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

When you compare the extract from the CHES list to that of PEPFAR, it is evident that there is a significant overlap in the elements but there are also omissions. The CHES 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 Criterion 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 Criterion 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 ratio 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 (DFID³, USAID, the EU and GIZ⁴) 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 a set 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/apdpr.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](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>