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 & Nilsson, 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 - Ethnographic techniques (focus group discussions, in-depth interviews, participant observations, case studies)
Summative/ Outcome	Impact, economic and synthesis evaluations	Quantitative - Statistical surveys - Experimental designs (randomised control trials) - 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 evaluation³

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 5.14⁴). 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 countries⁵

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-term 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