West*

Evidence from the database is showing that this is generally true. While there are variations by commissioning agency, and acknowledging that identity and country of residence can be complex, lead evaluators on donor projects are generally not African. Chapters 2 and 3 discovered the scale and scope of the phenomenon and discussed some of the questions this raises for how we support an African evaluation sector. It then becomes an urgent priority of research to better understand the implications this has on the content of evaluations and the ways that they are used.

### *Assertion 2 When reference is made to 'evaluation', most people really mean monitoring*

This is also generally true. While evaluations are growing in number in the region, Chapter 7 has found that even when M&E systems conduct evaluations, these generally serve the purpose of monitoring by looking at progress against predetermined results, and rarely manage to support the function of strategic decision-making. Evaluations have the potential to provide evidence in response to some of the most pressing development challenges the region faces and could steer effective governance systems and development practice. However, if evaluation instead becomes another cog in the wheel of malicious bureaucratic compliance, it will fall far short of these aspirations.

### *Assertion 3 Since evaluation has not yet professionalised, evaluations are not of sufficient quality to be useful*

The answer here is mixed. Chapter 5 found that there are certainly evaluations that do not meet minimum quality standards. However, a larger number are

high quality, and the majority sit somewhere in between. Chapters 1 and 7 also discuss how different stakeholders may have different definitions of quality. Accommodating this diversity is important to consider as the sector evolves.

### *Assertion 4 Evaluation in Africa is growing*

This seems mostly true. Acknowledging that the final year of the database had a drop-off in submissions, most likely due to time lag, there have been a large and consistent increase in the number of peer-reviewed journal articles and evaluations themselves, even since 2005. This perceived growth of the sector is empirically demonstrable. There is reason to believe (and needs to be verified) that as public-sector evaluation systems grow, the evaluation sector will grow even more rapidly.

## New directions for research

This book is the start of a research agenda rather than the conclusion of one. Due to the birth of data in the evaluation sector, each data set asks more questions than it answers. This was certainly the experience of the contributing authors to this book.

On evaluation capacity and practice there is absolutely a need to better understand the profile of current evaluators and commissioners in the region. While we now have some information about the qualifications, experiences, and interests of evaluators, more work is needed to interrogate whether these strengths match the needs, and how evaluation practitioners themselves understand both the strengths and weaknesses of the evaluation practice. The issue of appropriate evaluation skills and capacity urgently requires more research to better define what should inform capacity-building approaches in Africa.

On methods and approaches, it is clear that evaluation reports do not always robustly reflect the methods and approaches used, In future the database can only be one source of information. However, it can start a conversation about the process of selecting methods for evaluations. It is clear that methods and approaches chosen often sit at the nexus of contestation between evaluator and commissioner. It is evident that a range of methods are being used in the region, but more work is needed to better understand what this means in terms of evaluation quality and use. Additionally, it is important to bring

this data back to questions of evaluation capacity and practice to see whether methodological questions and needs are driving capacity-building practice, or whether capacity (or other gaps) are limiting methodological approaches in the region.

Finally, on the issue of purpose and use, it is clear that stakeholders vary tremendously in the reasons for conducting evaluations – and in their ability to use the results. Better understanding this will help uncover how the construction of evaluation systems can drive reform in the region. Identifying the opportunities and gaps in this area will be critical to determining the extent to which evaluation can contribute to developmental outcomes in Africa.

## References


Abrahams, M.A. 2015. A review of the growth of monitoring and evaluation in South Africa: Monitoring and evaluation as a profession, an industry and a governance tool. *African Evaluation Journal*, 3(1):1-8. https://doi.org/10.4102/aej.v3i1.142

Basheka, B.C. & Byamugisha, A. 2015. The state of monitoring and evaluation (M&E) as a discipline in Africa. *African Journal of Public Affairs*, 8(3):75-95. https://doi.org/10.4102/apsdpr.v3i2.80

Beguy, D. 2016. Poor data affects Africa's ability to make the right policy decisions. *The Conversation*.18 August. https://theconversation.com/poor-data-affects-africas-ability-to-make-the-right-policy-decisions-64064 [Accessed 2 November 2016].

Bezuidenhout, L.M.; Leonelli, S.; Kelly, A.H. & Rappert, B., 2017. Beyond the digital divide: Towards a situated approach to open data. *Science and Public Policy*, 44(4):464-475. https://doi.org/10.1093/scipol/scw036

Devarajan, S. 2013. Africa's statistical tragedy. *Review of Income and Wealth*, 59:S9-S15. https://doi.org/10.1111/roiw.12013

Görgens, M. & Kusek, J.Z. 2010. *Making monitoring and evaluation systems work: A capacity development tool kit.* Washington, DC: The World Bank. https://doi.org/10.1596/978-0-8213-8186-1

Hilbert, M. 2016. Big data for development: A review of promises and challenges. *Development Policy Review*, 34(1): 135-174. https://doi.org/10.1111/dpr.12142

Kitchin, R. 2014. *The data revolution.* London: SAGE.

Morkel, C. & Ramasobama, M. 2017. Measuring the effect of evaluation capacity building initiatives in Africa: A review. *African Evaluation Journal*, 5(1):1-11. https://doi.org/10.4102/aej.v5i1.187

Mouton, C. 2010. The history of programme evaluation in South Africa. PhD thesis, Stellenbosch University.

Mouton, J. 2011. Doctoral production in South Africa: Statistics, challenges and responses. *Perspectives in Education*, 29(1):13-29.

Porter, S. & Goldman, I. 2013. A growing demand for monitoring and evaluation in Africa. *African Evaluation Journal*, 1(1):1-9. https://doi.org/10.4102/aej.v1i1.25

Tarsilla, M. 2014. Evaluation capacity development in Africa: Current landscape of international partners' initiatives, lessons learned and the way forward. *African Evaluation Journal*, 2(1):1-13. https://doi.org/10.4102/aej.v2i1.89

World Bank. 2017. *Bulletin board on statistical capacity*. Washington, DC: World Bank. http://datatopics.worldbank.org/statisticalcapacity