

Johnson, Susan; Rasulova, Saltanat

Working Paper

Qualitative impact evaluation: Incorporating authenticity into the assessment of rigour

Bath Papers in International Development and Wellbeing, No. 45

Provided in Cooperation with:

Centre for Development Studies (CDS), University of Bath

Suggested Citation: Johnson, Susan; Rasulova, Saltanat (2016) : Qualitative impact evaluation: Incorporating authenticity into the assessment of rigour, Bath Papers in International Development and Wellbeing, No. 45, University of Bath, Centre for Development Studies (CDS), Bath

This Version is available at:

<https://hdl.handle.net/10419/179372>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

Centre for
Development Studies
(CDS)



Bath Papers in International Development and Wellbeing No:

45/2016

**Qualitative impact evaluation:
incorporating authenticity into the assessment of rigour**

Susan Johnson and Saltanat Rasulova

Centre for Development Studies, University of Bath

Oxford Policy Management and Centre for Development Studies, University of Bath

© Susan Johnson and Saltanat Rasulova, 2016

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means without the prior permission in writing of the publisher, nor be issued to the public or circulated in any form other than that in which it is published.

Published by:

The Centre for Development Studies, University of Bath
Claverton Down, Bath, BA2 7AY, UK

<http://www.bath.ac.uk/cds/>

ISSN 2040-.....3151

Series Editor:

Susan Johnson

The Centre for Development Studies at the University of Bath is an interdisciplinary collaborative research centre critically engaging with international development policy and practice.

UNIVERSITY OF BATH
CENTRE FOR DEVELOPMENT STUDIES
BATH PAPERS IN INTERNATIONAL DEVELOPMENT AND WELL-BEING
NO. 45 – MARCH 2016

Qualitative impact evaluation: incorporating authenticity into the assessment of rigour

Susan Johnson and Saltanat Rasulova

Centre for Development Studies, University of Bath

Oxford Policy Management and Centre for Development Studies, University of Bath

Abstract:

Recent developments in impact evaluation recognise the need to go beyond the intense debate over experimental techniques to incorporate theory-based approaches and qualitative research methods. Motivated by an underlying concern that qualitative research in this new wave of qualitative impact evaluation research is appropriately conducted, this paper reviews practical strategies to address rigour deploying Guba and Lincoln's principles of "trustworthiness" to do so. In particular we focus on the less discussed principle of 'authenticity' which responds to the demand for research orientations to be more transformative and emancipatory. In development impact evaluation, recent discussion has highlighted the frequent deficit of a transformative orientation and the problematic wider organisational contexts of aid relationships in which commissioned evaluations are conducted. We argue that embracing the authenticity principle offers commissioned researchers with a progressive orientation a rationale for making space for stakeholder interaction and negotiation within the rigour discourse. This in turn creates the scope to incorporate it into checklists of rigour so using the 'artefacts' of evaluation as a means to call commissioning organisations and other stakeholders to be more responsive to concerns for authenticity.

Key words:

qualitative research, rigour, trustworthiness, impact evaluation, evidence-based policy

Acknowledgement:

This paper is one of the outputs of a two-year collaborative research project investigating ways to improve rigour of qualitative research in impact evaluation between the University of Bath and Oxford Policy Management. Saltanat Rasulova was a Research Associate at the University of Bath from 2014-2016 when the main research for this paper was undertaken. This project operated through a Knowledge Transfer Partnership (KTP) funded jointly by the UK Technology Strategy Board and the Economic and Social Research Council. We would like to acknowledge this support as well as comments on earlier versions from Laura Camfield, James Copestake and Ian MacAuslan.

Introduction

The last decade has seen a huge shift to the view that policy-making in development should be evidence based, with this shift dominated by the “experimentalist surge” (Picciotto, 2014, p. 32) of randomised control trials (RCTs). These come with the particular claim to rigour as attribution arising from the use of specific counterfactuals. The elevation of RCTs as a ‘gold standard’ of rigour has led to something of a renewed paradigm war in the field of development evaluation in which RCTs have been viewed as the main legitimate form of evidence to the detriment of the potential contribution of qualitative research. This intense debate is now waning to a degree. Lead practitioners acknowledge the importance of qualitative methods to aid experimental design and understanding (Glennister & Takavarasha, 2013) although these are not yet well integrated into the discussion or interpretation of results. The claims of RCTs are increasingly subject to criticism that they are “off the gold standard” and there are calls for the debate to be refocussed on the core functions and purposes of impact evaluation and for the deployment of “relevant rigour” with rigour understood as needing negotiation (Camfield & Duvendack, 2014; Guijt & Roche, 2014, p. 50). Moreover, this narrow view of evidence for policy making, has provoked an important wider concern about its politics. When development is in fact conceptualised as a transformational process based on rights, with a concern for means and not only ends, and with politics and power relations at its centre (Eyben, Guijt, Roche, & Shutt, 2015), the question of how such - apparently value-free - evidence arises, is critical.

Practically also, recognition that a huge range of policy is not evaluable using experimental techniques, has led to moves to broaden the range of methods (Stern et al., 2012) and state-of-the-art thinking on evaluation involving theory-based and realist methods, along with ways of making qualitative work rigorous for causal inference are gaining traction (Stern, 2015). These approaches make use of a range of both quantitative and qualitative evidence but rely very extensively on qualitative methods. This in turn therefore leads to a broader concern regarding the rigour with which research using qualitative data collection and analysis is undertaken (Spencer, et al. 2003; DFID, 2014).

This paper therefore sets out to consider how rigour in qualitative impact evaluation¹ can be better addressed. According to the Oxford English Dictionary, rigour is defined as “the quality of being extremely thorough and careful”. For Guba and Lincoln this amounts to being able to assess the trustworthiness of the processes and procedures of qualitative research and we adopt their framework for doing so. Their trustworthiness framework is seen as “parallel” to positivist notions of rigour. They proposed a further principle for evaluation underpinned by the constructivist paradigm - also known as fourth generation evaluation (1989). This authenticity principle recognises that all judgement and

¹ This paper is primarily concerned with impact evaluation. Many of the same issues in managing qualitative data may arise in assessing other aspects in evaluation. Evaluation assesses a wider range of factors than impact alone, in particular (following Picciotto (2014)) (i) merit – as performance relative to quality standards and stakeholder aspirations – and relevance to beneficiary, country and global needs and priorities; (ii) worth – the net benefits of the intervention and efficacy taking into account relative importance of the objectives; (iii) value – doing things efficiently relative to other ways of doing it.

understanding has a value basis and hence is subject to political and ideological forces. The consequence for evaluation is that it must involve negotiation among stakeholders so that the basis of their understandings and values is part of the process. While their initial position was that third generation evaluation² based in a positivist perspective could not be mixed (2001) with such fourth generation approaches, their later position (2007, p25) recognises that the authenticity principle and its concerns for fairness were helping to address ethical and ideological problems “while at first appearing to be radical...are nevertheless becoming mainstream”.

We take the view that significant further efforts for making the trustworthiness and authenticity principles mainstream are still required. In particular the authenticity principles enable us to engage with the politics of evaluation by bringing in the voices of beneficiaries and engaging them with other stakeholders. While the underlying thinking is familiar within debates over participatory approaches to appraisal and evaluation, the scope for explicit participatory approaches in commissioned evaluations is often limited or - more specifically – not perceived as what the commissioner may be seeking. We review why it is a distinctive criterion of good impact evaluation and offer an initial approach for assessing how far different stages of qualitative research achieve different degrees of authenticity. Although this requires much more development, we indicate the mind-set and orientation that its use demands. We argue that the time is ripe for actively adopting it into checklists of rigour as evaluation “artefacts” which can be used for “playing the game to change the rules” (Guijt, 2015, p. 200).

Further in pursuing this agenda we recognise that there are many forms of qualitative social research and many forms of impact evaluation. We are approaching this discussion in the practical context of commissioned impact evaluations by aid agencies – which invariably fit the mould of what Macdonald called “bureaucratic” evaluation (Norris, 2015) . The paper arises from a collaborative project undertaken between the Centre for Development Studies at the University of Bath and Oxford Policy Management - a consulting firm working in the international development sector. The project has been undertaken as part of a Knowledge Transfer Partnership Grant funded by Innovate UK and the ESRC, which seeks to develop and improve practices of qualitative research and evaluation in the context of a development consultancy practice. In this context it is necessary to remain practical while recognising and remaining reflexive over the wider constraints (Groves & Hinton, 2004). Indeed, the intention is not to arrive at best practice since the variation in contexts, purposes, audiences, costs, time and so on are huge but to engage with the debate about how practice can be improved in this quite difficult terrain. A particular aspect of remaining practical is to address the question of “how much?” since neither Guba and Lincoln’s trustworthiness criteria nor the intense debates in this area engage with questions of resource availability and it is issues of budgets and time that are also critical in practice.

² Guba and Lincoln chart stages through which evaluation practice has developed. First generation evaluation was the simple measurement of intervention achievements that was quantitative and factual. Second generation evaluation involved how achievements related to objectives. The third generation involved the assessment of effects and the making of judgements about success or failure

The paper proceeds as follows. Since the challenge in approaching rigour in qualitative research is that it is a minefield of philosophical debate and disagreement and there are shifting criteria for judging quality in evaluation arising from different paradigms (e.g. Patton, 2002), we start with a review of the philosophical foundations of quantitative and qualitative research methodologies. Primarily the charge that rigour is weak in qualitative research is based on these differences – both by comparison to quantitative research but even also in its own terms. This mirrors the original paradigm wars of the late 20th century. With this philosophical background in place, the second section discusses the frameworks that have been proposed for developing and assessing rigour in qualitative research. This raises the question as to whether or not one can even agree that rigour can be assessed. The strongest adherents of the constructivist perspective suggest that this is in itself a flawed project. However, since the mainstream of qualitative impact evaluation research in consultancy and related fields is primarily approached from an implicitly - if not explicitly – objectivist perspective, it is necessary for the mixed methods rationale of pragmatism to bridge the divide. We then focus on the framework of Lincoln and Guba as one that can accommodate the range of practical strategies needed and review the range of practical ways in which these can be pursued.

Finally, we consider impact evaluation underpinned by transformative-emancipatory values that the principle of authenticity engages with. Indeed, OPM itself seeks to engage in a progressive way concerned about its own social impact and how its consulting activities make a lasting positive impact in reducing poverty and disadvantage. It seeks to do this through partnership, independence, analytical rigour and a commitment to development carried out with core values of collaboration and mutual support; integrity and respect; trust, empowerment and accountability.³ The organisational context of commissioned research therefore is one in which it must seek to practice these commitments and values and we discuss how the authenticity principle can be used critically as an “artefact” in this regard (Guijt, 2015).

1. The philosophical basis of qualitative research and the ‘problem’ of rigour

As widely agreed in the literature the aim of qualitative research is to understand some aspects of social life in an in-depth way by answering the ‘what’, ‘why’ and ‘how’ questions by generating narrative or observational data to do this. In particular, it is applied to study people’s attitudes, behaviours, meanings and interpretations of events and phenomena and to provide rich description and analysis based on these perspectives. The methods employed are suitable to explore the depth of the issues under scrutiny and therefore the findings are not sought to be statistically generalised.

But it is necessary to go beyond methods as a contrast to quantitative research in understanding debates around qualitative research. It is useful to start with the primary philosophical basis of quantitative research. Quantitative research is usually understood as

³ See <http://www.opml.co.uk/about-us/our-values>

operating with a positivist paradigm⁴ of the world as observable and quantifiable. The philosophy underpinning quantitative research is primarily one in which there is an objective reality (ontological objectivism) driven by natural laws which science seeks to discover. From the point of view of investigation - epistemology - the researcher and the object of inquiry are separate (duality) therefore the researcher does not influence the way this truth is experienced (Guba and Lincoln, 1989). Truth is therefore seen as being obtained through scientific methods in which theories are proposed and hypotheses derived from them are tested with the potential of their falsification (Teddie and Tashakkori, 2009). Thus, the findings or knowledge of the truth is presented as generalisable and relevant across time and context.

The philosophy underlying most qualitative research on the other hand is that there is no single truth or reality and that phenomena depend on our perceptions and interpretations of them. Therefore the nature of reality is not unique or objectively verifiable but relative (ontologically⁵ relativist) and is created by our interpretations of it (epistemologically constructionist / interpretivist / or subjective). The "truth" presented is then a result of the interaction and relationship between the researcher and the researched rather than simply of the research design, as it is constructed by individuals under particular conditions and in a particular context, and consequently cannot be generalised (Sandelowski, 1986).

Although this is a standard way of defining the contrast between qualitative and quantitative methodologies, the field of qualitative research is a diverse and expanding field with a range of approaches and methods such that it involves further debate and disagreement within it. For example, Creswell (1998) describes five common 'traditions' within qualitative research: phenomenology, ethnography, case study, biography and grounded theory⁶. Seale (2002)

⁴ It is known by different names such as positivist/post-positivist research, empirical science, and post-positivism. Post-positivism challenges the positive thinking about the absolute truth of knowledge. It represents a deterministic philosophy in which causes determine outcomes and therefore it is concerned about identifying and assessing the causes that affect outcomes. It also reduces the general ideas into small sets of ideas which are manageable to test. Therefore, the objective truth which is out there in the world is researched based on observation and measurement. Consequently, post-positivism involves developing numeric measures for observations and studying the behaviour of individuals. (Creswell, 2009).

⁵ Ontology concerns what we consider to be the nature of the world - whether it is objective and separate from our role in it (objectives or realist ontology) or that reality is a result of the way we perceive it (constructivist). Epistemology is concerned about the nature of the relationship between the researcher and the object of study and hence about how we know or find out about the world. Realist ontology and positivist epistemology underpin quantitative methods and is therefore seen as a scientific paradigm in which the reality of the world can be investigated through measurement; qualitative methods are seen as being underpinned by a relativist ontology and a constructionist (aka interpretivist) epistemology which takes the view that how we view and understand the world i.e. our interpretations and meaning differ and affect the way we interact with it.

⁶ Phenomenological studies examine human experiences through the descriptions that are provided by the people involved. Ethnographic studies use methods such as participant observation over prolonged periods to see the world through the eyes of the people they are studying. In grounded theory studies, data are collected and analysed, and then a theory is developed that is *grounded* in the data. A process called constant comparison is used, in which data are constantly compared to data that have already been gathered. Pertinent concepts are identified and assigned codes. Historical studies concern the identification, location, evaluation, and synthesis of data from the past. Historical data should be subjected to both external and internal criticism. Case studies are in-depth examinations of people, groups of people, or institutions. Action research is a type of

expands the list by adding symbolic interactionism, ethnomethodology, semiotic analysis, feminist and Marxist perspectives. But the list is debatable, it is argued that the aims of research in these methodologies may be very different: whereas the aim of phenomenology is *to understand*, the aim of feminist methods is to bring *about a social change* (Roberts, 1981; Sandelowski, 1986). Moreover, how qualitative researchers apply constructionism and/or interpretivism depends on what traditions they belong to, and some qualitative traditions can be closer to positivism and post-positivism (Racher and Robinson, 2002) in understanding the world as having objective reality. So a straightforward dichotomous understanding of quantitative as being premised on positivist/post-positivist paradigm and qualitative on interpretivism is misleading. For example, the recent emergence of critical realism as a theoretical position adopts an ontologically realist perspective on the world but acknowledges the epistemological problem so accepting that this must be an interpretive exercise.

2. Rigorous debates in qualitative research

Despite the widespread use of qualitative research and its contribution to understanding, it faces some prominent criticism. The very features of it that are its strength such as the depth and closeness of the researcher to the research and to the data in particular, are argued to be its weakness due their being subjective and lacking rigour (Denzin and Lincoln 2000; Van Maanen, 1995; Smith and Deemer, 2000; Morse et al, 2002). This criticism was in particular promoted by the work of Archie Cochrane who introduced a model of research evaluation in medical sciences in his publication *Effectiveness and Efficiency (1972)*, in which qualitative research was seen as being unscientific and unfit for purpose. In contrast, he proposed randomized control trials as the gold standard for evidence⁷.

On the other hand qualitative researchers point out that quantitative research suffers from subjectivism too when it comes to choosing categories and variables. Some regard this as particularly problematic because ‘the probability is high that ...subjectivism survives without being thoroughly corrected during the study and ...it may affect the results...as the quantitative ...researcher does not get as close to those under study ...and therefore is less likely to be corrected by the study objects “talking back” (Flyvbjerg, 2011; p. 310). In particular positivist epistemology is seen as problematic because it tends to claim that there is only one true view so suppressing the variety of truth(s) and understandings of the world (Pretty, 1994). One could argue that truth is complex and is affected by many factors. So, there is ‘no way that the researcher can in any sense capture the literal truth of events’ (Mays and Pope, 1995; p. 109). As Pretty suggests ‘We can only get a human idea of what is

qualitative research that seeks action to improve practice and study the effects of the action that was taken. Participatory action research (PAR) is a special kind of action research in which there is collaboration between the study participants and the researcher in all steps of the study.

⁷ Later Sackett (1993) classified the *standards for quality of evidence* and weighted randomized trials as grade A evidence; whereas qualitative data and research were classified as “mere opinion” i.e. the Grade C evidence which is of the lowest level of quality (Morse, 2006).

in the world, and so science itself can only be a human picture of the world.’ (p. 40). As Rolfe (2002) argues truth is when it ‘rings true’, that is when it resonates with our own experiences’ (p. 91). So the plurality of truth does not prevent there being characteristics of it that are common across contexts and people involved. What is key is to understand those characteristics and interpret them in a non-intrusive and a bias-free language. This is challenging for qualitative research given the frequently profound submersion of the researcher with the data collection from research participants and context given its paradigmatic foundation. Nevertheless, these features seem to be its advantages when it comes to enabling it to generate descriptions and interpretations that ‘ring true’ and hence present a view of reality.

Nevertheless it is qualitative research that has been on the back foot from the perspective of wider policy discussion and the debate over evidence.⁸ The response from qualitative researchers has been a debate over whether and how rigour can in fact be implemented. As might be expected, given the above discussion of the philosophical stance of qualitative research, the question of whether rigour can or should be a criterion is itself highly contested. While some take the view that neglecting rigour could undermine the whole existence of this paradigm as a systematic science (Morse et al., 2002; Morse, 1999); others argue that it does not need to be discussed at all (Sparkes, 2001). Moreover, given the range of qualitative methodologies there is little agreement as to whether a set of criteria needs to be general to all the fields of qualitative research or distinctive to each one of them (Sandelowski, 1993; Maggs-Rapport, 2001).

The debate over criteria involves three positions (Denzin, 2011). First, the foundationalists – who take the view that all research can conform to shared criteria (Teddlie and Tashakkori, 2003); second, the quasi-foundationalists who argue that a specific set of criteria unique to qualitative research is needed (eg. Sandelowski, 1993; Lincoln and Guba, 1985); and finally the non-foundationalists who argue that it is understanding rather than prediction that is necessary and that “inquiry and its evidence is always political and moral” (Denzin, 2011) and therefore that criteria are not appropriate.

For those who consider a set of criteria as useful, one starting point is the quality criteria used in quantitative research. The central ones are internal validity, objectivity, reliability and generalization (for debates see Bryman, 1988). Internal validity refers to the suitability of a chosen method to measure what it is used to measure i.e. the degree to which the results are attributable to the independent variable; objectivity is achieved by positioning the researcher outside of the measured activity; reliability is when the same results can be achieved using the same methods; generalisation is ensured through using statistical tests to control other factors (Hamberg et al., 1994).

The quality principles that have been developed such as those of Agar (1986), Guba (1981), Kirk and Miller (1986), Leininger (1985) and Brink (1991) are summarised in table 1.

⁸ The Cabinet Office guidelines on Quality in Qualitative Evaluation (Spencer, Ritchie, Lewis, & Dillon, 2003) could be seen as reflecting this concern.

Table 1. Examples of principles of rigour in qualitative research

Authors	Principles of quality and rigour
Agar (1986)	Credibility, accuracy of representation, and authority of the writer
Guba (1981)	Truth value, applicability, consistency, and neutrality
Kirk and Miller (1986)	Consistency of results, stability over time, similarity within a given time period
Brink (1991)	Stability, consistency, equivalence ⁹
Lincoln and Guba (1985)	Credibility, confirmability, dependability, transferability, authenticity

What becomes evident from this table is that there is considerable overlap in the nature and language of the criteria being presented as well as some mapping onto the quantitative criteria. Agar's (1986) credibility, accuracy of representation, and authority of the writer are versions of validity and reliability and are close in meaning to the two typologies of Guba (1981) and Lincoln and Guba (1985). Guba's (1981) model of four aspects of trustworthiness that are relevant to both quantitative and qualitative studies are very similar to the one by Lincoln and Guba (1985) except the principle of authenticity. The latter is based on the former and only differs by the language used. An exception is Kirk and Miller (1986) whose criteria argue for consistency of results and stability over time which seems contrary to a qualitative view that interpretations and understandings may change over time.

In the literature in this field it is Lincoln and Guba's (1985) criteria that have become the most used and cited with their language of "trustworthiness" rather than rigour also being adopted. It is laid out in a little more detail in Table 2. The comparative advantages of this typology are seen as its relevance across qualitative paradigms (Sale and Brazil, 2004) and having the advantage of parsimony (Bryman et al., 2008). The typology is the most developed and therefore is worth applying for the conceptual and analytical clarity and depth that it can offer.

Despite these advantages, it is unsurprising that, given the paradigmatic disagreements laid out above, there are of course criticisms. The first is that in order to argue for absolute trustworthiness it is necessary to take a positivistic view (Pretty, 1994). This is because truth is valid for a certain time and a certain context and is not static. Therefore, in order to ensure trustworthiness it is suggested that the criteria should include the process of inquiry, informing the reader about what has and has not happened before, during and after the inquiry (*ibid*).

In addition, Morse et al. (2002) argue that confirmability is not suitable for phenomenology, feminist research and critical theory 'in which the investigator's experience becomes part of data, and which perceive reality as dynamic and changing' (p. 19). Smith (1993 in Tobin and

⁹ Equivalence involves for example asking different kinds of questions within the single interview or questionnaire in order to establish the equivalence of the data elicited regardless of the form of question, or multiple observers comparing their observations of the same event.

Begley, 2004) argues that some procedures such as member or dependability checks¹⁰ are not appropriate to the philosophical idea of multiple realities that qualitative research is based on and therefore, target its fundamental epistemological and ontological premise.

Table 2. Lincoln and Guba’s principles for evaluating trustworthiness in qualitative research

Qualitative research	Questions that underpin the principles of qualitative research (Pretty, 1994; p. 42)	Quantitative research concepts
Credibility	How can we be confident about the ‘truth’ of the findings?	Internal validity
Confirmability	How can we be certain that the findings have been determined by the subjects and contexts of the inquiry, rather than the biases, motivations and perspectives of the investigator?	Objectivity
Dependability	Would the findings be repeated if the inquiry were replicated with the same (or similar) subjects in the same or similar context?	Reliability
Transferability	Can we apply these findings to other context or with other groups of people?	Generalisation
Authenticity	Have people been changed by the process? To what extent did the investigation prompt action?	

Source: own table based on Pretty (1994) and Lincoln and Guba (1985)

As indicated above, the non-foundationalists reject Lincoln and Guba’s criteria out of hand. As Barbour (2001) and Morse et al. (2002) claim, any technical solutions reducing qualitative research to a list of methodological procedures does not ensure rigour. Barbour (2001) argues that such procedures can only be successful if they are conducted within a proper understanding of qualitative research as creative and flexible, which makes it problematic to reduce it to what is seen as a box-ticking exercise. Moreover, it is argued that the procedural aspects must come second when evaluating the quality of qualitative research after ‘moral principles and ethical standards’ of how researchers relate to research participants (Rallis et al., 2007). This ethical argument is paramount since qualitative research is based on relationships which emerge between the researcher and data, the researcher and research participants, data and research participants and the wider circle of readers. Therefore, they argue that the trustworthiness of a qualitative study needs to be judged on how ethically it

¹⁰ This is to ensure that the researcher and the informant are viewing the data consistently. The research shares the conclusions with informants and ask for their feedback about the accuracy of the content

is done with relation to research participants, other stakeholders and the scientific community.

It was partly in response to such criticisms that Lincoln and Guba added the fifth criterion of authenticity in order to more clearly distinguish these principles from positivistic ones (Lincoln, 1990 cited in Pretty, 1994). The authenticity criterion has five elements (Guba & Lincoln, 1989) and was developed in accordance with a constructivist tradition i.e. knowledge is relative to time and place, subjective meanings matter and that ‘...‘truth is a matter of consensus among informed and sophisticated constructors...’ (Lay and Papadopoulos, 2007; p. 495). Table 3 gives the summary of the five criteria of authenticity including their short definitions.

Table 3. Authenticity principle and its criteria with explanations

Authenticity principle (Guba and Lincoln, 1989)	Nelson, et al. 2003 cited in (Wilson and Clissett, 2011)	Definition (Guba and Lincoln, 1989))
Fairness	All viewpoints are represented even-handedly	The extent to which all competing constructions have been accessed, exposed, and taken into account in the evaluation report, that is, in the negotiated emergent construction.
Ontological authenticity	Participants understand their situation in more informed ways as a result of participation in the research	The extent to which individual constructions (including those of the evaluator) have become more informed and sophisticated
Educative authenticity	Participants understand the situations of others in more informed ways as a result of participation in the research	The extent to which individuals (including the evaluator) have become more understanding (even if not more tolerant) of the constructions of others.
Catalytic authenticity	Participants have a greater insight into actions that they might take to change their situation as a result of participation in the research	The extent to which action (clarifying the focus at issue, moving to eliminate or ameliorate problems, sharpening values) is stimulated and facilitated by the evaluation.
Tactical authenticity	Participants feel empowered and enabled to act as a result of participation in the research	The extent to which individuals are empowered to take the action that the evaluation implies or proposes

Source: own table based on Wilson and Clissett, 2011 and Guba and Lincoln, 1989

The basic assumptions behind the authenticity principle are that people and groups have different value systems which affect their constructions. Therefore, the researcher should

make sure that different constructions emerge to allow conflicting constructions and value structures to express themselves (fairness). Moreover, people's consciousness i.e. 'conscious experiencing of the world' (Guba and Lincoln cited in Schwandt, Lincoln, and Guba 2007, p. 22) develops and their constructions change as people gain experience and interact with others. Participating in the research is also part of gaining experience and changing one's constructions as the research participants also develop a better understanding of the phenomenon under scrutiny (ontological authenticity). But research participants not only know more about the phenomenon, they also become educated about other research participants' values and constructions. Hence they will appreciate different opinions, judgments, and actions (educative authenticity). Once the diversity of constructions and values are achieved and recognised, and their understanding is enriched, as a result of participation in the research, the research participants would not only have developed consciousness but also be empowered to act upon the phenomenon (catalytic authentication). For all these components of the authenticity principle to be achieved, negotiation is essential. It is the negotiation for achieving consensus that drives the authenticity principle. The principles therefore respond to the concerns of the non-foundationalists and critical theorists that research in itself has political and moral dynamics whose purpose may be to affect the world and not simply to represent it in objective ways. They affect the world by changing understanding with the potential to lead to action.

In the context of impact evaluation this raises a particular issue regarding the rights of those affected by a policy or intervention to be engaged in the research and to engage with other policy stakeholders in negotiation over findings and implications. Indeed, as Guba (1987; p. 39) suggests for the case of evaluation then authentic evaluation will '*...essentially be about the process of negotiation with and among stakeholders, and that the product of evaluation is not '...a series of conclusions and recommendations, but as an agenda for further negotiation'* (emphasis in original). The principle therefore particularly responds to the recent concerns raised regarding the politics of evidence in development and fears that the dominance of experimentalist approaches has become increasingly hegemonic in the practices and processes of the allocation of aid resources (Eyben & Guijt, 2015).

In order to move forward, we propose adopting Lincoln and Guba's framework and the next sections discuss how its principles of trustworthiness can be operationalised in practice. The first four are discussed and then we address the fifth principle of authenticity in a separate section as it speaks to the core issue of ethics and power dynamics that arise in commissioned impact evaluation.

3. Principles of trustworthiness in practice

This section discusses in greater depth Lincoln and Guba's framework and draws into each of these areas specific methodological strategies for demonstrating qualitative rigour and

achieving these principles (Guba and Lincoln, 1981; Lincoln and Guba, 1985; Guba and Lincoln, 1982)¹¹.

Credibility

Credibility is argued to be similar to internal validity in quantitative research. Research is seen as credible when the researcher has confidence in the truth of the findings with regard to the subjects of research and the context where it was conducted. As there is understood to be a plurality of truths within the qualitative paradigm, then credibility is about 'representing those multiple realities revealed by informants as adequately as possible' and testing those realities against various groups of people (Krefting, 1991; p. 215), so as to draw out common themes, actions and issues. The description of realities is important, so that people who share similar experiences should recognise them and be able to relate them to their own realities (Sandelowski, 1986). Moreover, Sandelowski (1986) suggests that the credibility of qualitative work increases when the researcher discusses his/her own behaviour alongside the behaviour of other 'subjects of research' and reflects on their closeness to the data in order to recognise how she might be unduly giving weight to particular positions or interpretations over others when analysing and interpreting it. So the researcher needs to stay close to data but at a reflexive distance from it. This needs to result in such an interpretation of data which is recognisable for both the research participants and scientific community.

Relevant strategies:

Table 4 summarises the strategies to achieve the principle of credibility, their purposes and the stage of the research when they are most likely to be implemented. We discuss them in turn.

In order to enhance credibility, the research can apply a range of strategies. To start with, it needs to focus on the process of selecting research participants and explain how they were selected. Sampling of all possible situations relevant to the research such as different social settings, times of day, week, and season may be important in order to demonstrate possible variations in experiences and cases (*time sampling* (Krefting, 1991)). Of course, all of these are not likely to be possible, so consideration and justification of those likely to be most relevant is required.

Purposeful sampling is considered to be more suitable for qualitative research which allows the researcher to have information-rich cases for the study that will provide relevant data to the study question. It is important to choose the right type of purposeful sampling among

¹¹ The original list of strategies for the trustworthiness principle (Lincoln and Guba (1985) includes following strategies: prolonged engagement, persistent observation, triangulation, peer debriefing, negative case analysis referential adequacy, member check, thick description, stepwise replication, inquiry audit (of the process and product), audit trial and reflexive journal. The rest of the strategies adopted in this note come from other sources as Baxter and Eyles (1997) and Krefting (1991), while others are suggested by the authors (systematic coding and data reduction and ethics).

the many options available (Patton, 1990 in Baxter and Eyles, 1997), a relevant sample size and justify those choices. Indeed, sample size may not need to be predetermined, as it may depend on the data collection process itself, for example, whether saturation has been reached. Moreover, qualitative methods can have quite a different understanding of representativeness in which it is assumed that ‘any subject belonging to a specified group is considered to represent that group’ (Sandelowski, 1986) rather than the statistical representativity of a sample in quantitative methods. This demonstrates how important it is to apply purposeful sampling and to give detailed information about the setting and research participants as well as those who did not participate.

Table 4. A principle of credibility and strategies to achieve it

Principle	Stages of implementation	Strategies to achieve the principle	Purposes of strategies
Credibility (Authentic representation of experience)	Design	Purposeful sampling	To involve as many experiences as possible until thematic saturation is reached and to include negative cases
		Time sampling	To sample all possible situations, different social settings; times of day, week, and season; and interactions among different social groupings to see how research participants interact with one another in a number of contexts at different times. This strategy emphasizes the importance of the environment in which the data are collected in seeking to establish credibility.
	Data collection	Reflexivity (field journal)	To assess the influence of the researcher's own background, perceptions, and interests on the research process
		Disciplined subjectivity	To monitor the researcher's own influence on developing thematic constructions and be aware of their own biases (e.g. keep an analytical journal, recording thoughts in NVivo through memoing, talking to another

Principle	Stages of implementation	Strategies to achieve the principle	Purposes of strategies
			colleague doing the same analysis and discussing emerging codes)
		Prolonged engagement in the field	To identify reappearing patterns/themes and to build rapport
		Persistent observation	To observe phenomenon under scrutiny in the natural context
	Data analysis	Peer debriefing	To discuss data and interpretation with colleagues
		Triangulation (methods, sources, researchers, theories)	To cross check data and interpretation
		Negative case review	To constantly revise the hypothesis against all the texts until it accounts for all of the cases
		Referential adequacy	To verify '...the constructs developed through an interpretation of the bulk of the data by the subsequent analysis of a selection of previously collected data which has been archived (Baxter and Eyles, 1997; p.515)
	Writing up	Structural coherence	To ensure coherent structure of the story line and that there is no unexplained inconsistencies between the data and their interpretations
		Member checking	To constantly check the data, analysis and interpretation with informants
		Using quotes	To check the interpretation against verbatim accounts

Source: Author's own table adopted from Lincoln and Guba (1985 mentioned in Baxter and Eyles, 1997) and Krefting (1991)

The credibility of the research also depends to a great extent on the skills of the researcher in observing, interviewing, analysing and writing up. Of course, these tend to be particularly

constrained by time and report length.¹² So the researcher is the eyes and ears of the reader who should be given a detailed description of the fieldwork site, background of research participants, and do so reflexively without mixing her own outlook of events with those of research participants i.e. *disciplined subjectivity* (Erickson, 1973 cited in Baxter and Eyles, 1997).

Moreover, Tong et al. (2007) suggest that the researcher should inform the reader about her '...identity, credentials, occupation, gender, experience and training' (p. 351) so the reader can assess how a personal background of the researcher can affect the study. This is common in academic studies but the credentials of the consultancy team are rarely the subject of reflection in commissioned reports – a practice that could change. It requires honesty from the researcher to be truthful to the reader and research participants. Documenting reflexivity and reflectivity of the researcher is useful to analyse how the researcher can affect the research participants, how research participants can affect the researcher and how the experiences, feelings and background of the latter can affect her observations (Patton, 2003). It will also help the researcher realise how her relationships with the settings and people change over time and how it can potentially affect the research. The study design should clarify the degree of participation of the researcher in the study, nature of collaboration between her and her co-researcher, if any, and discuss possible tensions and biases. Consequently, *field journals* are important not only to work on the biases of the researcher but also to inform the reader about how decisions were made and how the researcher came to certain claims, which will be discussed below. Being reflexive in doing research is part of being honest and ethically mature that requires that the researcher is in constant reflection over research methods, methodology, limitations, own biases etc as the study goes on. Reflexivity is about methodological, observational, analytical thinking but also about personal critical analysis of own personal biases and limitations.

Prolonged engagement in the field and *persistent observation* are argued to be helpful in order to learn about the locality, build rapport with research participants and ensure quality data. Repeated interviews or group discussions are also important for building rapport and therefore need to be reported to the reader. The researcher should clarify and justify the nature and degree of her engagement with research participants; the duration of observation and engagement and the focus of observation. Reflexivity on how all these can affect the data is important to avoid the risk of 'going native' (Vidich et al. cited in Baxter and Eyles, 1997). The risk is seen as about getting too close to research participants and data which could result in the situation when the researcher jeopardises her objectivity and analytical perspective. This can result in bias of becoming '...coopted, going native, swallowing the agreed-upon or taken-for-granted version of local events.' (Miles and Huberman, 1984; p. 233). So the idea of objectivity is retained. In the context of commissioned work, unless there are long standing links with the people involved, the risk may be less of biases towards "going native" but of "friendship biases" towards the organisations being assessed (H. White & Phillips, 2012, p. 25). For non-foundationalists and those using critical theory, the argument is that the values which drive the researcher involve taking sides, and this is with those who are seen to be marginalised or oppressed by

¹² A critical issue if all the recommendations of (Spencer et al., 2003) for example are to be covered.

the structures of society and hence whose voices should be heard and 'objectivity' is understood to be part of the problematic. This aspect is dealt with in greater depth in the authenticity principle below.

Triangulation of methods, sources, theories and researchers are particularly helpful in order to cross validate analysis and findings and increase plausibility of findings. It also helps to look out for a new source of data i.e. an informant, a new group of informants, a new setting or event (Miles and Huberman, 1984). In reporting, the researcher can also use *quotes* which represent people's views in their own words to triangulate sources. Here it is important not to cherry pick the quotes and misuse them to show the researcher's own message not the true data. It addresses the question of how interpretations are arrived at as well as the boundaries between obtaining the insight and validating it. Qualitative studies are often criticised for being unclear when reporting claims about the extent to which a particular claim is representative and common among the research participants but the risk is of responding to the idea that only quantification of the responses is useful. While understanding the numerical ordering of phenomena in the data may be useful it is also about raising the level of analysis to accommodate variation and to explain it. In this case, a systematic coding scheme is important in order to be able to generate claims which are grounded in and generated by data, not by the researcher. This means finding ways to structure codes that use all the data and leaves audit trails of how coding is built up and managed in analysing findings. It also helps to differentiate *negative cases* (i.e. cases which are outliers and do not well with other data collected) which can reassure the reader about the accuracy and depth of analysis and also help the researcher to develop more inclusive theories (Seale and Silverman, 1997). As Miles and Huberman (1984) suggest negative cases are friends of the researcher. They '...test and strengthen the basic finding. It not only tests the generality of the finding, but protects against self-selecting biases (p. 237). In addition, the coded data of emerging themes can be verified against the data already archived to achieve *referential adequacy* (i.e. coding of new data and comparison with old data).

Furthermore, coding can be repeated when another researcher does coding of the same data in order to check for the accuracy of the original analysis. This is probably a basic quality control necessary in most studies - certainly larger-scale ones. Should any discrepancies occur in coding, then both researchers need to discuss those cases. *Peer* involvement can also extend to organising *debriefings* when the lead researcher is able to get expert-led feedback on data analysis and coding. However, this strategy might generate 'unresolvable' disagreements (Risteen Hasselkus, 1991 cited in Baxter and Eyles, 1997). All these methods will then contribute to the *structural coherence* (Krefting, 1991) of the analysis and interpretation which follows a clear line of argumentation and claims. This should demonstrate that interpretations follow from the qualitative data and elucidated meanings (Patton, 2003). As Miles and Huberman (1984) suggest a well written report can win the reader with vivid writing that brings the material alive, though even plausible and coherent accounts on their own are not a solution as they too can be biased.

To check whether the interpretation of data by the researcher reflects what the researcher was actually told, the latter can involve a *member check* technique which involves returning

to participants in order to validate the finding and analysis with them. However, in doing so one needs to bear in mind that the interpretation must be made in more analytical format and language, analysed through the prism of certain theories and concepts. Morse et al. (2002) argue that the risk of member (respondent) check is that the researcher may need to adhere to a more descriptive level of analysis and interpretation and therefore 'invalidate the work of the researcher and keep the level of analysis inappropriately close to the data' (p.16). A particular problem is when respondents are not organised or related to by the project in any way that makes access to groups of respondents to undertake such exercises more logistically straightforward. For example, if respondents are users of particular services which have focal points such as schools or health centres, then research feedback activities can be organised to a degree around these access points, where such services are more disparate (e.g. cash transfers) then this can be harder. Use of community level structures may be possible and the research can also be designed with these in mind as means through which member checking can be carried out.

In all of these areas the question of "how much" arises. Trade-offs of time, people and resources are inevitable and there is of course no easy rule of thumb for how such trade-offs are to be resolved. The perhaps unhelpful response to this is "as much as possible" but the critical orientation is an underlying concern for ethical integrity and an ideological awareness of the politics at hand and the range of potential biases that can arise.

Confirmability

Confirmability is about ensuring that the research process and findings are not biased, hence it refers to both the researcher *and* the interpretations (Baxter and Eyles, 1997). This raises the issue of distance from and influence of the researcher on data collection and analysis when the researcher is actively involved with research participants and constantly engages with the data. This closeness of the researcher to the object of the study is argued to be a unique feature of qualitative data, so it is challenging for the researcher as a 'positioned subject' (Rosaldo, 1989) to consciously reflect about her own acts and background in relation to the data. Confirmability is achieved when the interpretation of data is neutral and free from the researcher's personal bias.

Relevant strategies:

To achieve confirmability, the researcher has the same strategies used for credibility (Table 4) with some additional ones. The confirmability and credibility principles are related to being *ethical* in the sense of seeking to truthfully reflect what others are saying and to understand the basis of this which is a result of relationships with participants that embody respect and integrity and seek to understand their perspective and its causes. While certain procedures help to ensure this level of ethics through the suggested strategies, there is a certain intrinsic morality that the researcher should possess which goes beyond documenting, recording and reporting. Ethics in this sense is twofold. First, it is concerned with traditional ethical issues of confidentiality i.e. power relations of the research relationship with informed consent and ability to exit. But, second, it is about moral ethics,

which is harder to trace as an inside-out process happening within the researcher herself. This recognises that since the data that is collected from a respondent is a view of the world this must be treated with trust and respect. Both aspects of ethics are a focus of this paper, which is included in the list of proposed strategies for achieving rigour in qualitative research, though the latter part of ethics are a challenge to ensure through procedural measures. Although researchers apply for ethical clearance and meet the requirements procedurally, the moral ethics are still out of the scope. Having groups of researchers helps with ‘ethical triangulation’ as individual researchers can mutually check and balance. As ‘interpreters of human behaviour’ (Schwandt, Lincoln, and Guba; 2007), qualitative researchers find themselves in the midst of different kinds of relationships, as said earlier, where the task is to make sense of and understand research participants, be confident of the research findings and be ethical throughout.

Table 5. A principle of confirmability and strategies to achieve it

Principle	Stages of implementation	Strategies to achieve the principle	Purposes of strategies
Confirmability (Extent to which biases, motivations, interests or perspectives of the inquirer influence interpretations) True findings free of biases	Data collection	Audit trail	To follow through the progression of events and decisions made as well as the product, data, findings, interpretations to arrive at comparable findings
		Field notes, reports, journal/notebook	To keep notebooks, journals or a log of everyday events during the fieldwork as well as to reflect on the researcher’s thoughts, feelings, assumptions, etc.
	Data analysis	Ethics	Ethics in truthfully interpreting data, applying respectful attitudes to research participants and being reflexive of own personal biases
		Systematic coding and data reduction	To show how data was interpreted, reduced and how the main themes/claims emerged as a result

Source: Author’s own table adopted from Lincoln and Guba (1985 mentioned in Baxter and Eyles, 1997), Krefting (1991) and own strategy added (i.e. ethics)

In general, the researcher needs to provide the reader with as full and detailed a picture of how decisions were made and the progression of the study as possible so that the reader can understand what happened, when it happened and why from the very beginning to the very end of the study. Again in the context of commissioned reports which may be of limited

length, the question is how to bring the most important and pertinent aspects of the process into the account.

This is the so called *audit trail* which is basically about describing, explaining and justifying certain actions, decisions, and events. Guba and Lincoln are not specific about the form of the materials suitable for the audit trail, but they suggest that they can include some form of a diary or log. Indeed whatever is used must be suitable for the exercise and easily understandable for the second reader.¹³ *Field notes, reports and journals* would be particularly important to trace a certain change in the design, any problems faced and any decisions made during the process. These materials are close to the ones under the credibility principle in the way that they are meant to capture the researcher's thoughts, feelings, ideas, and hypotheses generated by her contact with the research participants and data, as well as questions, problems, and frustrations with regard to the overall research process (Krefting, 1991). As a result of this written reflection, the researcher can become aware of her own biases and assumptions. The challenge here would be a systematic ordering of fieldwork materials to avoid creating confusion and wasting time.

The confirmability principle can also include implementing systematic coding and sharing of *data reduction materials* (quantitative summaries, condensed notes, working hypotheses, thematic categories, interpretations, inferences), which demonstrate how certain themes, events, actions and codes emerged and how data was consequently reduced to arrive at the main findings. They can include code books and memos to record thoughts, questions, relationships between interviews, specific themes that emerged, and references to concepts and themes in the data. These materials are conceptual in intent and do not simply report but conceptualise links between data and concepts. The process of coding and identification of themes are argued to be closely linked with the quality of any qualitative research and therefore they must be made explicit (Tong et al., 2007). This strategy can be used for the principle of credibility as well, as this will be helpful in satisfying the criterion of authentic representation of the experience under scrutiny. The challenges with this strategy could be generated by the way that data coding can be done by various members of the team or those involved which can cause inconsistency in analysis. Moreover, the researcher needs to be careful with regard to revealing confidential data and personal information of research participants when displaying raw materials. The overall idea behind this principle is to demonstrate that the researcher remains true and ethical towards the research process and data that she works with. Table 5 gives the summary of the additional strategies. To extend the strategies suggested by Lincoln and Guba (1985), the paper suggests the strategies of *systematic coding* and *data reduction* to enhance the 'truth' and replication of the study.

Dependability

Dependability is to ensure consistent data collection without unnecessary variations to ensure repeatability of the research process. This is about being able to trace sources that

¹³ Of course such an approach is equally applicable to quantitative research. The question is now gaining increasing attention as requirements for data archiving and secondary use develop further (Irwin & Winterton, 2012)

the data comes from and about documenting the data, methods and decisions made during the fieldwork. So consistency in the entire research process is key for achieving dependability. This criterion is suggested to be closely linked to credibility and be equally important for qualitative research. Lincoln and Guba (1985) themselves argue that establishing dependability ensures credibility. Although the strategies used to prove these two principles can be similar, their meanings are different: credibility is about ‘accurate representation of experience while dependability focuses attention on the researcher-as-instrument and the degree to which interpretation is made in a consistent manner.’ (Baxter and Eyles, 1997; p. 517).

Table 6. A principle of dependability and strategies to achieve it

Principle	Stages of implementation	Strategies to achieve the principle	Purposes of strategies
Dependability (Minimization of idiosyncrasies in interpretation; Variability tracked to identifiable sources) Repeatability	Data collection	Low inference descriptors, mechanically recorded data	To check the level of ‘agreement’ between data and its interpretation through checking the field notes, quotations, and other narratives.
		Audit check	To document the methods of data gathering, analysis, and interpretation to make it repeatable and auditable for another researcher to understand the decision trail, the process and findings
		Triangulation, inquiry audit	Checking between researchers’ to check the process of the research in terms of relevant decision making done along the way as well as introducing alternative perspectives in data analysis prior to finalizing the set of theoretical constructs.
		Thick description of methods	To generate detailed descriptions of methods including their purposes, limitations, order of using them, matching them with research questions etc.
	Data analysis	A stepwise replication technique	Teams or researchers work separately on the data and compare results
		A code-recode procedure	A researcher codes a segment of data and checks its coding after a while (eg.1-2 weeks)

	Writing up	Peer examination, Multiple researchers	Colleagues check the research plan and implementation
--	------------	---	---

Source: Author's own table adopted from Lincoln and Guba (1985 mentioned in Baxter and Eyles, 1997) and Krefting (1991)

Relevant strategies:

To achieve dependability, there are several strategies available which are shared by the two previous principles (Table 5). Documenting the entire process of research design and implementation, as well as decisions made, will be crucial in order to give the reader full information. *Thick description of methods* (Krefting, 1991) used for data collection is also important. This should be to the extent which will enable her to conduct a similar study and arrive at similar the same findings if she wishes to. This should also allow for an *audit* check. Any field materials covering basic research information such as daily schedules and logistics, a methods log capturing decisions about methods and their rationale are important for auditability. Field notes and audio records are examples of *low-inference* descriptors (field notes and audio recordings which include verbatim accounts and narratives of behaviours, activities and events). These sources as well as peer debriefings as methods of triangulation of sources and methods are helpful in order to demonstrate the agreement between the data and findings, the plausibility of accounts and methods (Baxter and Eyles, 1997). The greater involvement of researchers i.e. *peer examination* in discussions and re/coding (*a code –recode procedure* (one researcher reads and codes a text and does so again 1-2 weeks later) *and stepwise replication* (two or more researchers analyse the same data and compare) help to ensure the objectivity and plausibility of findings. It brings definitional clarity and is a good reliability check (Miles and Huberman, 1984). Peer examination is about checking how decisions were made with regard to sampling, data collections and analysis methods and may be part of internal or external quality assurance overview. This strategy will be well complemented by the coding and data reduction materials as discussed in relation to confirmability. In general, what is important for coding data is some conceptual and structural order which will help to relate the codes as well as distinguish them from one another in a conceptually meaningful way (Miles and Huberman, 1984).

The difference between the audit and peer examination is, according to Baxter and Eyles (1997), that the former takes place all the way along during the study (sampling and selection of research participants, methods used and data analysis), whereas the latter takes place towards its end.

Transferability

This is similar to external validity or generalisability but is not reached through random sampling and probabilistic reasoning. Transferability is when the research descriptions and findings are sufficient to draw similarities with another context. To achieve this criterion, the researcher needs to provide detailed descriptive information. These details should enable the reader to judge the applicability of findings to her own settings. However, as discussed,

it is hard to generalise meanings constructed from a small group of respondents and this is usually not the goal of qualitative research.¹⁴ Nevertheless, the researcher has the (ethical) responsibility to describe the findings in the way that allows transferability and let the reader decide whether those meanings are transferrable to her context (Baxter and Eyles, 1997).

Table 7. A principle of transferability and strategies to achieve it

Principle	Stages of implementation	Strategies to achieve the principle	Purposes of strategies
Transferability (Fit within contexts outside the study situation)	Design	Purposeful sampling	To search for as many different cases as possible until new themes stop emerging
	Writing up	Thick description	To give detailed information about the research participants, contexts and settings
		Comparison of sample to demographic data	To compare the characteristics of the research participants against the demographic information available on the group under scrutiny.
Applicability			

Source: Author's own table adopted from Lincoln and Guba (1985 mentioned in Baxter and Eyles, 1997) and Krefting (1991)

Relevant strategies:

In doing so, the researcher needs to give a *thick description* of the setting under scrutiny to allow the reader to feel as if she knows the setting well. This will help her to decide whether the meanings and experiences of the target group are common to another group of people and if the findings can be transferable to a context outside the setting being studied.

Moreover, the relevant *purposeful (or purposive) sampling* is critical to ensure that various groups of people representing the wider community are covered by the study (Patton, 2003). Krefting (1991) suggests that a further *comparison of the characteristics of the informants to the demographic information* available on that group and filling gaps if certain demographic characteristics are not represented is another way of achieving transferability. Table 7 summarises the principle and strategies used.

4. Authenticity principle

¹⁴ The distinction can be made between generalising statistically and generalising analytically or to theory (see Yin 2003).

Finally we discuss Lincoln and Guba’s principle of authenticity established to respond to criticisms that their trustworthiness criteria did not adequately address transformative- emancipatory research approaches which take a more critical and constructive stance. Guba and Lincoln (1989) suggest certain strategies to achieve it which are presented in Table 8.

Table 8. Authenticity principle and strategies to achieve it

Criteria of authenticity principle	Strategies to achieve the criteria of authenticity
Fairness	Observing contradictions and tensions among constructions of stakeholders
	Negotiating these contradictions and tensions and establishing an ‘agenda for subsequent action’ (Guba & Lincoln, 1989, p. 246).
Ontological Authenticity (world around)	Analysing statements (e.g., testimonies) provided by the research participants and leaving audit trails that document the participant’s growth in consciousness and understanding of the world, as well as the growth in the researcher’s own ‘progressive subjectivity’ (p. 248).
	Observing the participants in action for collecting evidence of an elevation in the participant’s consciousness level.
Educative authenticity (other people’s constructions)	Analysing statements (e.g., testimonies) provided by the research participants and leaving audit trails that document the participant’s growth in understanding of and appreciation for [but not necessarily agreement of] the constructions of <i>others</i> outside their stakeholding group are enhanced” Guba & Lincoln, 1989, p. 248, italics in original).
Catalytic authenticity	Obtaining testimony from all participants and stakeholders regarding their interest in and willingness to turn their increased understanding into action.
	Obtaining testimony regarding the joint actions of participants who have come to resolutions stemming from negotiations of tensions invoked by contesting and contradictory constructions of the stakeholders’ (Onwuegbuzie, Leech, & Collins, 2008, p. 9).
	Assessing the extent to which the actions that stemmed from the increased understandings that emerged during the course of the study actually occurred
Tactical authenticity	Obtaining testimony from all the participants and stakeholders regarding whether/ how the emergent feelings of empowerment evolved and manifested themselves
	Systematically following up within a predetermined time frame to assess which participants and/or groups acted on their increased feelings of empowerment, and what actions came to the fore
	Researcher and participants (jointly) assess the degree of empowerment that evolved during the study.

Source: Authors' own table based on (Collins et al. 2013)

Shwandt et al argue (Schwandt, Lincoln, & Guba, 2007) that authenticity is an extension of the trustworthiness criteria because it enables questions to be asked about how interpretations are made and how this process has evolved. Indeed the authenticity principle recognises that inquiry and understanding are a process of learning, changing, negotiating and ultimately acting. Qualitative research affects the consciousness of the researcher and research participants to the extent that it can change the way they understand the truth(s). For the authenticity principle, the evaluation research should recognise and promote the diversity of the value systems and people's constructions of the world which change constantly as people interact with one another.¹⁵ Therefore the principle takes into account the process of forming interpretations from the value-point of respondents, their voice, their diversity, their positioning and empowerment towards other respondents and themselves. In this process, research participants as well as researchers learn to respect each other, to see the issues from different perspectives and consequently, negotiate the construction of truth. Based on a constructivist epistemology, the main drive for the authenticity principle is negotiation i.e. negotiation of constructs and values. Guba and Lincoln (2001) suggest that such an approach to the evaluation can deal with the conventional problems with evaluations which, among other weaknesses, favours the point of view of the funder, which is disempowering, unfair to other stakeholders, and fails to accommodate value-pluralism, and is overcommitted to the positivist paradigm of inquiry.

The authenticity principle gives considerable weight to the ability of the research process to incorporate the values and constructions of participants as well as empowering them to improve their situations. Such intentions are of course not without problems when it comes to practice. In the development field there has been a long discussion of participation in both development practice (Cooke & Kothari, 2001; Hickey & Mohan, 2004) and research (Chambers, 2004; Holland, 2013) which have highlighted the core problems of structural power dynamics in attempts to involve participants in empowering ways.

Apart from this, the practical context of increasing swathes of commissioned evaluation and research stands largely outside a participatory paradigm such that it faces real practical challenges regarding the amount of time spent in the field together with research participants and the ability to develop effective relationships with them such that these objectives can be achieved. While one-off interviews or group discussions and workshops for disseminating research findings can engage new thinking and help inform, they clearly fall short of this vision and rarely focus on evaluation participants and stakeholders becoming enlightened about each other's values and constructions. This makes it less likely that the research would be able to generate far reaching impact on research participants to fulfil the authenticity principle. More realistic would be to assume that it is rather the researchers or evaluators who would be more likely to be subject to any change during the

¹⁵ This resonates also with the concern to bring understanding of what wellbeing means for different people into the evaluation of development policy, recognising that what is valued requires wider deliberative debate along the lines of Sen's concern for the expansion of freedoms (Sen, 1999); (McGregor & Sumner, 2010); (S. White & Blackmore, 2016).

course of the study. However, it is hard to foresee that any change would be sustainable enough to lead to action – particularly given the usual distance at which stakeholder’s (especially commissioners) operate from this process.

Moreover, the authenticity principle promotes recognition of the diversity of values and constructions, but it is unclear how exactly consensus is to be reached or how disagreements are dealt with. There will always be power relations among different research participants and other groups of research participants. Indeed, qualitative research in international development evaluation rarely seeks to act as a solution provider, consensus builder or producer of tangible action. It is issues such as these that approaches to evaluation ranging from empowerment (Fetterman & Wandersman, 2012) to participatory (Coghlan & King, 2005) and democratic evaluation traditions (House & Howe, 2000; MacDonald & Kushner, 2005) endeavour to manage through specific ways of conducting them.

However, the point here is to ask how far can these aspects be taken in the context of commissioned assessments that are constrained by bureaucratic requirements? The point is to find ways that consultants can exercise their agency within these processes to raise the bar.

At the moment, the strategies suggested in Table 8 rather represent a range of ideal approaches for achieving such ethically complex ideas as participation and empowerment. In this sense, while the authenticity principle speaks to an important normative concern that research be undertaken in a transformative-emancipatory way, the principle needs to bring into view the wider research context and practical organisational framework for implementing its strategies. It is these wider dynamics that have been at the core of recent concern about the dominance of RCTs in the evidence-based movement and how it meets visions of development based in rights-based approaches and the transformation of power relations (Eyben et al., 2015).

Indeed, engaging with an assessment to open up the space for debate and learning is potentially dangerous terrain for participants and raises the potential for extremely unsatisfactory or even dangerous processes of engagement to result. This may be at a minimum by wasting the time of participants in processes of engagement whose potential for change is virtually non-existent or extremely marginal. Moreover, in the context of power relations surrounding the implementation of policies or programmes, the result may be even more problematic with the potential for negative consequences if findings lead to decisions and action that detrimentally rebounds on them. This could be the withdrawal of resources or worse, the potential for victimisation if particular actors feel aggrieved at information revealed during evaluation processes.

These issues raise in a specific way the questions of what have been called “macro-ethical issues” which go beyond the “micro-ethical issues” of consent, anonymity and so on and tip the areas of concern more systematically towards how evaluation relates to society and social processes more broadly; whose interests it serves; its role in deliberation and the use

of evidence and in public sector management and wider public debate (Barnett & Munslow, 2014, p. 13). Indeed it also raises the question of how development is done and the contribution of evaluation to ethical development (Barnett & Munslow, 2014, p. 13). While in principle therefore, authenticity incorporates some aspects of concern around the politics and power dynamics of impact evaluation, strategies for implementation are obviously a much more difficult question. This discussion therefore highlights the need for a wider ethical framework within which the authenticity principle must operate at two levels: the individual study and commissioned organisation, and the wider institutional context of evaluation practice.

For the commissioned researchers, no guidelines and procedural measures can fully mould their actions and they are largely driven by their own values and ethics within the context of the practical and structural constraints within which they are operating. Ethics were added as a feature of the confirmability principle in Table 5 and are particularly important as researchers and evaluators (especially commissioners) tend to have resources and power concentrated in their hands such that the scope of negotiations that can take place in the course of a study is likely to be extremely constrained by its design from the outset. Therefore, the issue of ethics is one that extends well beyond procedural ethical requirements such as consent. From the view of the authenticity principle it means that the researcher must be fair in allowing a diversity of constructions and values to be expressed, she must actively seek to become more informed of her own as well as other people's constructions through reflexivity and encourage the right actions towards the phenomenon by generating trustworthy and ethical qualitative findings.

Going further to create space for the practice of negotiation as part of the practical strategies can obviously be of value under the right circumstances and act to legitimate space for diversity of views within the evaluation in the context of problematic top-down power relations and in principle requires space for negotiation with commissioners themselves. How then can this be done?

As Guijt (2015) argues, it is necessary to “play the game to change the rules” in part by using the “artefacts” of evaluation themselves. That is, the processes, mechanisms and tools that are deployed in this field. It is necessary to understand them, “reframe” and “intelligently adopt them” (p200). We therefore propose that commissioned researchers deploy checklists for rigour in this way. That they include the authenticity principle as a means of raising the game with commissioners as to what is appropriate and ethical in the conduct of evaluation research and seek to make more space for engagement, negotiation and action. While we recognise that a list of strategies is not ideal if it is solely treated in the spirit of box ticking – including them can create space for those researchers and commissioned organisations who do indeed wish to add value and undertake qualitative impact evaluation in a more progressive way. So that while it is not possible to up-end power relations at a single turn, it is certainly the case that demonstrating a concern for these ethical dimensions and their practice can help confirm and give space for reflexivity over the intention of the exercise. This in turn calls for the practices of commissioning organisations to respond in terms of their own guidelines and checklists.

To achieve this, such checklists have to be underpinned by a particular mind-set and ethics. This mind-set is an orientation underpinned by a concern for ethical practice at both the “micro” and “macro” levels. It is this orientation that underpins the question of “how much?” of any of these strategies is appropriate. There is always space for more to be done and it is constant reflexivity on these issues that is required. Commissioned researchers can deploy such guidelines and checklists to constantly develop their approaches.

Table 9. A principle of authenticity and initial strategies to achieve it

Principle	Stages of implementation	Strategies to achieve the principle	Purposes of strategies
Authenticity (negotiation over constructions through diversity of views)	Design	Purposeful sampling	To have a wide ‘representation’ of people who have ‘relevant’ relationship with the phenomenon
		Enabling space for voice and negotiation of potential directions of the research	To include the views of research participants on the research design e.g. to identify hard to access respondents, to be responsive to the local mechanisms of consensus building and negotiation
	Data collection and analysis	Making space for a wide group of research participants to participate in data collection and analysis	To draw out diverse views and experiences of the phenomenon among research participants and negotiate the articulation of findings with them
	Dissemination and follow up	Offering means through which research participants can be engaged with commissioners as stakeholders in making recommendations and taking follow up actions	To make sure that research participants have their ‘say’ in recommending any changes to the intervention and taking part in their follow up. Any recommendations should be responsive to the diversity of values and opinions of research participants
	Review	Re-visiting commissioners, stakeholders and respondents to assess what has changed.	To establish how the evaluation has been used and examine what changes in understanding and action were achieved by all parties involved.

As a starting point then the orientation is to find ways (see Table 9) for: (i) making space for voice and negotiation by research participants as stakeholders within the design and conduct of the research through meetings and workshops and ensuring these involve excellent facilitation skills; (ii) finding ways to make space for participation in data collection and analysis; (iii) offer means through which representatives of these views can be engaged with commissioners in later stages of the evaluation. All of which must involve ethical considerations regarding the potential for harm. In this context, the minimum standard is always to have considered the issues against the checklist and explained why the approach taken is appropriate within the resources available and how it avoids potential harm. Table 9 presents a starting checklist for orienting research to use the authenticity principle.

The final point in Table 9, addresses the question of what ultimately changed both in terms of understanding of the issues and the actions of commissioners and stakeholders. This effectively needs a further review. In many programme areas this is in fact possible since there are frequently follow on programmes whose planning stages could deploy a review approach, but of course it is rarely done. Commissioners usually start again and do not wish to revisit past programmes or projects, especially if these were not perceived to be 'successful'.

Moving beyond this checklist means experimenting with approaches to expand the scope for authenticity and engaging with commissioners over the processes and resources involved. The argument here is that checklists for rigour are one means for doing this and raising these wider issues with commissioners in and around processes of commissioning as well as overall standards and evaluation guidelines themselves.

5. Conclusion

This paper has been motivated by the recent turn towards qualitative impact evaluation methodologies to address the concern for rigour. The paper has therefore reviewed the literature on rigour in qualitative research, starting with a review of the philosophical basis of qualitative research in order to explain the disagreements within the debate between qualitative methodologists themselves over the feasibility of rigour.

Notwithstanding, considerable efforts have been made to develop standardised criteria for evaluating rigour and the context of impact evaluation research is sufficiently practical to require them. This paper adopts Lincoln and Guba's framework for trustworthiness for the reason that it is the most developed and used and is appropriate to the context of much impact evaluation research - particularly that in development consultancy - which operates within a more objectivist ontology and post-positivist epistemology where qualitative methodologies and evidence are still assessed through criteria of representation and objectivity. It then used this framework as an organising basis for presenting practical strategies to demonstrate its application in more depth. It is important that researchers

implement strategies across the four main areas of credibility, confirmability, dependability and transferability, rather than focus on one and also justify their choice of strategies.

However we go beyond the standard trustworthiness framework to include the principle of authenticity. This is embedded in concerns of the constructivist camp for research that is transformative-emancipatory in orientation, and fits well with concerns in impact evaluation about the politics of evidence and the wider macro-ethical concerns as to the role of evaluation in society more broadly. In order to move forward practically we propose that adding authenticity to checklists of rigour is therefore a strategic and useful way to deploy the “artefacts” of evaluation protocols and guidelines to demonstrate approaches to rigour to commissioners. At the same time there is the potential to use this approach to challenge the commissioners and institutional frameworks surrounding evaluation itself.

With the concern that checklists themselves are inadequate, we argue that using such a principle requires an orientation to better practice with these ethical concerns at their source. It is this orientation that underpins the question of “how much?” of any of these strategies is appropriate. There is always space for more to be done and it is constant reflexivity on these issues that is required. The minimum standard is always to have considered the issues against the checklist and explained why the approach taken is appropriate within the resources available and the avoidance of potential harm. Moving beyond this, the principle of authenticity demands the creation of space for learning among all stakeholders, for interaction and negotiation. This means experimenting with approaches to expand the scope for authenticity and engaging with commissioners over the processes and resources involved.

References:

- Agar, M. (1986). *Speaking of ethnography*. Beverly Hills, CA: Sage.
- Barbour, R. (2001) Checklists for improving rigour in qualitative research: a case of the tail wagging the dog? *BMJ*. 2001; 322: 1115–1117
- Barnett, C., & Munslow, T. (2014). Workshop Report: Framing Ethics in Impact Evaluation: Where are we? Which route should we take? *Evidence Report No 98*. Institute for Development Studies, Sussex www.ids.ac.uk.
- Baxter, J. and J. Eyles, (1997) Evaluating qualitative research in social geography: establishing 'rigour' in interview analysis. *Transactions of the Institute of British Geographers*, 22(4): 505–525.
- Brink, P. (1991) Issues of reliability and validity. In: Morse, J. (ed) *Qualitative nursing research: a contemporary dialogue*. Sage, London, pp. 164–186
- Bryman, A. (1988) *Quantity and Quality in Social Research*, London: Unwin Hyman
- Bryman, A. Becker, S. and J. Sempik (2008) Quality Criteria for Quantitative, Qualitative and Mixed Methods Research: A View from Social Policy' *Int. J. Social Research Methodology*, 11 (4): 261–276
- Camfield, L., & Duvendack, M. (2014). Impact Evaluation - Are We 'Off the Gold Standard'? [Article]. *European Journal of Development Research*, 26(1), 1-11. doi: 10.1057/ejdr.2013.42
- Chambers, R. (2004). Ideas for development: reflecting forwards. IDS Working Paper 238.
- Coghlan, A. T., & King, J. A. (2005). Participatory Evaluation. In S. Mathison (Ed.), *Encyclopedia of Evaluation* (pp. 292-297). Thousand Oaks, CA: Sage Publications, Inc.
- Cooke, B., & Kothari, U. (2001). *Participation: the new tyranny?* London: Zed Press.
- Creswell, J. W. (2009) *Research Design: Qualitative, quantitative, and mixed methods approaches*. London: SAGE 3rd Edition.
- Chambers, Robert. "Ideas for Development: Reflecting Forwards." IDS Working Paper 238, 2004.
- Cochrane A. L. (1972) *Effectiveness and Efficiency: Random Reflections on Health Services*, The Nuffield Provincial Hospitals Trust. (Reprinted in 1989 in association with the BMJ, Reprinted in 1999 for Nuffield Trust by. London: The Royal Society of Medicine Press; 1972. ISBN 1-85315-394-X
- Cooke, Bill, and Uma Kothari. *Participation: The New Tyranny?* Edited by Bill Cooke and Uma Kothari London: Zed Press, 2001.
- Creswell, J. W. (1998) *Qualitative Inquiry and Research Design: Choosing Among Five Traditions*. Thousand Oaks, CA: Sage Publications.
- Denzin, Norman K. (2011) "The Politics of Evidence." In *The Sage Handbook of Qualitative Research*, edited by Norman K. Denzin and Yvonna S. Lincoln, 645-58. London: SAGE, 2011.
- Denzin, N.K. and Lincoln, Y.S. (2011) Introduction. The discipline and practice of qualitative research. In Denzin, K. N. and Lincoln, S. Y. (eds) *The sage handbook of qualitative research*. USA: Sage Publications. pp. 1-19.

Denzin, N.K. and Lincoln, Y.S. (2000) Introduction: the discipline and practice of qualitative research. In Denzin N.K. and Y.S. Lincoln (eds) *Handbook of Qualitative Research*. Sage, Thousand Oaks, CA, pp. 1–28.

Eyben, R., & Guijt, I. (2015). Introduction. In R. Eyben, I. Guijt, C. Roche & C. Shutt (Eds.), *The politics of evidence and results in international development : playing the game to change the rules?* Rugby: Practical Action Publishing.

Eyben, Rosalind, Irene Guijt, Chris Roche, and Cathy Shutt (2015) *The Politics of Evidence and Results in International Development: Playing the Game to Change the Rules?* Rugby: Practical Action Publishing,

Fetterman, D. M., & Wandersman, A. (2012). *Empowerment Evaluation Principles in Practice*: Guilford Publications.

Flyvbjerg, B. (2011) Case study. In Denzin, K. N. and Lincoln, S. Y. (eds) *The sage handbook of qualitative research*. USA: Sage Publications. pp. 301-316

Glennerster, R., & Takavarasha, K. (2013). *Running randomized evaluations a practical guide*. Princeton: Princeton University Press.

Groves, Leslie Christine, and Rachel Barbara Hinton. *Inclusive Aid Changing Power and Relationships in International Development*. Edited by Leslie Christine Groves and Rachel Barbara Hinton London: Earthscan, 2004.

Guba, E. G. (1981). Criteria for assessing the trustworthiness of naturalistic inquiries. *Educational Communication and Technology Journal*, 29 (2), 75-91.

Guba, E. G., and Lincoln, Y. S. (1981). *Effective evaluation: Improving the usefulness of evaluation results through responsive and naturalistic approaches*. San Francisco, CA: Jossey-Bass.

Guba, E. G., and Lincoln, Y. S. (1982). Epistemological and methodological bases of naturalistic inquiry. *Educational Communication and Technology Journal* 30 (4): 233-252

Guba, E. G. and Lincoln, Y. S. (1989) *Fourth Generation Evaluation*, Newbury Park, CA: Sage Publications.

Guba, E.G. and Lincoln, Y.S. (1994) Competing paradigms in qualitative research. In. Denzin, N.K and Y.S. Lincoln (eds.), *Handbook of qualitative research*. pp. 105-117

Hamberg K, Johansson E, Lindgren G and Westman G. (1994) Scientific rigour in qualitative research—examples from a study of women's health in family practice. *Family Practice* 11: 176-181.

Hickey, Samuel, and Giles Mohan. *Participation: From Tyranny to Transformation? Exploring New Approaches to Participation in Development*. Edited by Samuel Hickey and Giles Mohan London: London: Zed, 2004.

Holland, Jeremy. (2013) *Who Counts?: The Power of Participatory Statistics*. Edited by Jeremy Holland Rugby: Rugby: Practical Action, 2013.

House, E. R., & Howe, K. R. (2000). Deliberative democratic evaluation. *New Directions for Evaluation*, 2000(85), 3-12. doi: 10.1002/ev.1157

- Irwin, S., & Winterton, M. (2012). Qualitative Secondary Analysis and Social Explanation. *Sociological Research Online*, 17(2), 4
- Kirk, J. and Miller, M. (1986) *Reliability and validity in qualitative research*. Beverly Hills, CA: Sage.
- Krefting, L. (1991) Rigor in qualitative research: The assessment of trustworthiness. *American Journal of Occupational Therapy*, 45: 214-222.
- Leininger, M. M. (1985). Nature, rationale and importance of qualitative research methods in nursing. In Leininger, M. M. (ed), *Qualitative research methods in nursing*. New York: Grune & Stratton, pp. 1-28
- Lincoln, Y. (2005). Institutional review boards and methodological conservatism. The challenge to and from phenomenological paradigms. In Denzin, N. and Y. Lincoln (eds.), *The Sage handbook of qualitative research* Thousand Oaks, CA: SAGE. 3rd ed., pp. 165-181
- Lincoln, Y. and E. Guba (1985) *Naturalistic Inquiry*. Beverly Hills: Sage
- MacDonald, B., & Kushner, S. (2005). Democratic Evaluation. In S. Mathison (Ed.), *Encyclopedia of Evaluation* (pp. 110-114). Thousand Oaks, CA: Sage Publications, Inc.
- Mays N, and Pope C. (1995) Qualitative Research: Rigour and qualitative research. *BMJ*. 1995; 311(6997):109-12
- Maggs-Rapport F. (2001) Best research practice: in pursuit of methodological rigor. *Journal of Advanced Nursing*, 35(3): 373–383.
- McGregor, A., & Sumner, A. (2010). Beyond Business as Usual: What Might 3D Wellbeing Contribute to MDG Momentum? *IDS Bulletin*, 41(1), 104-112. doi: 10.1111/j.1759 5436.2010.00111.x
- Miles, B.M. and Huberman, A. M. (1984) *Qualitative data analysis: a sourcebook of new methods*. Sage Publications: Beverly Hills
- Morse J.M., Barrett M., Mayan M., Olson K. and Spiers J. (2002) Verification strategies for establishing reliability and validity in qualitative research. *International Journal of Qualitative Methods* 1(2), Article 2.
- Morse J.M. (1999) Myth 93: Reliability and validity are not relevant to qualitative inquiry. *Qualitative Inquiry* 9: 717–718
- Norris, N. (2015). Democratic evaluation: The work and ideas of Barry MacDonald (Vol. 21, pp. 135-142).
- Patton, M. Q. (2003). *Qualitative evaluation checklist*. Retrieved March 24, 2014, from http://www.wmich.edu/evalctr/archive_checklists/gec.pdf
- Picciotto, Robert. "Is Impact Evaluation Evaluation?" *European Journal of Development Research* 26, no. 1 (2014): 31-38.
- Pretty, J. N. (1994) Alternative systems of inquiry for sustainable agriculture. *IDS Bulletin*, 25(2): 37-48. IDS: University of Sussex

- Racher F. and Robinson S. (2002) Are phenomenology and postpositivism strange bedfellows? *Western Journal of Nursing Research* 25: 464–481
- Rallis, S., Rossman, B and Gajda, R. (2007) Trustworthiness in evaluation practice: An emphasis on the relational. *Evaluation and Program Planning* 30: 404–409
- Roberts H. (1981) *Doing Feminist Research*. Routledge, London.
- Rolfe G (2002) 'A Lie that Helps us See the Truth': Research, truth and fiction in the helping professions. *Reflect Pract* 3:89–100.
- Rolfe G. (2006) Validity, trustworthiness and rigour: quality and the idea of qualitative research. *Journal of Advanced Nursing* 53(3): 304–310.
- Rosaldo, R. 1993 (1989). Grief and a headhunter's rage. *In: Culture and truth*. London & New York, Routledge: 1-21.
- Sale, J. and K. Brazil (2004) A Strategy to Identify Critical Appraisal Criteria for Primary Mixed-Method Studies *Quality & Quantity* 38: 351–36
- Sandelowski M. (1993) Rigor or rigor mortis: the problem of rigor in qualitative research revisited. *Advances in Nursing Science*, 16:1–8.
- Sandelowski, M. (1986) Problem of rigor in qualitative research. *Advancing in Nursing Science* 8(3): 27-37.
- Schwandt, T. A., Lincoln, Y. S., & Guba, E. G. (2007). Judging interpretations: But is it rigorous? trustworthiness and authenticity in naturalistic evaluation. *New Directions for Evaluation*, 2007(114), 11-25. doi: 10.1002/ev.223
- Seale, C. (1999) Quality in Qualitative Research. *Qualitative Inquiry* 5(4): 465-478.
- Seale, C. (2002) 'Quality Issues in Qualitative Inquiry'. In Atkinson, P. and S. Delamont (eds) *SAGE Qualitative Research Methods*. Thousand Oaks: SAGE. pp. 98-275.
- Seale, C. and D. Silverman (1997) Ensuring rigour in qualitative research. *European Journal of Public Health* 7: 379-384
- Sen, A. (1999). *Development as Freedom*. Oxford: Oxford University Press.
- Sparkes, C. (2001). Myth 94: Qualitative health researchers will agree about validity. *Qualitative Health Research*, 11(4): 538-552.
- Spencer, L., Ritchie, J., Lewis, J., & Dillon, L. (2003). Quality in Qualitative Evaluation: a framework for assessing research evidence: UK Cabinet Office.
- Smith J. & Deemer D. (2000) The problem of criteria in the age of relativism. In Denzin N.K. and Lincoln Y.S. (eds) *Handbook of Qualitative Research*. Sage, Thousand Oaks CA, pp. 877–896
- Stern, Elliot, Nicoletta Stame, John Mayne, Kim Forss, Rick Davies, and Barbara Befani. "Broadening the Range of Designs and Methods for Impact Evaluations" In *Report of a study commissioned by the Department for International Development* London: DFID Working Paper No. 38, 2012.

Stern, E. (2015). Editorial. *Evaluation*, 21(2), 131-134. doi: 10.1177/1356389015580670

Teddie C. and A. Tashakkori (2009) *Foundations of Mixed Methods Research: Integrating Quantitative and Qualitative Approaches in the Social and Behavioral Sciences*. Thousand Oaks, CA: SAGE

Tobin, G and Begley C (2004) Methodological rigour within a qualitative framework. *J Adv Nurs* 48(4): 388–96

Tong, A., Sainsbury, P. and 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

Van Maanen J. (ed.) (1995) *Representation in Ethnography*. Thousand Oaks, CA: Sage

Wainwright D. (1997) Can sociological research be qualitative, critical and valid? *The Qualitative Report* 3(2), 2

White, H., & Phillips, D. (2012). Addressing attribution of cause and effect in small n impact evaluations: towards an integrated framework: 3ie <http://www.3ieimpact.org/>.

White, S., & Blackmore, C. (Eds.). (2016). *Cultures of wellbeing method, place, policy*. Basingstoke: Basingstoke : Palgrave Macmillan.

Yin, R 2003, *Case study research: Design and methods*, Applied Social Research Methods Series, Volume 5, Sage, Thousand Oaks, CA.

The Centre for Development Studies (CDS), University of Bath

The Centre for Development Studies aims to contribute to combating global poverty and inequality through primary research into the practical realities of global poverty; and, critical engagement with development practice and policy making. In December 2011, the Bath Papers in International Development (BPD) working paper series was merged with the Wellbeing in Developing Countries (WeD) Working Paper Series, which has now been discontinued. The new series, Bath Papers in International Development and Well-Being continues the numbering of the BPD series.

Bath Papers in International Development and Well-Being (BPIDW)

Bath Papers in International Development and Well-Being publishes research and policy analysis by scholars and development practitioners in the CDS and its wider network. Submissions to the series are encouraged; submissions should be directed to the Series Editor, and will be subject to a blind peer review process prior to acceptance.

Series Editors: Susan Johnson and Althea-Maria Rivas Website: <http://www.bath.ac.uk/cds/publications>
Email: s.z.johnson@bath.ac.uk or a.rivas@bath.ac.uk

2016

No. 44 Financial Capability for Wellbeing: An alternative perspective from the Capability Approach
Author(s): Silvia Storchi and Susan Johnson

2015

No. 43 Relational Wellbeing: A Theoretical and Operational Approach
Author(s): Sarah C. White

No.42 Humanitarian NGOs: Dealing with authoritarian regimes
Author(s): Oliver Walton

No.41 'Upliftment', friends and finance: Everyday concepts and practices of resource exchange
Underpinning mobile money adoption in Kenya
Author(s): Susan Johnson and Froukje Krijtenburg

No.40 Towards a plural history of microfinance
Author(s): James Copestake, Mateo Cabello, Ruth Goodwin-Groen,
Robin Gravestijn, Julie Humberstone, Susan Johnson, Max Nino-Zarazua, Matthew Titus

No.39 Theological resources and the transformation of unjust structures: The case of Argentine informal economy workers
Author(s): Séverine Deneulin, Centre for Development Studies, University of Bath

No.38 Coloniality and Indigenous Territorial Rights in the Peruvian Amazon: A Critique of the Prior Consultation Law

Author(s): Roger Merino Acuña, Centre for Development Studies, University of Bath

No.37 Micro-foundations of producer power in Colombia and the Philippines: towards a political understanding of rents

Author(s): Charmaine G. Ramos, Centre for Development Studies, University of Bath

2014

No.36 “Whither development studies?” Reflections on its relationship with social policy

Author(s): James Copestake, Centre for Development Studies, University of Bath

No.35 Assessing Rural Transformations: Piloting a Qualitative Impact Protocol in Malawi and Ethiopia

Author(s): James Copestake and Fiona Remnant, Centre for Development Studies, University of Bath

No.34 “We don’t have this is mine and this is his”: Managing money and the character of conjugality in Kenya

Author(s): Susan Johnson, Centre for Development Studies, University of Bath

No.33 Can civil society be free of the natural state? Applying North to Bangladesh Author(s): Geof Wood, Centre for Development Studies, University of Bath

No.32 Creating more just cities: The right to the city and the capability approach combined

Author(s): Séverine Deneulin, Centre for Development Studies, University of Bath

No.31 Engaging with children living amidst political violence: Towards an integrated approach to protection

Author(s): Jason Hart, Centre for Development Studies, University of Bath

No.30 Competing visions of financial inclusion in Kenya: The rift revealed by mobile money transfer

Author(s): Susan Johnson, Centre for Development Studies, University of Bath

No.29 Can’t buy me happiness: How voluntary simplicity contributes to subjective wellbeing Author(s):

Nadine van Dijk, United Nations Research Institute for Social Development, Switzerland

2013

No.28 Challenge funds in international development

Author(s): Anne-Marie O’Riordan, James Copestake, Juliette Seibold & David Smith, Triple line Consulting and University of Bath

No.27 From the Idea of Justice to the Idea of Injustice: Mixing the Ideal, Non-ideal and Dynamic Conceptions of Injustice

Author(s): Oscar Garza, Centre for Development Studies, University of Bath

No.26 Understanding Policy and Programming on Sex-Selection in Tamil Nadu: Ethnographic and Sociological Reflections

Author(s): Shahid Perwez, Centre for Development Studies, University of Bath

No.25 Beyond the grumpy rich man and the happy peasant: Subjective perspectives on wellbeing and food security in rural India

Author(s): Sarah C. White, Centre for Development Studies, University of Bath

No.24 Behind the aid brand: Distinguishing between development finance and assistance Author(s):

James Copestake, Centre for Development Studies, University of Bath

No.23 The political economy of financial inclusion: Tailoring policy to fit amid the tensions of market development

Author(s): Susan Johnson, Centre for Development Studies, University of Bath; and Richard Williams, Oxford Policy Management, Oxford

No.22 'Everything is Politics': Understanding the political dimensions of NGO legitimacy in conflict-affected and transitional contexts

Author(s): Oliver Walton, Centre for Development Studies, University of Bath

No.21 Informality and Corruption

Author(s): Ajit Mishra, University of Bath; and Ranjan Ray, Monash University, Australia

No.20 The speed of the snail: The Zapatistas' autonomy *de facto* and the Mexican State Author(s): Ana

C. Dinerstein, Centre for Development Studies, University of Bath

No.19 Patriarchal investments: Marriage, dowry and economic change in rural Bangladesh Author(s):

Sarah C White, Centre for Development Studies, University of Bath

2012

No.18 Political economy analysis, aid effectiveness and the art of development management Author(s): James Copestake and Richard Williams, Centre for Development Studies, University of Bath

No.17 Justice and deliberation about the good life: The contribution of Latin American *buen vivir* social movements to the idea of justice

Author(s): Séverine Deneulin, Centre for Development Studies, University of Bath

No.16 Limits of participatory democracy: Social movements and the displacement of disagreement in South America; *and,*

Author(s): Juan Pablo Ferrero, Department of Social and Policy Sciences, University of Bath

No.15 Human rights trade-offs in a context of systemic unfreedom: The case of the smelter town of La Oroya, Peru

Author(s): Areli Valencia, University of Victoria, Canada

No.14 Inclusive financial markets: Is transformation under way in Kenya?

Author(s): Susan Johnson, Centre for Development Studies, University of Bath; and Steven Arnold, Department of Economics, University of Bath

No.13 Beyond subjective well-being: A critical review of the Stiglitz Report approach to subjective perspectives on quality of life

Author(s): Sarah C. White, Centre for Development Studies, University of Bath, Stanley O. Gaines, Department of Psychology, Brunel University; and Shreya Jha, Centre for Development Studies, University of Bath

2011

No.12 The role of social resources in securing life and livelihood in rural Afghanistan Author(s): Paula Kantor, International Centre for Research on Women; and Adam Pain, Afghanistan Research and Evaluation Unit

2010

No.11 Côte d'Ivoire's elusive quest for peace
Author(s): Arnim Langer, Centre for Peace Research and Strategic Studies, University of Leuven

No.10 Does modernity still matter? Evaluating the concept of multiple modernities and its alternatives
Author(s): Elsje Fourie, University of Trento

No.9 The political economy of secessionism: Inequality, identity and the state Author(s):
Graham K. Brown, Centre for Development Studies, University of Bath

No.8 Hope movements: Social movements in the pursuit of development
Author(s): Séverine Deneulin, Centre for Development Studies, University of Bath; and Ana C. Dinerstein, Centre for Development Studies, University of Bath

No.7 The role of informal groups in financial markets: Evidence from Kenya Author(s): Susan Johnson, Centre for Development Studies, University of Bath, Markku Malkamäki, Decentralised Financial Services Project, Kenya; and Max Niño-Zarazua, Independent Consultant, Mexico City

2009

No.6 'Get to the bridge and I will help you cross': Merit, personal connections, and money as routes to success in Nigerian higher education
Author(s): Chris Willott, Centre for Development Studies, University of Bath

No.5 The politics of financial policy making in a developing country: The Financial Institutions Act in Thailand
Author(s): Arissara Painmanakul, Centre for Development Studies, University of Bath

No.4 Contesting the boundaries of religion in social mobilization
Graham K. Brown, Centre for Development Studies, University of Bath,
Author(s): Séverine Deneulin and Joseph Devine, Centre for Development Studies, University of Bath

No.3 Legible pluralism: The politics of ethnic and religious identification in Malaysia

Author(s): Graham K. Brown, Centre for Development Studies, University of Bath

No.2 Financial inclusion, vulnerability, and mental models: From physical access to effective use of financial services in a low-income area of Mexico City

Author(s): Max Niño-Zarazua, Independent Consultant, Mexico City; and James G. Copestake, Centre for Development Studies, University of Bath

No. 1. Financial access and exclusion in Kenya and Uganda

Author(s): Susan Johnson, Centre for Development Studies, University of Bath; and Max Niño-Zarazua, Independent Consultant, Mexico City