6

APPRAISING QUALITATIVE EVALUATIONS

**Takunda Chirau, Precious Tirivanhu
& Mokgophana Ramasobana**

Introduction

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

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

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

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

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

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

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

Quality and evaluations – a background

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

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

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

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

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

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

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

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

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

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

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

The quality assessment framework

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

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

The relevance and application of the Quality Framework

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

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

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

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

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

Content of the framework

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

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

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

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

Methodological approach to the study

Sampling strategy

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

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

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

Figure 6.1 Multi-stage stratified sampling process

Development of assessment scores and the quality scoring process

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

Analysis of findings

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

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

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

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

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

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

Ethics

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

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

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

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

Design and methodological limitations

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

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

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

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

Sampling

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

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

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

Data collection

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

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

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

Reporting

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

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

Reflexivity and neutrality

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

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

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

Discussion

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

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

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

Figure 6.2 Quality rating of evaluation reports

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

Conclusion

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

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

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

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

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

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

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

References

- AIR (American Institutes for Research). 2015. *Performance Evaluation of Reading for Ethiopia's Achievement Developed Institutional Improvement (READ II)*. Washington, DC.
- ALNAP (Active Learning Network for Accountability and Performance in Humanitarian Action). 2006. *Evaluating humanitarian action using the OECD-DAC criteria: An ALNAP guide for humanitarian agencies*. London: Overseas Development Institute.
- Armiger, B. 1997. Ethics in nursing research: Profile, principles, perspective. *Nursing Research*, 26(5):330-333. <https://doi.org/10.1097/00006199-197709000-00003>
- Bamberger, M. 2004. Influential evaluations: Evaluations that improved performance and impacts of development programs. Operations Evaluation Department, the World Bank. Washington, DC: The World Bank.
- Bartlett, J.E.; Kotrlik, J.W. & Higgins, C.C. 2001. Organizational research: Determining appropriate sample size in survey research. *Information Technology, Learning, and Performance Journal*, 19(1):43.
- Bober, C.E. & Bartlett, K.R. 2004. The utilization of training program evaluation in corporate universities. *Human Resource Development Quarterly*, 15(4):363-383. <https://doi.org/10.1002/hrdq.1111>
- Bowbrick, P. 2015. The use of voice recorders in economic research. *Journal of Agricultural Economics*, 26(2):n.p. <https://doi.org/10.1111/j.1477-9552.1975.tb01466.x>
- Burns, N. & Grove, S.K. 2003. *The practice of nursing research: Conduct, critique and utilization*. Toronto: W.B. Saunders.
- Burry, J.; Alkin, M.C. & Ruskus, J. 1985. Organizing evaluations for use as a management tool. *Studies in Educational Evaluation*, 11(2):131-157. [https://doi.org/10.1016/0191-491X\(85\)90020-3](https://doi.org/10.1016/0191-491X(85)90020-3)
- Chianca, T.C. 2007. International development evaluation: An analysis and policy proposals. PhD thesis, Western Michigan University, Kalamazoo.
- Chianca, T. 2008. The OECD/DAC criteria for international development evaluations: An assessment and ideas for improvement. *Journal of MultiDisciplinary Evaluation*, 5(9):41-51.

- Chilisa, B. & Malunga, C. 2014. Made in Africa Evaluation: Uncovering African roots in evaluation theory and practice. In: F. Cloete, B. Rabie, & C. de Coning (eds). *Evaluation management in South Africa and Africa*. Stellenbosch: African Sun Media. <https://doi.org/10.18820/9781920689513>
- Connolly, T. & Porter, A.L. 1980. A user-focused model for the utilization of evaluation. *Evaluation and Program Planning*, 3(2):131-140. [https://doi.org/10.1016/0149-7189\(80\)90061-0](https://doi.org/10.1016/0149-7189(80)90061-0)
- Dassah, M.O. & Uken, E.A. 2007. Towards more effective utilisation of evaluations. *Journal of Public Administration*, 42(2):119-135.
- De Vos, A.A.; Strydom, H.; Fouché, C.B. & Delpont, C.S.L. 2011. *Research at Grassroots for Social Sciences and Human Service Professions*. Fourth edition. Pretoria: Van Schaik.
- DPME (Department of Planning, Monitoring and Evaluation). 2014. *Standards for evaluation in government*. Pretoria: Government Printing Works.
- Farrugia, P.; Petrisor, B.A.; Farrokhyar, F. & Bhandari, M. 2010. Research questions, hypotheses and objectives. *Canadian Journal of Surgery*, 53(4):278-281.
- Flick, U.; Von Kardorff, E. & Steinke, I. 2004. *A companion to qualitative research*. London: SAGE.
- Gilgun, J.F. 2008. Lived experience, reflexivity, and research on perpetrators of interpersonal violence. *Qualitative Social Work*, 7(2):181-197. <https://doi.org/10.1177/1473325008089629>
- Ginsberg, A. & Rhett, N. 2003. Building a better body of evidence: New opportunities to strengthen evaluation utilization. *American Journal of Evaluation*, 24(4):489-498. <https://doi.org/10.1016/j.ameval.2003.10.005>
- Guba, E.G. & Lincoln, Y.S. 1989. *Fourth generation evaluation*. Newbury Park: SAGE.
- Hinkin, T.R.; Tracey, J.B. & Enz, C.A. 1997. Scale construction: Developing reliable and valid measurement instruments. *Journal of Hospitality & Tourism Research*, 21(1):100-120. <https://doi.org/10.1177/109634809702100108>
- Johnson, K.; Greenseid, L.O.; Toal, S.A.; King, J.A.; Lawrenz, F. & Volkov, B. 2009. Research on evaluation use: A review of the empirical literature from 1986 to 2005. *American Journal of Evaluation*, 30(3):377-410. <https://doi.org/10.1177/1098214009341660>
- Johnston, W.P., Jr. 1986. A study of the acceptance of management performance evaluation recommendations by federal agencies: Lessons from GAO reports issued in FY 1983. Dissertation Abstracts International: Section A: The Humanities and Social Sciences, 48, 2157 (UMI 8725055).
- Kibombo, R.; Owarwo, V.; Nuamanya, A.; Jessani, N.; Otikal, K. & Mubangizi, P. 2016. Evaluation: USAID/Uganda Strengthening Decentralization for Sustainability (SDS) Program. Prepared for United States Agency for International Development. http://ghpro.dexonline.com/sites/default/files/136%20SDS_Eval_Final_Format%207.18.16_508.compressed.pdf
- Kumar, R. 2005. *Research Methodology – A Step-by-Step Guide for Beginners*. Second edition. Singapore: Pearson Education.
- Leedy, P.D. & Ormrod, J.E. 2005. *Practical Research: Planning and Design*. Seventh edition. New Jersey: Merrill Prentice Hall.
- Leslie, M.; Moodley, N.; Goldman, I.; Jacob, C.; Podems, D.; Everett, M. & Beney, T. 2015. Developing evaluation standards and assessing evaluation quality. *African Evaluation Journal* 3(1), Art. #112, 13 pages. <https://doi.org/10.4102/aej.v3i1.112>

- Lewandowsky, S.; Ecker, U.K.H. & Cook, J. 2017. Beyond misinformation: Understanding and coping with the "Post-Truth" era. *Journal of Applied Research in Memory and Cognition*, 6(4):353-369. <https://doi.org/10.1016/j.jarmac.2017.07.008>
- MDPMED (Ministry of Devolution and Planning, Monitoring and Evaluation Department). 2016. *Guidelines for the development of county integrated monitoring and evaluation systems*. Nairobi: Government of the Republic of Kenya.
- Meyrick, J. 2006. What is Good Qualitative Research? A First Step towards a Comprehensive Approach to Judging Rigour/Quality. *Journal of Health Psychology*, 11(5):799-808. <https://doi.org/10.1177/1359105306066643>
- Morris, M. 2008. *Evaluation ethics for best practices*. London: Guilford Press.
- Neuman, W.L. 1994. *Social research methods: Qualitative and quantitative approaches*. London: Allyn and Bacon.
- O'Brien, B.C.; Harris, I.B.; Beckman, T.J.; Reed, D.A. & Cook, D.A. 2014. Standards for reporting qualitative Research: A Synthesis of Recommendations. *Academic Medicine*, 89(9):1245-1250. <https://doi.org/10.1097/ACM.0000000000000388>
- OECD (Organisation for Economic Cooperation and Development). 2010. *DAC guidelines and reference series quality standards for development evaluation*. Paris: OECD Secretariat.
- Ofir, Z. 2017. *Updating the DAC evaluation Criteria, Part 2. Why change the status quo?* <http://zendaofir.com/updating-the-dac-evaluation-criteria-part-2/> [Accessed 12 January 2018].
- Patel, M. 2002. A meta-evaluation, or quality assessment, of the evaluations in this issue, based on the African Evaluation Guidelines: 2002. *Evaluation and Program Planning*, 25(4):329-332. [https://doi.org/10.1016/S0149-7189\(02\)00043-5](https://doi.org/10.1016/S0149-7189(02)00043-5)
- Patel, M. 2013. African evaluation guidelines. *African Evaluation Journal* 1(1), Art. #51. <https://doi.org/10.4102/aej.v1i1.51>
- Patton, M.Q. 2015. *Qualitative Research and Evaluation Methods: Integrating Theory and Practice*. Fourth edition. Thousand Oaks, CA: SAGE.
- Potts, S.A.K. 1998. Impact of mixed method designs on knowledge gain, credibility, and utility of program evaluation findings. Dissertation Abstracts International: Section A: The Humanities and Social Sciences, 59, 1942 (UMI 9837695).
- Probst, B. 2015. The eye regards itself: Benefits and challenges of reflexivity in qualitative social work research. *Social Work Research*, 39(1):37-48. <https://doi.org/10.1093/swr/svu028>
- Puhan, M.A.; Akl, E.A.; Bryant, D.; Xie F.; Apolone, G. & Ter Riet, G. 2012. Discussing study limitations in reports of biomedical studies – the need for more transparency. *Health and Quality of Life Outcomes*, 10:23. <https://doi.org/10.1186/1477-7525-10-23>
- Rockwell, S.K.; Dickey, E.C. & Jasa, P.J. 1990. The personal factor in evaluation use: A case study of a steering committee's use of a conservation tillage survey. *Evaluation and Program Planning*, 13:389-394. [https://doi.org/10.1016/0149-7189\(90\)90024-Q](https://doi.org/10.1016/0149-7189(90)90024-Q)
- Rog, D.J. & Bickman, L. 1984. The feedback research approach to evaluation: A method to increase utility. *Evaluation and Program Planning*, 7(2):169-175. [https://doi.org/10.1016/0149-7189\(84\)90042-9](https://doi.org/10.1016/0149-7189(84)90042-9)
- Roller, M.R. & Lavrakas, P.J. 2015. *Applied Qualitative Research Design: A Total Quality Framework Approach*. New York, NY: Guilford Press.
- Sandelowski, M. 1995. Sample size in qualitative research. *Research in Nursing & Health*, 18(2):179-183. <https://doi.org/10.1002/nur.4770180211>

- Shea, M.P. 1991. Program evaluation utilization in Canada and its relationship to evaluation process, evaluator and decision context variables. University of Windsor (Canada). Dissertation Abstracts International: Section B: The Sciences & Engineering, 53, 597.
- Shenton, A.K. 2004. Strategies for ensuring trustworthiness in qualitative research projects. *Education for Information*, 22(2):63-75. <https://doi.org/10.3233/EFI-2004-22201>
- Silverman, D. 2000. *Doing qualitative research: A practical handbook*. London: SAGE.
- Sobal, J. 2001. Sample extensiveness in qualitative nutrition education research. *Journal of Nutrition Education*, 33(4): 184-192. [https://doi.org/10.1016/S1499-4046\(06\)60030-4](https://doi.org/10.1016/S1499-4046(06)60030-4)
- Spencer, L.; Ritchie, J.; Lewis, J. & Dillon, L. 2003. *Quality in qualitative evaluation: A framework for assessing research evidence*. London: National Centre for Social Research, Government Chief Social Researcher's Office.
- Stuckey, H.L. 2014. The first step in data analysis: Transcribing and managing qualitative research data. *Journal of Social Health and Diabetes*, 2(1):6-8. <https://doi.org/10.4103/2321-0656.109828>
- Tong, A.; Sainsbury, P. & Craig, J. 2007. Consolidated criteria for reporting qualitative research (COREQ): A 32-item checklist for interviews and focus groups. *International Journal for Quality in Health Care*, 19(6):349-357. <https://doi.org/10.1093/intqhc/mzm042>
- UNEG (United Nations Evaluation Group). 2016. *Norms and standards for evaluation*. New York: UNEG.
- UNFPA (United Nations Population Fund). 2017. *Tools and guidance: Dimensions of evaluation quality at UNFPA*. New York: UNFPA Evaluation Office.
- Veal, A.J. 2011. *Research methods for leisure and tourism*. Fourth edition. London: Pearson Education.
- Walsh, D. & Downe, S. 2006. Appraising the quality of qualitative research. *Midwifery*, 22(2):108-119. <https://doi.org/10.1016/j.midw.2005.05.004>
- Weiss, C.H. 1998. *Evaluation: Methods for Studying Programs and Policies*. Second edition. Upper Saddle River, NJ: Prentice Hall.
- WFP (World Food Programme). 2015. *Evaluation Quality Assurance System (EQAS): Guidelines for operation evaluations*. Rome: WFP Office of Evaluation.
- White, C.; Woodfield, K. & Ritchie, J. 2003. Reporting and presenting qualitative data. In: J. Ritchie & J. Lewis (eds). *Qualitative Research Practice*. London: SAGE. 287-320.

ANNEXURE

Quality Framework			
	A. Appraisal questions	B. Quality indicators (possible features for consideration)	C. Notes on study being appraised
1. Findings	How credible are the findings?	<p>a. Findings/conclusions are supported by data/study evidence (i.e. the reader can see how the researcher arrived at his/her conclusions; the 'building blocks' of analysis and interpretation are evident).</p> <p>b. Findings/conclusions 'make sense/have a coherent logic'.</p> <p>c. Findings/conclusions are resonant with other knowledge and experience (this might include peer or member review).</p> <p>d. Use of corroborating evidence to support or refine findings (i.e. other data sources have been used to examine phenomena; other research evidence has been evaluated).</p>	
2. Findings	How has knowledge/understanding been extended by the research?	<p>a. Literature review (where appropriate) summarising knowledge to date/key issues raised by previous research.</p> <p>b. Aims and design of study set in the context of existing knowledge/understanding; identifies new areas for investigation (for example, in relation to policy/practice/substantive theory).</p> <p>c. Credible/clear discussion of how findings have contributed to knowledge and understanding (e.g. of the policy, programme or theory being reviewed); might be applied to new policy developments, practice or theory.</p> <p>d. Findings presented or conceptualised in a way that offers new insights/alternative ways of thinking.</p> <p>e. Discussion of limitations of evidence and what remains unknown/unclear or what further information/research is needed.</p>	

Quality Framework			
	A. Appraisal questions	B. Quality indicators (possible features for consideration)	C. Notes on study being appraised
3. Findings	How well does the evaluation address its original aims and purpose?	<ul style="list-style-type: none"> a. Clear statement of study aims and objectives; reasons for any changes in objectives. b. Findings clearly linked to the purposes of the study – and to the initiative or policy being studied. c. Summary or conclusions directed towards aims of study. d. Discussion of limitations of study in meeting aims (e.g. are there limitations because of restricted access to study settings or participants, gaps in the sample coverage, missed or unresolved areas of questioning, incomplete analysis; time constraints?). 	
4. Findings	Scope for drawing wider inference – how well is this explained?	<ul style="list-style-type: none"> a. Discussion of what can be generalised to wider population from which sample is drawn/case selection has been made. b. Detailed description of the contexts in which the study was conducted to allow applicability to other settings/contextual generalities to be assessed. c. Discussion of how hypotheses/propositions/findings may relate to wider theory; consideration of rival explanations. d. Evidence supplied to support claims for wider inference (either from study or from corroborating sources). e. Discussion of limitations on drawing wider inference (e.g. re-examination of sample and any missing constituencies: analysis of restrictions of study settings for drawing wider inference). 	
5. Findings	How clear is the basis of evaluative appraisal?	<ul style="list-style-type: none"> a. Discussion of how assessments of effectiveness/evaluative judgements have been reached (i.e. whose judgements are they and on what basis have they been reached?). b. Description of any formalised appraisal criteria used, when generated and how and by whom they have been applied. c. Discussion of the nature and source of any divergence in evaluative appraisals d. Discussion of any unintended consequences of intervention, their impact and why they arose. 	

Quality Framework			
	A. Appraisal questions	B. Quality indicators (possible features for consideration)	C. Notes on study being appraised
6. Design	How defensible is the research design?	<ul style="list-style-type: none"> a. Discussion of how overall research strategy was designed to meet aims of study b. Discussion of rationale for study design. c. Convincing argument for different features of research design (e.g. reason given for different components or stages of research; purpose of particular methods or data sources, multiple methods, time frames). d. Use of different features of design/data sources evident in findings presented. e. Discussion of limitations for study evidence. 	
7. Sample	How well defended is the sample design/target selection of cases/documents?	<ul style="list-style-type: none"> a. Discussion of how overall research strategy was designed to meet aims of study. b. Discussion of rationale for study design. c. Convincing argument for different features of research design (e.g. reasons given for different components or stages of research; purpose of particular methods or data sources, multiple methods, time frames etc.). d. Use of different features of design/data sources evident in findings presented. e. Discussion of limitations of research design and their implications for the study evidence. 	
8. Sample	Sample composition/case inclusion – how well is the eventual coverage described?	<ul style="list-style-type: none"> a. Detailed profile of achieved sample/case coverage. b. Maximising inclusion (e.g. language matching or translation; specialised recruitment; organised transport for group attendance). c. Discussion of any missing coverage in achieved samples/cases and implications for study evidence (e.g. through comparison of target and achieved samples, comparison with population etc.). d. Documentation of reasons for non-participation among sample approached/non-inclusion of selected cases/documents. e. Discussion of access and methods of approach and how these might have affected participation/coverage. 	

Quality Framework			
	A. Appraisal questions	B. Quality indicators (possible features for consideration)	C. Notes on study being appraised
9. Data Collection	How well was the data collection carried out?	<p>a. Discussion of:</p> <ul style="list-style-type: none"> • who conducted data collection; • procedures/documents used for collection/recording; • checks on origin/status/authorship of documents. <p>b. Audio or video recording of interviews/discussions/conversations (if not recorded, were justifiable reasons given?).</p> <p>c. Description of conventions for taking field notes (e.g. to identify what form of observations were required/to distinguish description from researcher commentary/analysis).</p> <p>d. Discussion of how fieldwork methods or settings may have influenced data collected.</p> <p>e. Demonstration, through portrayal and use of data, that depth, detail and richness were achieved in collection.</p>	
10. Analysis	How well has the approach to, and formulation of, the analysis been conveyed?	<p>a. Description of form of original data (e.g. use of verbatim transcripts, observation or interview notes, documents, etc.).</p> <p>b. Clear rationale for choice of data management method/tool/package.</p> <p>c. Evidence of how descriptive analytic categories, classes, labels etc. have been generated and used (i.e. either through explicit discussion or portrayal in the commentary).</p> <p>d. Discussion, with examples, of how any constructed analytic concepts/typologies etc. have been devised and applied.</p>	
11. Analysis	Contexts of data sources – how well are they retained and portrayed?	<p>a. Description of background or historical developments and social/organisational characteristics of study sites or settings.</p> <p>b. Participants' perspectives/observations placed in personal context (e.g. use of case studies/vignettes/individual profiles, textual extracts annotated with details of contributors).</p> <p>c. Explanation of origins/history of written documents.</p> <p>d. Use of data management methods that preserve context (i.e. facilitate within case description and analysis).</p>	

Quality Framework			
	A. Appraisal questions	B. Quality indicators (possible features for consideration)	C. Notes on study being appraised
12. Analysis	How well has diversity of perspective and content been explored?	<ul style="list-style-type: none"> a. Discussion of contribution of sample design/ case selection in generating diversity. b. Description and illumination of diversity/multiple perspectives/alternative positions in the evidence displayed. c. Evidence of attention to negative cases, outliers or exceptions. d. Typologies/models of variation derived and discussed. e. Examination of origins/influences on opposing or differing positions. f. Identification of patterns of association/linkages with divergent positions/groups. 	
13. Analysis	How well has detail, depth and complexity (i.e. richness) of the data been conveyed?	<ul style="list-style-type: none"> a. Use and exploration of contributors' terms, concepts and meanings. b. Unpacking and portrayal of nuance/subtlety/intricacy within data. c. Discussion of explicit and implicit explanations. d. Detection of underlying factors/influences. e. Identification and discussion of patterns of association/conceptual linkages within data. f. Presentation of illuminating textual extracts/observations. 	
14. Reporting	How clear are the links between data, interpretation and conclusions – i.e. how well can the route to any conclusions be seen?	<ul style="list-style-type: none"> a. Clear conceptual links between analytic commentary and presentations of original data (i.e. commentary and cited data relate; there is an analytic context to cited data, not simply repeated description). b. Discussion of how/why particular interpretation/significance is assigned to specific aspects of data – with illustrative extracts of original data. c. Discussion of how explanations/ theories/conclusions were derived – and how they relate to interpretations and content of original data (i.e. how warranted); whether alternative explanations explored. d. Display of negative cases and how they lie outside main proposition/theory/hypothesis etc.; or how proposition etc. revised to include them. 	

Quality Framework			
	A. Appraisal questions	B. Quality indicators (possible features for consideration)	C. Notes on study being appraised
15. Reporting	How clear and coherent is the reporting?	<ul style="list-style-type: none"> a. Demonstrates link to aims of study/research questions. b. Provides a narrative/story or clearly constructed thematic account. c. Has structure and signposting that usefully guide reader through the commentary. d. Provides accessible information for intended target audience(s). e. Key messages highlighted or summarised. 	
16. Neutrality & Reflexivity	How clear are the assumptions/theoretical perspectives/values that have shaped the form and output of the evaluation?	<ul style="list-style-type: none"> a. Discussion/evidence of the main assumptions/hypotheses/theoretical ideas on which the evaluation was based and how these affected the form, coverage or output of the evaluation (<i>the assumption here is that no research is undertaken without some underlying assumptions or theoretical ideas</i>). b. Discussion/evidence of the ideological perspectives/values/philosophies of research team and their impact on the methodological or substantive content of the evaluation (<i>again, may not be explicitly stated</i>). c. Evidence of openness to new/alternative ways of viewing subject/theories/assumptions (<i>e.g. discussion of learning/concepts/ constructions that have emerged from the data; refinement restatement of hypotheses/theories in light of emergent findings; evidence that alternative claims have been examined</i>). d. Discussion of how error or bias may have arisen in design/data collection/analysis and how addressed, if at all. e. Reflections on the impact of the researcher on the research process. 	

Quality Framework			
	A. Appraisal questions	B. Quality indicators (possible features for consideration)	C. Notes on study being appraised
17. Ethics	What evidence is there of attention to ethical issues?	<ul style="list-style-type: none"> a. Evidence of thoughtfulness/sensitivity about research contexts and participants. b. Documentation of how research was presented in study settings/to participants (<i>including, where relevant, any possible consequences of taking part</i>). c. Documentation of consent procedures and information provided to participants. d. Discussion of confidentiality of data and procedures for protecting. e. Discussion of how anonymity of participants/sources was protected. f. Discussion of any measures to offer information/advice/services etc. at end of study (<i>i.e. where participation exposed the need for these</i>). g. Discussion of potential harm or difficulty through participation, and how avoided. 	
18. Auditability	How adequately has the research process been documented?	<ul style="list-style-type: none"> a. Discussion of strengths and weaknesses of data sources and methods. b. Documentation of changes made to design and reasons; implications for study coverage. c. Documentation and reasons for changes in sample coverage/data collection/analytic approach; implications. d. Reproduction of main study documents (<i>e.g. letters of approach, topic guides, observation templates, data management frameworks etc.</i>). 	

(Source: Spencer et al., 2003:9-15)