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

Trustworthiness reporting in peer-reviewed qualitative evaluations in African contexts: Patterns and implications

Authors:
Dominic A. Okoliko[1]
Lauren P. Wildschut[1]

Affiliations:
[1]Centre for Research on Evaluation, Science and Technology (CREST), Faculty of Arts and Social Sciences, Stellenbosch University, Stellenbosch, South Africa

Corresponding author:
Dominic Okoliko,
okolikoda@sun.ac.za

Dates:
Received: 19 Nov. 2025
Accepted: 09 July 2026
Published: 20 Aug. 2026

How to cite this article:
Okoliko, D.A. & Wildschut, L.P., 2026, 'Trustworthiness reporting in peer-reviewed qualitative evaluations in African contexts: Patterns and implications', *African Evaluation Journal* 14(1), a884. https://doi.org/10.4102/aej.v14i1.884

Copyright:
© 2026. The Authors. Licensee: AOSIS. This work is licensed under the Creative Commons Attribution 4.0 International (CC BY 4.0) license (https://creativecommons.org/licenses/by/4.0/).

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

Background: Qualitative approaches are widely applied across disciplines, yet their role in programme evaluation in Africa, where evidence is critical for accountability and learning, remains underexplored.

Objectives: This study examines how trustworthiness strategies are reported in peer-reviewed qualitative evaluation studies in African contexts, and how reporting varies across methodological orientations, thematic focus areas and Computer-Assisted Qualitative Data Analysis Software (CAQDAS) use.

Method: We conducted a scoping review of 83 Web of Science (WoS)-indexed evaluation studies in African contexts, applying directed content analysis to map trustworthiness-reporting strategies and multiple correspondence analysis to examine variation.

Results: Findings reveal a consistent pattern: the WoS-indexed evaluation studies are most transparent about structurally visible methodological elements, notably thick description and data collection, while trustworthiness strategies that demonstrate deeper analytic or relational accountability are systematically under-reported. The corpus shows that strategies requiring greater methodological accountability, such as analytic details, reflexivity and sampling, are rarely comprehensive. Participant engagement strategies (e.g. prolonged engagement) and analytic mechanisms (e.g. audit trails) are particularly underused. Multiple correspondence analysis identified three clusters: (1) comprehensive reporting in health and standalone qualitative studies, (2) partial engagement strategies in CAQDAS-supported studies and (3) limited reporting in agriculture and environment evaluations.

Conclusion: This study contributes empirical grounding to qualitative evaluation practice in Africa, showing that reporting is structured by sectoral context and technological adoption more than by methodological orientation, with strategies critical to interpretive accountability remaining under-reported.

Contribution: Addressing the gaps requires capacity building that engages evaluators, journals, professional bodies and funders with the epistemological commitments of qualitative inquiry, not merely its reporting conventions.

Keywords: qualitative evaluation; trustworthiness; methodological transparency; CAQDAS; Africa.

Introduction

In an era of rising scrutiny of scientific inquiry, the credibility of research processes used in evaluation studies is more critical than ever. Evaluation relies on systematic inquiry to produce evidence for decision-making (Patton 2018). Quantitative and qualitative approaches reflect distinct epistemological traditions: quantitative inquiry, rooted in positivism, upholds rigour through objectivity, replicability and statistical verification, whereas qualitative inquiry, grounded in interpretivism, treats meaning as socially constructed and the researcher as an instrument, rendering subjectivity inevitable and making transparency and reflexivity central to establishing rigour (Guba, Lincoln & Lynham 2018; Lincoln & Guba 1985). While quantitative approaches benefit from established principles of rigour, qualitative approaches often face concerns about trustworthiness (Curtin & Fossey 2007; Lub 2015), especially in evaluations where judgements about merit and value are central (Patton 2018).

Note: Additional supporting information may be found in the online version of this article as Online Appendix 1 and Online Appendix 2.

Three concepts are central to this study. *Trustworthiness* refers to the extent to which a qualitative inquiry is perceived as credible, dependable and authentic – the overall confidence placed in its findings and processes (Curtin & Fossey 2007; Lincoln & Guba 1985). Trustworthiness is enacted through deliberate practices that address two interrelated dimensions: *rigour*, which refers to the systematic and methodical conduct of inquiry, including decisions about design, sampling, data collection and analysis; and *transparency*, which refers to the clarity with which these processes are reported, enabling scrutiny by readers and stakeholders (Kapiszewski & Karcher 2021; Lub 2015; Morse 2015). In evaluation, where findings inform judgements about merit and value, transparent reporting of these practices is not merely a methodological concern but a condition for credibility and use.

These concerns reflect long-standing debates about research quality in social sciences (Hammersley 2007; Kapiszewski & Karcher 2021) and evaluation (Anastas 2004; Spencer et al. 2003). In Africa, rising demand for evaluation (Porter & Goldman 2013; Wildschut & Silubonde 2020) has spurred professionalisation and efforts to enhance evaluation credibility. Regional evaluation principles (AfrEA 2021) and country-specific standards, such as South Africa's six-dimension framework (Leslie et al. 2015), emphasise methodological robustness to ensure credible results (Sibanda, Sibanda & Sibanda 2023). However, how qualitative evaluation reports communicate strategies for ensuring trustworthiness remains underexplored.

Effective reporting is central to trustworthiness demonstration (Lub 2015). As Leslie et al. (2015) note, evaluation reports should detail evaluation questions, intervention logic, methods and validity or reliability considerations. While qualitative traditions vary, there is consensus on core principles for demonstrating trustworthiness, including detailed methods, reflexivity and systematic procedures for ensuring methodological integrity (Creswell & Miller 2000; Spencer et al. 2003). However, empirical research on how these core principles are reported in African evaluation studies is scarce. Existing empirical reviews tend to be sector-specific, such as health (Raskind et al. 2019) and education (Liao & Hitchcock 2018). In Africa, most studies analysed grey literature rather than peer-reviewed publications (Ndhlovu, Khumalo & Krakra 2019; Ngwabi & Wildschut 2019). This study extends prior works by examining trustworthiness reporting in 83 peer-reviewed evaluation articles on Africa which used qualitative or mixed methods.

Academic literature, through peer review, encourages trustworthiness reporting (McWilliam 2000; Novak & Jen 2024), which elevates quality standards often lacking depth in non-academic evaluation reports (Ndhlovu et al. 2019; Ngwabi & Wildschut 2019). This study maps reported trustworthiness strategies in 83 Web of Science (WoS)-indexed evaluation articles in African contexts, analysing their prevalence using content and multiple correspondence analysis (MCA). The study responds to calls for research on evaluation to improve practice (Christie 2011). As Christie (2011) noted, rigorous, systematic evidence to guide evaluation and support evaluator reflection remains scarce. This study offers insights to improve the reporting of qualitative evaluations and support ongoing professionalisation and capacity building in African evaluation.

Evaluation, qualitative inquiry and trustworthiness in Africa

Programme evaluation has followed different trajectories across regions. In Africa, its expansion has been driven by non-governmental organisations and the private sector (Basheka 2015), though governments increasingly recognise its value (Abrahams 2019). While the practice grows in relevance across sectors, evaluation is often conflated with monitoring, and capacity gaps persist in national monitoring and evaluation (M&E) systems (Abrahams 2015, 2019). Non-governmental organisations and civil society implement numerous projects but often struggle to meet stakeholders' expectations for rigorous evaluation (Chaplowe & Engo-Tega 2007). Many practitioners also lack formal training in evaluation studies (Agonnoude et al. 2021). Although academic and non-academic training initiatives are growing, offerings remain limited, with only 13 of 26 South African universities offering such programmes (Mouton, Wildschut & Leslie 2018). Regional voluntary organisations for professional evaluation (VOPEs) continue to play a key role in fostering evaluation capacity.

Ensuring evaluation integrity is a global and continental priority (Agonnoude et al. 2021). Professional guidelines emphasise ethical and technical standards. The American Evaluation Association highlights systematic inquiry and competence (Rallis, Rossman & Gajda 2007). Similarly, the African Evaluation Association's (AfrEA) Evaluation Principles (2021) promote evaluations that are trustworthy, ethical and contextually grounded. Principle 2 underscores the need for rigour, transparency, reflexivity and feasibility, while Principle 3 stresses ethics and equity (AfrEA 2021). These guidelines align with debates on quality standards for qualitative approaches, increasingly relevant in evaluation. Qualitative methods provide contextual insights into how and why interventions work (Spencer et al. 2003). In evaluation, quantitative approaches focus on measuring outcomes and impacts, while qualitative approaches offer interpretive insight into these processes.

Lincoln and Guba (1985) introduced 'trustworthiness' as the qualitative counterpart to reliability and validity in the quantitative tradition. They proposed four criteria: credibility, transferability, dependability and confirmability – each linked to specific techniques. Since Lincoln and Guba's (1985) work, perspectives on qualitative rigour have evolved. While some advocate for applying traditional criteria to qualitative work (Creswell & Miller 2000), others argue for distinct standards aligned with qualitative epistemologies (Curtin & Fossey 2007; Novak & Jen 2024). Consequently, multiple frameworks exist, though consensus remains on the need for rigorous reporting (Hammersley 2007).

These debates are not merely academic: in the African evaluation context, where the field is still professionalising, practitioner training is uneven, and evaluation findings directly inform resource allocation and development decisions (Agonnoude et al. 2021; Tarsilla 2014), how trustworthiness is established and communicated carries particular practical weight. Evaluators operating within complex accountability environments – including donor-driven pressures, stakeholder expectations and internal learning demands (Tirivanhu 2022) – often have limited institutional support for methodological reflection (Agonnoude et al. 2021). Clear standards for demonstrating rigour are therefore necessary, yet the proliferation of competing frameworks (Dixon-Woods et al. 2004) risks obscuring rather than guiding practice. Continental frameworks such as AfrEA's Evaluation Principles (2021), which explicitly foreground rigour, transparency and reflexivity as constitutive of evaluation quality, signal that the African evaluation community recognises this need. The question this study addresses is whether that recognition is reflected in how qualitative evaluations are reported.

Discussions on the quality and trustworthiness of qualitative inquiry have influenced fields including education (Liao & Hitchcock 2018), social work (Enworo 2023), health sciences (Santiago-Delefosse et al. 2016) and programme evaluation (Spencer et al. 2003). Still, empirical work assessing how trustworthiness is demonstrated in practice remains limited. Liao and Hitchcock (2018) analysed 118 higher education evaluations, identifying commonly used trustworthiness techniques and offering a reporting framework that informs this study. We focus specifically on peer-reviewed articles reporting on qualitative components of evaluations in African contexts.

Three studies provide a foundation for understanding evaluation reporting in Africa. Ndhlovu et al. (2019) analysed 338 reports in the African Evaluation Database (AfrED) (2005–2016), finding high use of qualitative (55.60%) and mixed methods (35.80%) but limited methodological detail. Ngwabi and Wildschut (2019) reviewed 142 donor-commissioned reports, finding low adherence to Organisation for Economic Co-operation and Development (OECD) standards, especially in sampling and methodology. They also highlighted the underuse of African evaluators. Chirau, Tirivanhu and Ramasobana (2019) examined 20 qualitative reports using an 18-item checklist. Most were rated medium quality, with major gaps in reflexivity, ethics and methodological alignment.

Together, these studies highlight persistent gaps in evaluation reporting. However, they focus on grey literature. We argue that peer-reviewed publications, guided by journal standards and a review process, offer stronger insights into reporting practices, where consistent rigour is expected (McWilliam 2000; Novak & Jen 2024). Additionally, evaluators rely on academic research to guide practice: a survey of American Evaluation Association members found 80.95% regularly read published research (Coryn et al. 2016). This study examines how African evaluation authors respond to reporting expectations, offering lessons for improving practice.

To guide the inquiry, we asked:

- Which strategies for ensuring trustworthiness in qualitative evaluation processes are most frequently reported in African studies?
- How detailed and comprehensive are the descriptions of these strategies?
- How do these strategies vary by methodology (qualitative-only vs. mixed methods), focus area (e.g. health, education, agriculture) and Computer-Assisted Qualitative Data Analysis Software (CAQDAS) use (software-supported vs. non-software-supported studies)?

Research design and methods

This study adopts a scoping review design to examine how peer-reviewed qualitative evaluation studies conducted in African contexts report trustworthiness practices. Scoping reviews are suited to mapping evidence, identifying gaps and synthesising methodological trends (Peters et al. 2020; Pollock et al. 2022). This approach aligns with our objective of examining how trustworthiness strategies are applied and communicated, rather than evaluating individual studies. The review was guided by a Population–Concept–Context (PCC) framework (Peters et al. 2020). The population comprised peer-reviewed evaluation studies employing qualitative or mixed methods; the concept focused on trustworthiness, encompassing the rigour of inquiry and the transparency of its reporting; and the context was studies conducted in African settings.

This framework informed the formulation of the research questions and the development of the search strategy. Consistent with Lub (2015), who argues that research reports are the most plausible site for evaluating methodological processes, our analysis focused on reported evidence of trustworthiness. This aligns with Farrington's (2003) notion of descriptive validity, emphasising reporting adequacy. Accordingly, the study examines how published evaluation articles communicate trustworthiness strategies, rather than assessing underlying methodological practice. While rigour and transparency are conceptually distinct, we recognise that they are empirically inseparable at the level of reporting: our analysis assessed communicative transparency – the depth and clarity with which strategies were documented – from which the rigour of the underlying inquiry is inferred, as what evaluators make visible in published texts is the most plausible available evidence of how systematically their inquiry was conducted (Farrington 2003; Lub 2015).

Search strategy

Guided by the PCC framework, we searched the WoS database in July 2022 using the query:

TS = (qualitative research AND program* AND evaluation) AND CU = (Africa).

The search terms were derived from the core components of the review: qualitative evaluation approaches (Population), trustworthiness and methodological practices (Concept) and African study contexts (Context). Web of Science was selected for its extensive coverage of peer-reviewed literature, structured indexing and suitability for systematic searching. We note that in WoS, the CU field restricts searches to the country or region. CU = 'Africa' was therefore used as a broad retrieval parameter to capture studies indexed as relating to African contexts. While this approach supports exploratory coverage, it may not fully capture all country-specific records and therefore represents a limitation of the search strategy.

Consistent with the exploratory nature of scoping reviews, the search aimed to identify a broad yet manageable corpus rather than achieve exhaustive coverage. We note that the resulting dataset represents WoS-indexed, peer-reviewed evaluation studies conducted in African contexts, rather than the full universe of African evaluation practice. We acknowledge that additional sources, including the AfrED, may yield further studies. However, our focus on WoS-indexed publications reflects the study's exploratory objective of examining reporting practices within academically published evaluation research, complementing prior AfrED-based and institutional evaluation studies (Chirau et al. 2019; Ndhlovu et al. 2019; Ngwabi & Wildschut 2019).

The search yielded 189 studies. After applying inclusion criteria (evaluation studies using qualitative methods as stand-alone or mixed and focused on Africa), 83 publications were retained. We excluded duplicates, studies not based on African contexts, non-empirical papers and those lacking qualitative data. While the search does not claim exhaustive coverage, the systematically derived sample is sufficient to support the study's objective of mapping trustworthiness reporting practices. All included studies are listed in Online Appendix 1: Table 1. The review was conducted in line with established guidance for scoping reviews (Peters et al. 2020), with methodological adaptations appropriate to its exploratory scope and analytical focus. A pre-registered protocol was not developed and full compliance with reporting standards such as Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews (PRISMA-ScR) is not claimed; rather, scoping review principles were selectively applied to suit the study's objective.

Analytical framework

To analyse how trustworthiness is reported in qualitative evaluation studies, we developed an analytical framework based on established literature on qualitative rigour and trustworthiness (e.g. Morse 2015; Spencer et al. 2003). The framework draws primarily on Liao and Hitchcock (2018), who identify key techniques used to establish trustworthiness in qualitative evaluation. As summarised in Table 1, these sources collectively provide a comprehensive foundation for identifying and categorising trustworthiness strategies.

TABLE 1: Trustworthiness appraisal framework.

Primary design techniques	Confirmatory citation
Design: Describe and justify the study design and approaches, ensuring their relevance to the research questions.	Morse 2015; Spencer et al. 2003
Sampling: Describe and justify the sampling strategy, size and procedures. Provide context about the study setting and participants.	Spencer et al. 2003; Walsh & Downe 2006
Data collection: Describe the methods and procedures employed for data collection.	Spencer et al. 2003
Analytic details: Describe data analysis methods, including coding approaches, progression and reliability checks. Demonstrate appropriateness of the methods.	Anastas 2004; Tong, Sainsbury & Craig 2007
Thick description: Provide detailed descriptions of findings and use quotes to present real data supporting transferability.	Creswell & Miller 2000; Tong et al. 2007
Reflexivity: Demonstrate researcher awareness of their influence on the research, including efforts to control personal biases.	Walsh & Downe 2006; Tong et al. 2007
Limitations and delimitations: Acknowledge the study's limitations and delimitations.	Tong et al. 2007
***Ethical considerations**: Address ethical considerations related to the study.	AfrEA 2021; Anastas 2004
Additional trustworthiness techniques	
Triangulation: Use multiple data sources, methods, theories or investigators to converge evidence.	Lub 2015; Tong et al. 2007
Audit trail: Systematically document all procedures and data relevant to the study and indicate availability.	Anastas 2004; Lub 2015
Expert checking: Report the involvement of external experts in assessing study quality.	Tong et al. 2007
Member checking: Share data and findings with participants to obtain feedback on data, interpretations or conclusions.	Walsh & Downe 2006; Lub 2015
Peer debriefing: Engage professional colleagues in analytic discussions to enhance interpretation quality.	Creswell & Miller 2000
Negative case analysis: Identify and discuss disconfirming evidence to refine themes.	Creswell & Miller 2000; Lub 2015
Prolonged engagement: Spend significant time in the field to deepen understanding of the context.	Morse 2015; Lub 2015
Persistent observation: Conduct in-depth investigations with participants for accurate understanding.	Morse 2015; Tong et al. 2007
Exploring rival explanations: Rule out alternative explanations to enhance study robustness.	Tong et al. 2007
Pattern match: Compare themes across and within cases for consistency.	Liao & Hitchcock 2018
Weighting the evidence: Prioritise stronger data during analysis.	Liao & Hitchcock 2018
Comparison and contrast: Use comparison and contrast to enrich data interpretation.	Walsh & Downe 2006
Multivocality: Incorporate multiple perspectives to enhance understanding.	Spencer et al. 2003
***Inter-coder agreement**: Report on inter-coder reliability or consensus discussions when multiple coders are involved.	Morse 2015
***Code system developed**: Describe the development and use of a coding system to enhance internal validity.	Tong et al. 2007
***Other techniques**: Indicated additional strategies outside this list (e.g. pilot study).	-

Source: Adapted from Liao, H. & Hitchcock, J., 2018, 'Reported credibility techniques in higher education evaluation studies that use qualitative methods: A research synthesis', *Evaluation and Program Planning* 68, 157–165. https://doi.org/10.1016/j.evalprogplan.2018.03.005

Note: *, represents categories that emerged from our analysis, with corresponding recognition within the literature. The trustworthiness strategies were operationalised through a structured coding scheme with defined indicators and categorical coding of reporting. Full definitions are provided in Online Appendix 1: Table 2, Table 3 and Table 4.

As originally presented in Liao and Hitchcock (2018), trustworthiness strategies were organised into two categories: primary design techniques and additional trustworthiness techniques. The former represents techniques that are fundamental to the design phase and corresponding to the structurally visible elements of research reports, including research design, sampling, data collection, data analysis and

thick description. The latter techniques comprise supplementary strategies whose application varies depending on the nature of the research, including triangulation, audit trails, member checking, peer debriefing, negative case analysis and others (see Table 1), and which often require deeper analytic or relational accountability.

This approach shifts focus from classic categories such as credibility, transferability, dependability and confirmability (Lincoln & Guba 1985; Novak & Jen 2024), and associated debates about their parallels with quantitative quality criteria (Morse 2015), to observable practices through which trustworthiness is enacted and reported in published studies (Curtin & Fossey 2007). For analytical purposes, all strategies were treated as conceptually relevant but not hierarchically weighted, as the study aims to map reporting patterns rather than evaluate the relative importance of individual techniques.

Our adaptation further incorporated additional techniques that emerged during the coding process, extending Liao and Hitchcock's (2018) framework to reflect how trustworthiness is operationalised in the reviewed studies. These include ethical considerations under primary design techniques and inter-coder agreement and coding system description under additional trustworthiness techniques. As Lub (2015) and AfrEA (2021) emphasise, ethical considerations are crucial given the relational dynamics. Similarly, when multiple investigators are involved, inter-coder agreement and clear coding system descriptions support research rigour.

Additionally, we recorded descriptive characteristics for each publication: year of publication, author's country of affiliation, evaluation site (country), thematic focus area (e.g. health, education, agriculture), qualitative data type (e.g. interviews, observations) and use of CAQDAS to characterise methodological breadth.

Three document variables – methods, focus area and CAQDAS use – were used to explore variations in trustworthiness reporting through MCA, a method suited to revealing patterns in categorical data (Kienstra & Van der Heijden 2015). Multiple correspondence analysis enables the exploration of multiple categorical associations simultaneously, making it appropriate for our purposes (details provided in a later section). Computer-Assisted Qualitative Data Analysis Software use has received particular attention due to its relevance in supporting analytic rigour and transparency (Kapiszewski & Karcher 2021). Based on experience using and teaching ATLAS.ti, we expected CAQDAS-supported studies to show stronger trustworthiness reporting.

Coding and analysis

Figure 1 outlines our analytic workflow for the 83 publications using ATLAS.ti. Following the directed qualitative content analysis method (Hsieh & Shannon 2005), we applied a structured framework developed from the literature,

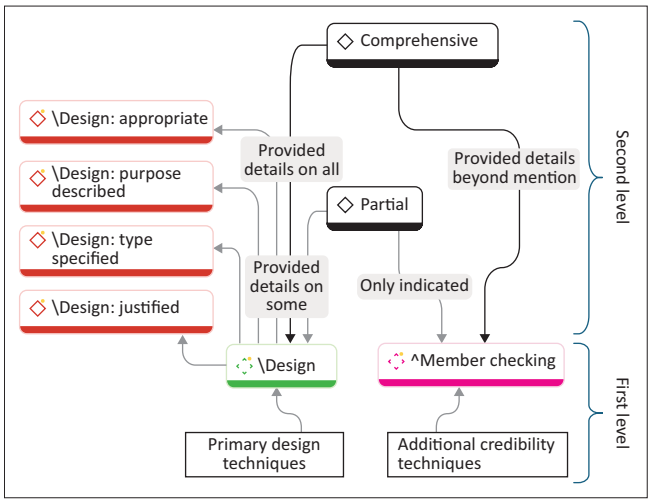


FIGURE 1: Analytic schemata for the study.

complemented by inductive refinement and both latent (interpretive) and manifest (descriptive) coding. Directed content analysis enables systematic classification of textual data to identify patterns. Our approach blended predefined concepts with data-driven interpretation, allowing iterative refinement of the initial coding framework to incorporate emergent categories (e.g. ethics, inter-coder agreement).

We used a two-level coding strategy:

- **Level 1**: Binary coding to indicate presence or absence of each technique.
- **Level 2**: Detailed coding to capture how each technique was implemented, the category code and sub-code functions in ATLAS.ti, enabling the determination of whether reporting was comprehensive or partial.

For instance, a study's research design was coded as present if it included any of the following: description, justification, alignment with purpose or relevance to research questions. A subsequent step involved identifying and sub-coding specific elements (e.g. design type, justification, stated purpose, and alignment with research questions). This detailed breakdown was applied across all techniques where granularity was necessary. The second-level analysis determined whether a technique was comprehensively or partially reported, based on the sub-codes (see the sub-code indicators in Online Appendix 1: Table 2, Table 3 and Table 4) illustrated above. Comprehensive reporting was recorded when all sub-category indicators were present for a strategy and partial when one or more elements were missing. Final classification was conducted in Excel, using ATLAS.ti's Code-Document Table outputs. The coding structure is fully documented in the accompanying Online Appendix 1.

Results visualisation and presentation

Our results include descriptive statistics and MCA. Firstly, we present frequencies and percentages summarising the use of trustworthiness techniques across documents. Secondly, we used MCA to explore how strategies varied by methodology, thematic focus and CAQDAS use.

Two key functions informed the MCA analysis: the eigenvalue, which indicates the variance (inertia) explained by each dimension, and the squared cosine (Cos^2), which assesses the quality of a category's representation in a dimension (Kienstra & Van der Heijden 2015). Eigenvalues help identify the most influential dimensions. Dimensions 1 and 2, explaining 8.86% and 7.33% of the variance, respectively (16.19% total), were retained due to their interpretive strength (see Figure 2). While these dimensions explained a modest proportion of variance, they were prioritised due to their interpretive relevance, consistent with exploratory analysis where low-dimensional solutions capture meaningful patterns (Husson, Le & Pagès 2017). A Cos^2 of 0.1 threshold was used to exclude weakly represented categories and minimise overinterpretation, consistent with exploratory multivariate analysis practices that retain interpretively meaningful associations (Husson et al. 2017).

Multiple correspondence analysis was conducted in R using *Factoshiny* and *FactoMineR*. The dataset comprised 26 variables (three document-level, 23 trustworthiness strategies) across 83 studies. Document features were inputted as supplementary variables, and trustworthiness strategies as active. Results are presented as a biplot and summary table, with full statistical results available in the accompanying Online Appendix 2 (Excel file).

Trustworthiness strategies for the study

To strengthen trustworthiness, we piloted the review using the full dataset and presented initial findings at the World Conference on Qualitative Research (January 2024). Feedback informed the shift to a full-text review. Author 1 conducted all coding and analysis, while Author 2 independently reviewed coded segments, coding decisions and category definitions. Disagreements were discussed and resolved through iterative comparison and refinement of the coding framework, supported by analytic memos and an audit trail documenting coding decisions. Given our interpretive approach, formal inter-coder agreement metrics were not applied; instead, consistency was enhanced through systematic verification and transparent documentation.

We also acknowledge the researchers' positionality in shaping the analysis. Both authors have experience in qualitative methods and CAQDAS (including ATLAS.ti), which informed initial expectations regarding analytic rigour and its transparent reporting. These assumptions were critically reflected upon during the analysis, particularly where findings did not align with expectations.

The accompanying Online Appendix 1 provides a detailed account of the dataset and analytic procedures, including the bibliography of the 83 studies and the coding structure, while Online Appendix 2 (Excel file) provides the structured dataset exported from ATLAS.ti and the MCA outputs (see Online Appendix 1: Table 5, for brief descriptions of the nine spreadsheets). We also maintained an audit trail, provided thick description in this report and conducted peer debriefing (early findings were shared with colleagues for feedback), reflecting established trustworthiness strategies (Curtin & Fossey 2007; Novak & Jen 2024).

Ethical considerations

This study relied solely on published literature and did not involve human participants or animals. An ethics-exempt approval letter was obtained from Stellenbosch University's Research Ethics Committee: Social Behavioural and Education Research (REC: SBE). The study was conducted in accordance with recognised ethical standards for research, including the principles outlined in the Singapore Statement on Research Integrity (2010).

Results

Our exploration of trustworthiness reporting across the 83 publications reveals a consistent pattern: these WoS-indexed evaluation studies in African contexts are most transparent about structurally visible methodological elements, while strategies requiring deeper analytic or relational accountability are systematically under-reported. Below, we present descriptive results before examining how these strategies are reflected in reporting.

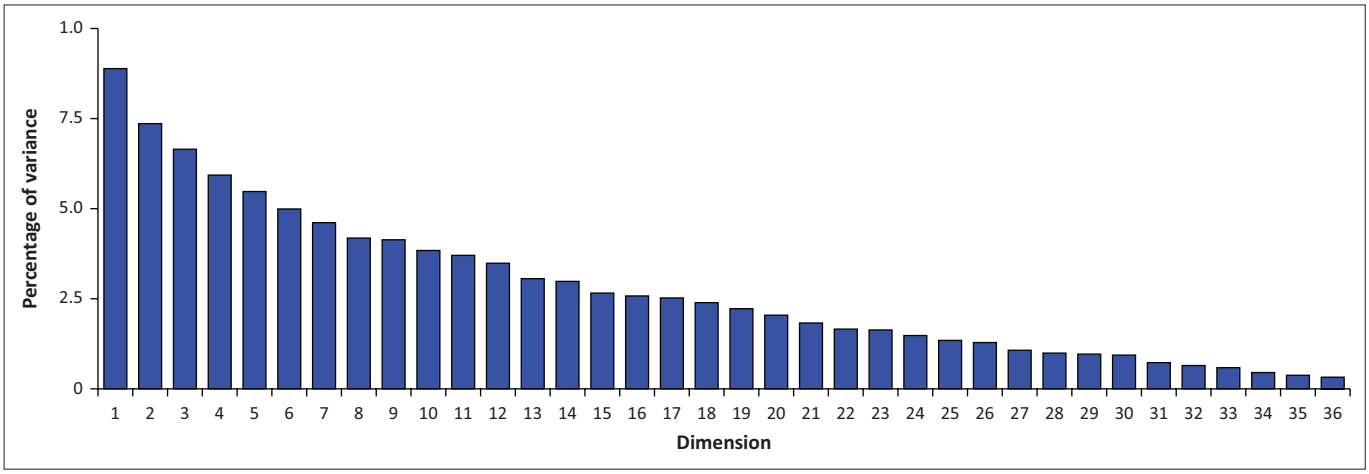
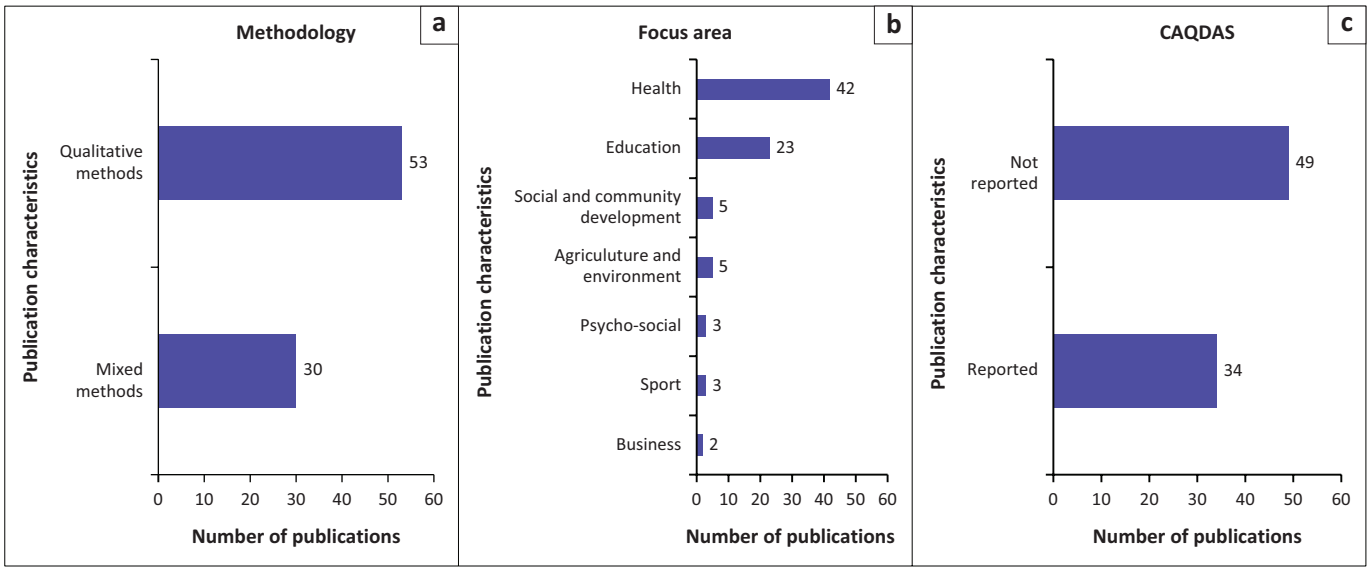


FIGURE 2: Percentage of explained variances of the overall dimensions.



CAQDAS, Computer-Assisted Qualitative Data Analysis Software.

FIGURE 3: The three coded publication characteristics (a) methodology, (b) focus area and (c) CAQDAS.

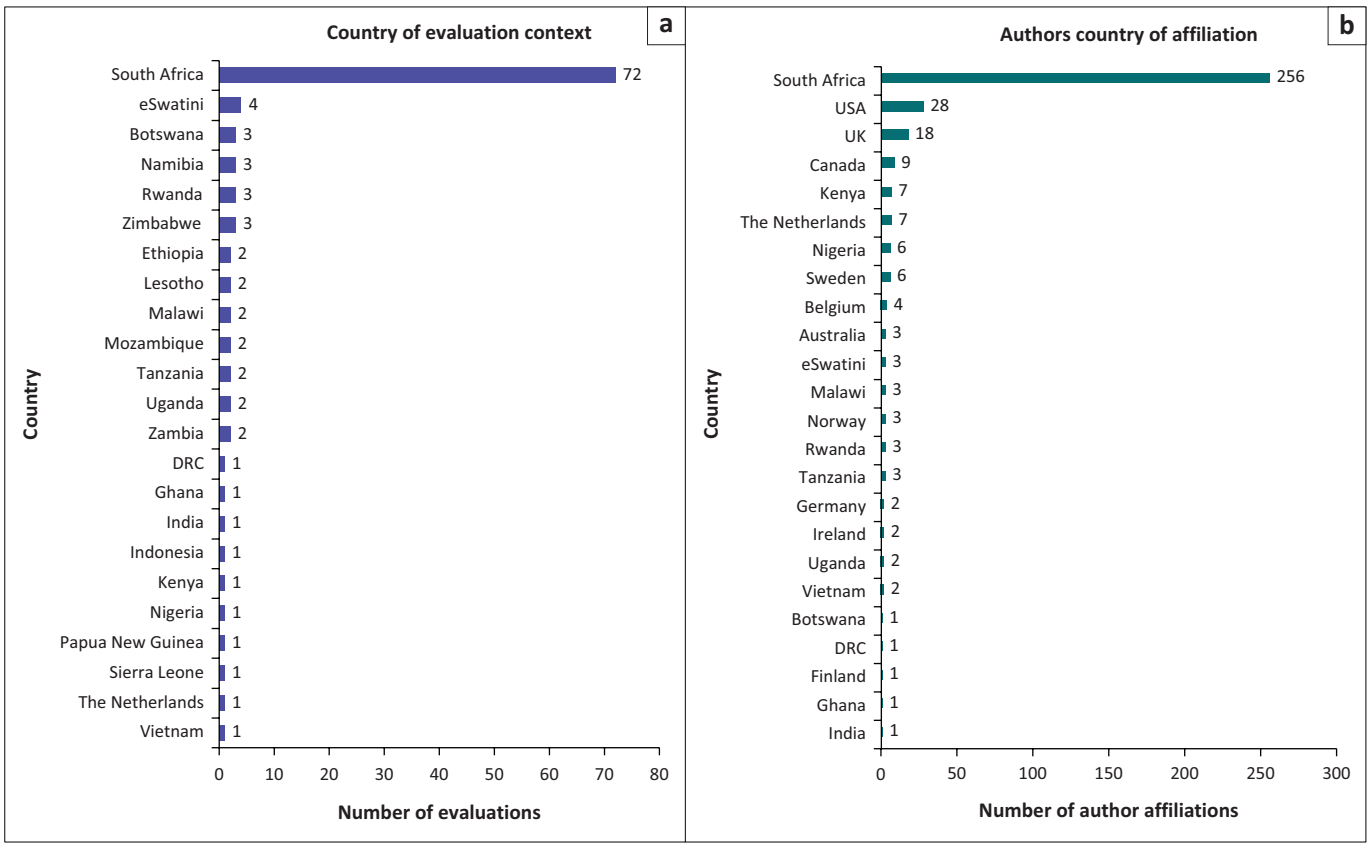


FIGURE 4: Evaluation context and authors' country of affiliation (a) country of evaluation context and (b) authors country of affiliations.

General description of the studies and qualitative approaches identified

Our analysis covered 83 qualitative evaluation studies conducted in Africa, grouped by methodological orientation, thematic focus and CAQDAS use. As shown in Figure 3, more than half of the articles (63.9%) used qualitative methods, while others (36.1%) adopted mixed methods, combining qualitative and quantitative data. Interviews were the predominant data source (77.1%), frequently supported by focus group discussions (34.9%), reflecting participatory tendencies (see Figure 3).

Evaluations spanned seven thematic domains, with health (50.6%) and education (27.7%) dominating the corpus, while sectors like business (two studies) and sports (three studies) were underrepresented (Figure 3). Geographically, the corpus was heavily concentrated in South Africa, which accounted for 72 evaluations, with 90% of authors (n = 286) affiliated

with South African institutions. In contrast, only a small number of studies were conducted in eSwatini (four studies) or involved cross-national collaborations (five studies) (Figure 4). This concentration reflects South Africa's established prominence in African evaluation scholarship (Erasmus, Jordaan & Stewart 2020; Mouton et al. 2018). These findings should therefore be interpreted with caution, as the dominance of South African cases may limit broader generalisation across Africa.

Overall, the data highlight growing use of qualitative approaches in African evaluation, both as standalone methods and in combination with quantitative methods. Notably, 41% of studies employed CAQDAS, particularly NVivo (18 studies) and ATLAS.ti (14 studies), underscoring the growing role of digital tools in supporting qualitative workflows. No use of artificial intelligence (AI)-assisted qualitative analysis tools was identified, consistent with the July 2022 data collection timeframe, prior to their wider adoption.

Trustworthiness strategies

The analysis identified 23 trustworthiness strategies, categorised as primary design techniques (fundamental to methodological rigour) and additional trustworthiness techniques (supporting interpretive accountability). Table 2 presents these strategies by comprehensiveness of reporting. The findings reveal substantial variability in reporting practices, with strong adherence to foundational strategies but notable gaps in the reporting of an additional category.

Primary design techniques

Reporting on primary design techniques was strongest where strategies are foundational and characteristic of published text – notably thick description, data collection procedures and study design – but weakened considerably where deeper methodological reflection was required. Thick description (89.2%) and data collection (86.8%) were the most comprehensively reported, while ethical considerations, though addressed in over 60% of studies, were absent in more than a quarter.

By contrast, the strategies most critical to methodological accountability – analytic details and sampling – were the least comprehensively reported. Mention was common, but substantive explanation was rare: while 83.1% of studies referenced some form of analysis, only 13.3% offered detailed accounts of coding approaches, analytic progression or verification procedure. The gap between presence and depth was equally stark for sampling, which appeared in 90.4% of studies but was comprehensively described in only 9.6% – with strategy justification, sample size rationale and recruitment procedures rarely elaborated. One study's account that 'themes emerged through open coding', without clarifying how codes were consolidated or validated, illustrates a broader pattern: analytic processes were acknowledged but not made transparent enough for independent scrutiny. These omissions are consequential as they affect assessments of transferability and rigour.

TABLE 2: Trustworthiness strategies identified.

Techniques	Key	*N* out of 83% of total					
		Comprehensive		Partial		Not reported	
		n	%	*n*	%	*n*	%
Triangulation	Add.	77	92.77	-	-	6	7.23
Thick description	Pri.	74	89.16	9	10.84	-	-
Multivocality	Add.	74	89.16	2	2.41	7	8.43
Data collection	Pri.	72	86.75	11	13.25	-	-
Design	Pri.	63	75.90	20	24.10	-	-
Comparison & contrast	Add.	58	69.88	15	18.07	10	12.05
Ethical consideration	Pri.	52	62.65	9	10.84	22	26.51
Limitations and delimitations	Pri.	37	44.58	18	21.69	28	33.73
#Inter-coder agreement	Add.	30	36.14	-	-	45	54.22
Member checking	Add.	18	21.69	3	3.61	62	74.70
Persistent observation	Add.	13	15.66	-	-	70	84.34
Rival explanations	Add.	12	14.46	-	-	71	85.54
Others	Add.	12	14.46	4	4.82	67	80.72
Reflexivity	Pri.	11	13.25	10	12.05	62	74.70
Analytic details	Pri.	11	13.25	69	83.13	3	3.61
Audit trail	Add.	9	10.84	1	1.20	73	87.95
Sampling	Pri.	8	9.64	75	90.36	-	-
Expert checking	Add.	8	9.64	-	-	75	90.36
Weighting evidence	Add.	8	9.64	-	-	75	90.36
Codebook	Add.	6	7.23	10	12.05	67	80.72
Peer debriefing	Add.	4	4.82	-	-	79	95.18
Negative case	Add.	2	2.41	24	28.92	57	68.67
Prolonged engagement	Add.	1	1.20	2	2.41	80	96.39

Note: #Inter-coder reliability was assessed in 75 studies, excluding eight studies with a sole coder or analyst.
Pri., Primary design technique; Add., Additional trustworthiness technique.

Additional trustworthiness techniques

Among additional techniques, uptake followed a clear pattern: strategies that leave visible traces in a research report were well-represented, while those requiring sustained relational or analytic investment were largely absent. Triangulation was the most consistently applied (92.8%), typically through integration of data sources (67 studies) or methods (62 studies). For example, some evaluations combined interviews, focus groups and documents, while others paired qualitative case studies with surveys. Multivocality (89.2%) and comparison and contrast (69.9%) were similarly strong, reflecting a tendency to incorporate diverse voices and perspectives within the written account of findings.

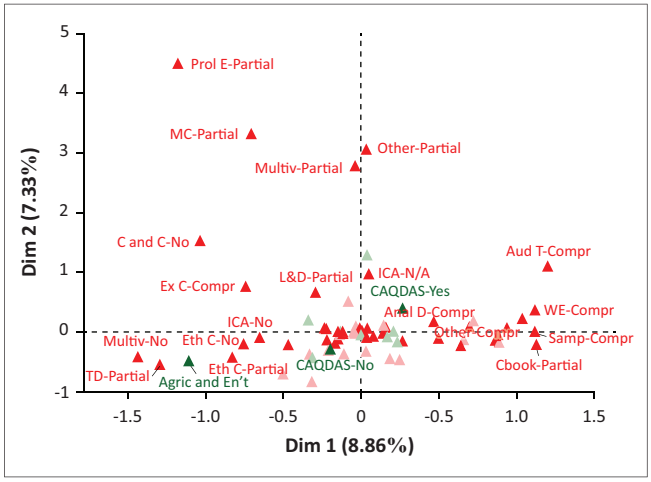
The contrast with participant-centred engagement strategies is striking. Member checking – one of the most direct mechanisms for validating data and interpretations with those who provided the data – was fully reported in only 21.7% of studies and absent in nearly three quarters. Prolonged engagement and persistent observation were almost entirely missing (1.2%). These underuses are striking, given that nearly all studies involved direct human interaction. This indicates preferences for strategies that leave visible textual traces over those requiring documented relational engagement.

Analytic transparency was the most consistently underdeveloped dimension of trustworthiness reporting under the additional techniques category. Audit trails, codebook descriptions, peer debriefing and engagement with rival explanations were rarely evidenced, suggesting limited documentation of analytic decision-making processes. Notably, over half of studies involving multiple coders offered no account of how coding discrepancies were resolved undermining confidence in the consistency of analytic procedures.

Patterned variation in trustworthiness reporting

In Figure 5, our MCA results focus on categories with $\cos^2 \geq 0.1$. The MCA reveals that variation in trustworthiness reporting is patterned rather than random, structured by sectoral context and the presence or absence of digital analytical tools. CAQDAS use ($\cos^2 = 0.12$ for Yes; 0.12 for No) and agriculture and environment focus ($\cos^2 = 0.10$) were the only document characteristics with statistically significant contributions to the dimensions ($\cos^2 \geq 0.1$), while other characteristics – thematic focus and methodological orientation – informed broader thematic interpretation without reaching statistical significance.

Three clusters emerge from the structure (summarised in Table 3). The first, defined by the positive pole of Dimension 1, groups studies with the most comprehensive reporting profiles, particularly in sampling, analytic detail, ethics and audit trails. Health evaluations and standalone qualitative studies are thematically proximate to this cluster, though their weak statistical loadings ($\cos^2 < 0.1$) indicate that sectoral affiliation does not reliably predict rigour. Comprehensive reporting in this cluster appears to reflect



Note: Dim represents dimensions. The active categories are in red, representing the categories of the trustworthiness variables. The supplementary categories in green represent the categories of the document variables. The biplot uses a $\cos^2$ threshold (≥ 0.1) to filter out categories with weak representation on the dimensions.

ICA, inter-coder agreement; CAQDAS, Computer-Assisted Qualitative Data Analysis Software; MCA, multiple correspondence analysis.

FIGURE 5: Multiple correspondence analysis biplot of the 27 variable categories.

TABLE 3: Key relationships between trustworthiness strategies and document characteristics.

Cluster	Key strategies (Coordinates, $\cos^2$)	Document characteristics (Coordinates, $\cos^2$)	Interpretation
1. Comprehensive strategies (Dim 1 +)	- *Sampling-Comprehensive* (1.11, 0.13) - *Analytic Details- Comprehensive* (1.03, 0.16) - *Ethical Considerations- Comprehensive* (0.46, 0.36) - *Audit Trail- Comprehensive* (1.20, 0.17)	**Focus Area:** - *Health* (0.21, 0.04) **Method:** - *Qualitative* (0.16, 0.06)	Health evaluations and qualitative studies are thematically linked to comprehensive strategies (sampling, analysis, ethics). However, their statistical contributions to Dim 1 are weak ($\cos^2 < 0.1$), suggesting rigour is prioritised inconsistently.
2. Partial strategies (Dim 2 +)	- *Member Checking-Partial* (3.32, 0.41) - *Prolonged Engagement-Partial* (4.49, 0.50)	**Focus Area:** - *Psycho-social* (1.28, 0.06) **Method:** - *Mixed* (-0.34, 0.06) **Tool:** - *CAQDAS-Yes* (0.39, 0.12)	CAQDAS use (Dim 2 +) correlates with partial reporting of engagement strategies (e.g. member checking). Psycho-social evaluations and mixed methods are weakly associated (low $\cos^2$) but align thematically with this cluster.
3. Absent and partial strategies (Dim 1 –)	- *Multivocality-No* (-1.43, 0.19) - *Ethical Considerations-No* (-0.76, 0.21) - *ICA-No* (-0.47, 0.26) - *TD-Partial* (-1.29, 0.20)	**Focus Area:** - *Agriculture & Environment* (-1.22, 0.10) **Tool:** - *CAQDAS-No* (-0.20, 0.05) **Method:** - *Mixed* (-0.34, 0.06)	Agriculture and environment studies are statistically linked to absent strategies (multivocality-No, ethical considerations-No) and partial thick descriptions (TD-Partial). Non-CAQDAS use is strongly loaded on Dim 2 ($\cos^2 = 0.12$) but weakly linked with this cluster, along with mixed methods.

CAQDAS, Computer-Assisted Qualitative Data Analysis Software; ICA, inter-coder agreement; TD, thick description.

deliberate methodological commitment rather than structural characteristics.

The second cluster, organised along Dimension 2, is characterised by partial rather than absent strategies, notably member checking and prolonged engagement. Computer-Assisted Qualitative Data Analysis Software use ($cos^2 = 0.12$) is the only statistically significant characteristic associated with this cluster. Rather than aligning with analytic accountability mechanisms (e.g. audit trails, reflexivity, codebook development or inter-coder agreement), software use is associated with partial reporting of engagement strategies. This suggests that CAQDAS may support the organisation and documentation of analytic work, and this potential is not reflected in the reporting of analytic processes, nor does it extend to strengthening relational and field-based practices central to trustworthiness.

The third cluster, anchored at the negative pole of Dimension 1, presents the starkest pattern. Agriculture and environment evaluations are statistically linked to the absence of multivocality, ethical reporting and inter-coder agreement, and to only partial thick description – a concentration of omissions indicative of sector-specific capacity constraints. While CAQDAS-No appears (–0.20, 0.05) in proximity to this cluster, it is more strongly associated with Dimension 2, indicating contextual co-occurrence rather than a direct causal role.

Together, the clusters suggest trustworthiness reporting in the WoS-indexed evaluation studies analysed is shaped less by methodological orientation than by sectoral and technological contexts, with agriculture and environment emerging as areas of particular concern.

Discussion

This study examined how WoS-indexed qualitative evaluations in African contexts report trustworthiness strategies. We used a qualitative examination of strategy distribution and MCA to analyse variation by methodology, sector and CAQDAS use. Four key insights emerged: increased use of qualitative methods and CAQDAS, uneven application of trustworthiness strategies, sectoral patterning of reporting and limited contribution of CAQDAS to deeper trustworthiness practices.

Firstly, we observed increasing reliance on qualitative approaches, either standalone or combined with quantitative methods. This aligns with prior findings (Ndhlovu et al. 2019) highlighting growing use of qualitative approaches in the African subfield. Computer-Assisted Qualitative Data Analysis Software tools, especially NVivo and ATLAS.ti, are also becoming more common, echoing global trends (O'Kane et al. 2023). Given the historical dominance of quantitative methods in evaluation, this shift marks a significant turn towards interpretive approaches (Smith, Pophiw & Tirivanhu 2019). However, increased uptake has not been matched by greater depth in trustworthiness reporting.

The distribution of evaluations and author affiliations remains uneven within the WoS-indexed corpus analysed, with South Africa producing the most studies and author contributions. While this concentration likely reflects institutional strength (Erasmus et al. 2020; Mouton et al. 2018), it raises questions about whether observed reporting patterns generalise beyond this context. Expanding the evidence base beyond South Africa-dominated literature is therefore both a methodological and a developmental priority.

Secondly, our results reveal a clear divide between primary design techniques and the complementary strategies, mirroring Liao and Hitchcock (2018) and extending their findings to the African evaluation context. Primary design elements like data collection (86.8%), thick description (89.2%), research design (75.9%) and ethics (62.7%) were frequently reported comprehensively. By contrast, additional strategies were often under-reported, despite some exceptions such as triangulation (92.8%) and multivocality (89.2%). Many strategies were rarely reported: prolonged engagement (96.4% unreported), peer debriefing (95.2%), expert checking and weighting evidence (90.4%), audit trails (88.0%) and rival explanations (85.5%).

These patterns are meaningful indicators of how trustworthiness is communicated in evaluation studies, even if reported strategies do not fully capture the underlying scope of practice. Journal word limits and reporting conventions may constrain what is documented, meaning the absence of reporting does not imply absence in practice. Our analysis, therefore, assesses communicative transparency (Lub 2015) – what the evaluation authors make visible in published texts. Within this scope, these patterns provide consequential insight into how trustworthiness is demonstrated and interpreted.

Evaluation authors tend to report what is structurally expected in published research texts – design, data collection and thick description – while treating strategies requiring analytic and relational accountability as discretionary. This produces what Lub (2015:7) terms a 'procedural charade' in reporting. This discrepancy may reflect a tacit assumption that primary design techniques are obligatory across paradigms, while additional techniques are context-dependent or optional (Hammersley 2007). Regardless of its source, this pattern results in reporting that is visible but not sufficiently transparent, limiting the ability to assess interpretive rigour (Anastas 2004). In this respect, the African evaluation literature examined largely aligns with broader qualitative evaluation patterns (see Liao & Hitchcock 2018), though the need to embed greater analytic and relational transparency and reflexivity is particularly urgent where evaluation credibility informs development decisions. This finding reinforces concerns by Chirau et al. (2019), Ndhlovu et al. (2019) and Ngwabi and Wildschut (2019) about inconsistency in methodological application evidenced in evaluation reports from Africa.

TABLE 4: Functional categorisation of trustworthiness strategies identified.

Category	Reported with higher depth (≥ 60% Comprehensive Reporting)	Under-reported (< 20% Comprehensive Reporting)
Strategies related to designs and data gathering	Triangulation (92.77%) Data collection (86.75%) Design (75.90%)	Sampling (9.64%) Limitations and delimitations (44.58%) (moderate but not highly reported)
Strategies related to the analytical process	Multivocality (89.16%) Thick description (89.16%) Comparison and contrast (69.88%)	Inter-coder agreement (36.14%) Analytic details (13.25%) Reflexivity (13.25%) Audit trail (10.84%) Rival explanations (14.46%) Negative case analysis (2.41%) Codebook (7.23%) Weighting evidence (9.64 %)
Study participants-related strategies	Ethical considerations (62.65%)	Member checking (21.69%) Persistent observation (15.66%) Prolonged engagement (1.20%)

To further illustrate variation in reporting depth, we categorised the strategies into three functional groups – design and data gathering, analytic process and participant-related strategies – as shown in Table 4.

Strategies related to designs and data gathering demonstrate the strongest performance, with triangulation (92.8%), data collection (86.8%) and study design (75.9%) achieving high comprehensive reporting rates. However, sampling strategies remain critically under-reported (9.64%), consistent with Ngwabi and Wildschut (2019). This gap is consequential, as sampling decisions underpin assessments of transferability and stakeholder trust. Good practice from the dataset typically provides sufficient information on programme contexts, sampling strategies, justifications, procedures, sample sizes and participant characteristics (e.g. Jarvis et al. 2023; Stern et al. 2019). For example, Stern et al. (2019:6), justified purposeful sampling 'to represent a diversity of environments including rural and peri-urban locations', an account that enables readers to assess representativeness in ways that most studies in our corpus did not (Gentles et al. 2015).

Strategies related to the analytical process exhibited the most pronounced disparities. While multivocality (89.2%), thick description (89.2%) and comparison and contrast (69.88%) were well-demonstrated, analytic accountability mechanisms – audit trails (10.8%), reflexivity (13.3%), codebook use (7.2%) and negative case analysis (2.4%) – were rarely reported. This reflects a preference for descriptive over verificatory analytic practices, where studies demonstrated findings more readily than they accounted for how those findings were produced. These patterns align with Chirau et al. (2019) and point to a central limitation: without documented analytic processes, evaluation users cannot assess whether interpretive claims are systematically grounded.

Participants-related strategies show the poorest performance. Only ethical considerations reached moderate reporting levels (62.7%) while core relational practices – member checking (21.7%), persistent observation (15.7%) and prolonged engagement (1.20%) – were rarely documented sufficiently. This is particularly significant in the African context, where 'Made in Africa' evaluation frameworks emphasise community engagement and relational accountability (Abrahams, Masvaure & Morkel 2022; Pophiwa & Saidi 2022). The limited reporting of these practices, which is also observed in Chirau et al.'s (2019), raises a critical question: whether the adoption of qualitative approaches reflects their epistemological commitments or primarily their procedural forms.

Our MCA reinforces these patterns by showing that variation in trustworthiness reporting is structurally patterned rather than incidental. The comprehensive strategies cluster (Dim 1+) confirms that stronger methodological reporting – particularly in sampling, analytic detail, ethics and audit trails – is associated with health sector evaluations and standalone qualitative studies, although these associations remain statistically weak. This is unsurprising, given the strong qualitative traditions in health studies and rigorous methodological expectations in health research (Hadi & José Closs 2016; Orr et al. 2020). Comprehensive reporting in this cluster appears to reflect methodological commitment rather than structural feature of the studies themselves – a finding that underscores the role of researcher capacity and editorial standards in shaping reporting quality.

The Partial Strategies cluster (Dim 2+) highlights incomplete reporting of engagement strategies, particularly member checking and prolonged engagement. Computer-Assisted Qualitative Data Analysis Software use is associated with this cluster, indicating that software adoption aligns with partial rather than comprehensive enactment of these strategies. This is notable given expectations that CAQDAS analytic rigour – through audit trails, coding systems, and reflexivity – as well as transparency (Kapiszewski & Karcher 2021; Woods, Macklin & Lewis 2016). Our experience teaching and using ATLAS.ti confirm these embedded capabilities, yet this potential is not reflected in reporting. Computer-Assisted Qualitative Data Analysis Software use shows no clear link to analytic accountability and only partial alignment with engagement practices, suggesting tool adoption alone does not translate to communicative transparency. Rather, limited methodological grounding may conflate technical tools with analytic approaches (Woods et al. 2016), reinforcing the observation that CAQDAS does not resolve methodological deficiencies (Niedbalski & Ślęzak 2022).

The absent and partial strategies cluster (Dim 1–), dominated by agriculture and environment studies, reflects the most pronounced reporting gaps. The association with missing or partial reporting on analytic and relational accountability mechanisms points to sector-specific capacity constraints. Its link to mixed-methods studies further indicates that qualitative components may be treated as secondary, consistent with prior research on qualitative underreporting

in mixed-methods designs (O'Cathain, Murphy & Nicholl 2017; Wisdom et al. 2012).

Together, these findings underscore a dual narrative that has both methodological and developmental implications. The studies in our WoS-indexed corpus demonstrate strong adherence to structurally visible elements of qualitative rigour, yet several strategies critical to interpretive accountability remain under-reported. Addressing this gap requires more than improved tools; it demands sustained capacity building that engages evaluators with the epistemological foundations of qualitative inquiry, not only its reporting conventions.

Limitations and delimitations

Several caveats should be considered when interpreting the findings of this study. While our analysis focuses on qualitative evaluation research in African contexts, it is limited to WoS-indexed, peer-reviewed publications, excluding unpublished and internal organisational evaluation reports that may provide additional insights. This study complements previous investigations into these other areas in the African context (Ndhlovu et al. 2019; Ngwabi & Wildschut 2019). The focus on WoS-indexed and English-language publications may also exclude relevant studies, particularly from non-English-speaking contexts, with implications for geographic representation. All coding was conducted by a single coder. While several trustworthiness strategies were applied, coding decisions inevitably reflect interpretive judgement, as is standard in qualitative research.

Additionally, our analysis is based on reported methodological practices, meaning actual evaluation practices may differ from what is documented. The findings should therefore be interpreted as indicative of reporting patterns rather than definitive assessments of practice or quality. The study also does not assess the substantive findings of the evaluations themselves.

Finally, consistent with its exploratory design, MCA was employed to identify patterns in the data. The results therefore reflect associations rather than causal relationships. Future research could benefit from larger and more exhaustive samples that explore causal links between methodological rigour and factors such as donor influence, commissioner-evaluator relationships and institutional contexts. Future research could extend this work using larger and more diverse samples to examine how trustworthiness relates to factors such as donor influence, commissioner–evaluator relationships and institutional contexts.

Conclusion

This study assessed how WoS-indexed qualitative evaluations in African contexts report trustworthiness strategies, identifying key gaps, sectoral disparities, and the influence of methodological and technological factors on reporting practices. As qualitative approaches gain prominence across fields, concerns have shifted from *whether* trustworthiness standards are necessary to *how* they are implemented to ensure credible and systematic inquiry (Megheirkouni & Moir 2023). Given the pivotal role of evaluation in decision-making, understanding how trustworthiness is embedded in evidence generation is critical, particularly in African contexts where capacity building remains a priority (Abrahams 2019; Mouton et al. 2018).

Our findings reveal a consistent pattern: studies are most transparent about structurally visible methodological elements, while strategies critical to interpretive accountability remain under-reported. Although qualitative and mixed-methods approaches and CAQDAS use (41% of studies) are increasing, uptake has not been matched by depth in trustworthiness reporting. The corpus is geographically concentrated, with South Africa accounting for most studies by context and authorship, which limits broader generalisation. Reporting gaps are most evident in sampling (9.6%), analytic detail (13.3%) and participant engagement strategies, reflecting a pattern of reporting that is visible but not sufficiently transparent.

Multiple correspondence analysis further shows that these patterns are structured rather than incidental. Comprehensive reporting clusters with health and standalone qualitative studies but lacks strong statistical association, suggesting that rigour reflects deliberate methodological commitment within these contexts. Computer-Assisted Qualitative Data Analysis Software use correlates with partial reporting of engagement strategies rather than analytic accountability, highlighting a gap between tool adoption and methodological depth. Agriculture and environmental studies show the weakest reporting, while mixed-methods studies tend to underreport qualitative components, consistent with broader concerns about their marginalisation (Wisdom et al. 2012).

In addition, this study offers a consolidated set of trustworthiness strategies, organised into three functional categories: design-related strategies (e.g. research design, sampling, data collection), analytic accountability strategies (e.g. audit trails, reflexivity, codebooks, inter-coder agreement) and relational strategies (e.g. member checking, prolonged engagement, multivocality). This functional framing provides a practical entry point for linking methodological principles to their application in qualitative evaluation.

Addressing the identified gaps requires more than improved tools, even amid the rapid adoption of CAQDAS and emerging AI tools. Greater emphasis is needed on under-reported strategies – sampling, analytic transparency, reflexivity and participant engagement – to align practice with the epistemological commitments of qualitative inquiry. Targeted training and sector-specific capacity building, particularly in agriculture, environmental, and mixed-methods contexts, are therefore essential.

These responsibilities extend across the evaluation ecosystem. Journals can strengthen reporting standards through clearer trustworthiness criteria; professional bodies and VOPEs can support capacity building; and funders can incentivise rigorous qualitative reporting through commissioning requirements. Strengthening these practices will enhance confidence in evaluation evidence and its use in policy and decision-making. Ultimately, the credibility of evaluation depends not on the adoption of qualitative methods, but on the extent to which their commitments are enacted and transparently reported.

Acknowledgements

The authors express their gratitude to Dr Mpho Mfata and the anonymous reviewers for their valuable input in improving this article.

The authors also appreciate the feedback received following the presentation of preliminary findings from the pilot phase of this research at the 8th World Conference on Qualitative Research (WCQR2024), held in January 2024 in São Miguel, Azores, Johannesburg and online. The present article is based on a subsequent analysis that differs from the work presented at the conference.

Competing interests

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

CRediT authorship contribution

Dominic A. Okoliko: Conceptualisation, Data curation, Formal analysis, Investigation, Methodology, Project administration, Software, Visualisation, Writing – original draft, Writing – review & editing. Lauren P. Wildschut: Conceptualisation, Methodology, Supervision, Writing – review & editing. All authors reviewed the article, contributed to the discussion of results, approved the final version for submission and publication, and take responsibility for the integrity of its findings.

Funding information

Funding for the publication of this article was provided by the Centre for Research on Evaluation, Science and Technology (CREST). During the research period, Dominic A. Okoliko also received postdoctoral funding from CREST.

Data availability

The data analysed in this study are available from the corresponding author, Dominic A. Okoliko, upon reasonable request.

Disclaimer

The views and opinions expressed in this article are those of the authors and are the product of professional research. They do not necessarily reflect the official policy or position of any affiliated institution, funder, agency or that of the publisher. The authors are responsible for this article's results, findings and content.

References

Abrahams, M., Masvaure, S. & Morkel, C., 2022, 'Made in Africa evaluation', *African Evaluation Journal* 10(1), 1–3. https://doi.org/10.4102/aej.v10i1.665

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

Abrahams, M.A., 2019, 'Editorial-2019: Omniscience of monitoring and evaluation', *African Evaluation Journal* 7(1), 1–2. https://doi.org/10.4102/aej.v7i1.433

AfrEA, 2021, *African evaluation principles*, African Evaluation Association, viewed 26 March 2026, from https://afrea.org/african-evaluation-principles-2/.

Agonnoude, T.M., Houeto, S.D., Zannou, G.S., Agbo, M., Behanzin, L. & Houehanou, C.Y., 2021, 'Professionalisation of program evaluation in Africa: An imperative for effectiveness and accountability for public policy', *Central African Journal of Public Health* 7(4), 189. https://doi.org/10.11648/j.cajph.20210704.16

Anastas, J.W., 2004, 'Quality in qualitative evaluation: Issues and possible answers', *Research on Social Work Practice* 14(1), 57–65. https://doi.org/10.1177/1049731503257870

Basheka, B.C., 2015, 'The state of Monitoring and Evaluation (M&E) as a discipline in Africa from infancy to adulthood?', *African Journal of Public Affairs* 8(3), 75–93.

Chaplowe, S.G. & Engo-Tega, R.B., 2007, 'Civil society organizations and evaluation lessons from Africa', *Evaluation* 13(2), 257–274. https://doi.org/10.1177/1356389007075227

Chirau, T., Tirivanhu, P. & Ramasobana, M., 2019, 'Appraising qualitative evaluations', in C.B. Mapitsa, P. Tirivanhu & N. Pophiwa (eds.), *Evaluation landscape in Africa: Context, methods and capacity*, 1st edn., pp. 139–169, African Sun Media, Stellenbosch.

Christie, C.A., 2011, 'Analyzing the practice of evaluation: What do these cases tell us about theory?', in J.L. Fitzpatrick, C.A. Christie & M.M. Mark (eds.), *Evaluation in action: Interviews with expert evaluators*, Sage, Thousand Oaks, CA.

Coryn, C.L.S., Ozeki, S., Wilson, L.N., Greenman II, G.D., Schröter, D.C., Hobson, K.A., et al., 2016, 'Does research on evaluation matter? Findings from a survey of American Evaluation Association Members and prominent evaluation theorists and scholars', *American Journal of Evaluation* 37(2), 159–173. https://doi.org/10.1177/1098214015611245

Creswell, J.W. & Miller, D.L., 2000, 'Determining validity in qualitative inquiry', *Theory into Practice* 39(3), 142–130. https://doi.org/10.1207/s15430421tip3903_2

Curtin, M. & Fossey, E., 2007, 'Appraising the trustworthiness of qualitative studies: Guidelines for occupational therapists', *Australian Occupational Therapy Journal* 54(2), 88–94. https://doi.org/10.1111/j.1440-1630.2007.00661.x

Dixon-Woods, M., Shaw, R.L., Agarwal, S. & Smith, J.A., 2004, 'The problem of appraising qualitative research', *Quality and Safety in Health Care* 13(3), 223–225. https://doi.org/10.1136/qshc.2003.008714

Enworo, O.C., 2023, 'Application of Guba and Lincoln's parallel criteria to assess trustworthiness of qualitative research on indigenous social protection systems', *Qualitative Research Journal* 23(4), 372–384. https://doi.org/10.1108/QRJ-08-2022-0116

Erasmus, Y., Jordaan, S. & Stewart, R., 2020, 'Scoping the impact evaluation capacity in sub-Saharan Africa', *African Evaluation Journal* 8(1), 11. https://doi.org/10.4102/aej.v8i1.473

Farrington, D.P., 2003, 'Methodological quality standards for evaluation research', *The ANNALS of the American Academy of Political and Social Science* 587(1), 49–68. https://doi.org/10.1177/0002716202250789

Gentles, S.J., Charles, C., Ploeg, J. & Mckibbon, K.A., 2015, 'Sampling in qualitative research: Insights from an overview of the methods literature', *The Qualitative Report* 20(11), 1772–1789. https://doi.org/10.46743/2160-3715/2015.2373

Guba, E.G., Lincoln, Y.S. & Lynham, S.A., 2018, 'Paradigmatic controversies, contradictions, and emerging confluences, revisited', in N.K.D.Y.S. Lincoln (ed.), *The handbook of qualitative research*, pp. 352–394, Sage, London.

Hadi, M.A. & José Closs, S., 2016, 'Ensuring rigour and trustworthiness of qualitative research in clinical pharmacy', *International Journal of Clinical Pharmacy* 38(3), 641–646. https://doi.org/10.1007/s11096-015-0237-6

Hammersley, M., 2007, 'The issue of quality in qualitative research', *International Journal of Research & Method in Education* 30(3), 287–305. https://doi.org/10.1080/17437270701614782

Hsieh, H.F. & Shannon, S.E., 2005, 'Three approaches to qualitative content analysis', *Qualitative Health Research* 15(9), 1277–1288. https://doi.org/10.1177/1049732305276687

Husson, F., Le, S. & Pagès, J., 2017, *Exploratory multivariate analysis by example using R*, 2nd edn., Chapman and Hall/CRC, New York, NY.

Jarvis, M.A., Pillay, S.R., Norton, L.M., Hiraman, N. & Baloyi, O.B., 2023, 'Delivering a psychosocial program for older people living in retirement homes during the Covid-19 pandemic: A process evaluation and recommendations for community interventions', *Journal of Community Psychology* 51(1), 31–50. https://doi.org/10.1002/jcop.22876

Kapiszewski, D. & Karcher, S., 2021, 'Transparency in practice in qualitative research', *PS: Political Science & Politics* 54(2), 285–291. https://doi.org/10.1017/S1049096520000955

Kienstra, N.H.H. & Van der Heijden, P.G.M., 2015, 'Using correspondence analysis in multiple case studies', *BMS Bulletin of Sociological Methodology/Bulletin de Methodologie Sociologique* 128(1), 5–22. https://doi.org/10.1177/0759106315596920

Leslie, M., Moodley, N., Goldman, I., Jacob, C., Podems, D., Everett, M. et al., 2015, 'Developing evaluation standards and assessing evaluation quality', *African Evaluation Journal* 3(1), 13. https://doi.org/10.4102/aej.v3i1.112

Liao, H. & Hitchcock, J., 2018, 'Reported credibility techniques in higher education evaluation studies that use qualitative methods: A research synthesis', *Evaluation and Program Planning* 68, 157–165. https://doi.org/10.1016/j.evalprogplan.2018.03.005

Lincoln, Y.S. & Guba, E.G., 1985, *Naturalistic inquiry*, Sage Publications, Beverly Hills, CA.

Lub, V., 2015, 'Validity in qualitative evaluation: Linking purposes, paradigms, and perspectives', *International Journal of Qualitative Methods* 14(5), 1609406915621406. https://doi.org/10.1177/1609406915621406

McWilliam, R.A., 2000, 'Author and reviewer guideline series: Reporting qualitative studies', *Journal of Early Intervention* 23(2), 77–80. https://doi.org/10.1177/105381510002300202

Megheirkouni, M. & Moir, J., 2023, 'Simple but effective criteria: Rethinking excellent qualitative research', *The Qualitative Report* 28(3), 848–864. https://doi.org/10.46743/2160-3715/2023.5845

Morse, J.M., 2015, 'Critical analysis of strategies for determining rigor in qualitative inquiry', *Qualitative Health Research* 25(9), 1212–1222. https://doi.org/10.1177/1049732315588501

Mouton, J., Wildschut, L. & Leslie, M., 2018, *Monitoring and evaluation capacity: A landscape analysis*, Zenex Foundation, viewed n.d., from https://www.zenexfoundation.org.za/monitoring-and-evaluation-capacity-a-landscape-analysis/monitoring-and-evaluation-research/.

Ndhlovu, L., Khumalo, L. & Krakra, J., 2019, 'An assessment of methods used in evaluations from the African evaluation database', in C.B. Mapitsa, P. Tirivanhu & N. Pophiwa (eds.), *Evaluation landscape in Africa: Context, methods and capacity*, 1st edn., pp. 111–138, African Sun Media.

Ngwabi, N. & Wildschut, L., 2019, 'Scandinavian donors in Africa: A reflection on evaluation reporting standards', in C.B. Mapitsa, P. Tirivanhu & N. Pophiwa (eds.), *Evaluation landscape in Africa: Context, methods and capacity*, 1st edn., pp. 88–110, African Sun Media, Stellenbosch.

Niedbalski, J. & Ślęzak, I., 2022, 'Encounters with CAQDAS: Advice for beginner users of computer software for qualitative research', *The Qualitative Report* 27(4), 1114–1132.

Novak, A.M. & Jen, E., 2024, 'The editorial word: Trustworthiness', *Journal of Advanced Academics* 35(4), 718–728. https://doi.org/10.1177/1932202X241288384

O'Cathain, A., Murphy, E. & Nicholl, J., 2017, 'The quality of mixed methods studies in health services research', *Journal of Health Services Research & Policy* 13(2), 92–98. https://doi.org/10.1258/jhsrp.2007.007074

O'Kane, P., Ott, D.L., Smith, A.D. & Brown, T.C., 2023, 'Understanding computer-assisted qualitative data analysis software as a tool to enhance systematic literature reviews in human resource development', *Human Resource Development Review* 22(2), 291–307. https://doi.org/10.1177/15344843221144668

Orr, E.R., Ballantyne, M., Gonzalez, A. & Jack, S.M., 2020, 'Visual elicitation: Methods for enhancing the quality and depth of interview data in applied qualitative health research', *Advances in Nursing Science* 43(3), 202–213. https://doi.org/10.1097/ANS.0000000000000321

Patton, M.Q., 2018, 'Evaluation science', *American Journal of Evaluation* 39(2), 183–200. https://doi.org/10.1177/1098214018763121

Peters, M.D.J., Marnie, C., Tricco, A.C., Pollock, D., Munn, Z., Alexander, L. et al., 2020, 'Updated methodological guidance for the conduct of scoping reviews', *JBI Evidence Synthesis* 18(10), 2119–2126. https://doi.org/10.11124/JBIES-20-00167

Pollock, D., Tricco, A.C., Peters, M.D.J., McInerney, P.A., Khalil, H., Godfrey, C.M. et al., 2022, 'Methodological quality, guidance, and tools in scoping reviews: A scoping review protocol', *JBI Evidence Synthesis* 20(4), 1098–1105.

Pophiwa, N. & Saidi, U., 2022, 'Approaches to embedding indigenous knowledge systems in Made in Africa evaluations', *African Evaluation Journal* 10(1), 1–9. https://doi.org/10.4102/aej.v10i1.623

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

Rallis, S.F., Rossman, G.B. & Gajda, R., 2007, 'Trustworthiness in evaluation practice: An emphasis on the relational', *Evaluation and Program Planning* 30(4), 404–409. https://doi.org/10.1016/j.evalprogplan.2007.06.006

Raskind, I.G., Shelton, R.C., Comeau, D.L., Cooper, H.L.F., Griffith, D.M. & Kegler, M.C., 2019, 'A review of qualitative data analysis practices in health education and health behavior research', *Health Education & Behavior: The Official Publication of the Society for Public Health Education* 46(1), 32–39. https://doi.org/10.1177/1090198118795019

Santiago-Delefosse, M., Gavin, A., Bruchez, C., Roux, P. & Stephen, S.L., 2016, 'Quality of qualitative research in the health sciences: Analysis of the common criteria present in 58 assessment guidelines by expert users', *Social Science & Medicine* 148, 142–151. https://doi.org/10.1016/j.socscimed.2015.11.007

Sibanda, A., Sibanda, T.D. & Sibanda, T.N., 2023, 'Towards equitable evaluation through the use of the African evaluation principles', *African Evaluation Journal* 11(1), 11. https://doi.org/10.4102/aej.v11i1.697

Smith, L., Pophiw, N. & Tirivanhu, P., 2019, 'Introduction', in C.B. Mapitsa, P. Tirivanhu & N. Pophiwa (eds.), *Evaluation landscape in Africa: Context, methods and capacity*, 1st edn., pp. 1–17, African Sun Media, Stellenbosch.

Spencer, L., Ritchie, J., Lewis, J. & Dillon, L., 2003, *Quality in qualitative evaluation: A framework for assessing research evidence*, Government Chief Social Researcher's Office, London.

Stern, E., Gibbs, A., Willan, S., Dunkle, K. & Jewkes, R., 2019, '"When you talk to someone in a bad way or always put her under pressure, it is actually worse than beating her": Conceptions and experiences of emotional intimate partner violence in Rwanda and South Africa', *PLoS One* 14(11), e0225121. https://doi.org/10.1371/journal.pone.0225121

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

Tirivanhu, P., 2022, 'Whither Made in Africa evaluation: Exploring the future trajectory and implications for evaluation practice', *African Evaluation Journal* 10(1), 10. https://doi.org/10.4102/aej.v10i1.614

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 Health Care* 19(6), 349–357. https://doi.org/10.1093/intqhc/mzm042

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

Wildschut, L. & Silubonde, T.R., 2020, 'Evaluation education in South Africa: Characteristics and challenges in a changing world', *African Evaluation Journal* 8(1), 1–11. https://doi.org/10.4102/aej.v8i1.476

Wisdom, J.P., Cavaleri, M.A., Onwuegbuzie, A.J. & Green, C.A., 2012, 'Methodological reporting in qualitative, quantitative, and mixed methods health services research articles', *Health Services Research* 47(2), 721–745. https://doi.org/10.1111/j.1475-6773.2011.01344.x

Woods, M., Macklin, R. & Lewis, G.K., 2016, 'Researcher reflexivity: Exploring the impacts of CAQDAS use', *International Journal of Social Research Methodology* 19(4), 385–403. https://doi.org/10.1080/13645579.2015.1023964