

Validity in Qualitative Evaluation: Linking Purposes, Paradigms, and Perspectives

International Journal of Qualitative Methods
2015: 1–8
© The Author(s) 2015
Reprints and permissions:
sagepub.com/journalsPermissions.nav
DOI: 10.1177/1609406915621406
ijqm.sagepub.com



Vasco Lub^{1,2}

Abstract

This article provides a discussion on the question of validity in qualitative evaluation. Although validity in qualitative inquiry has been widely reflected upon in the methodological literature (and is still often subject of debate), the link with evaluation research is underexplored. Elaborating on epistemological and theoretical conceptualizations by Guba and Lincoln and Creswell and Miller, the article explores aspects of validity of qualitative research with the explicit objective of connecting them with aspects of evaluation in social policy. It argues that different purposes of qualitative evaluations can be linked with different scientific paradigms and perspectives, thus transcending unproductive paradigmatic divisions as well as providing a flexible yet rigorous validity framework for researchers and reviewers of qualitative evaluations.

Keywords

Validity, evaluation research, validity checklist, reliability, policy research

Introduction

Since the days of ethnographic pioneers such as the Anthropologist Franz Boas and members of the Chicago School of Urban Sociology, a vast literature has developed on the procedures and underlying philosophies of qualitative research. Focusing on the natural behavior of people and their perceptions of the social world (Denzin and Lincoln 2005; Yin, 1994), one relatively recent development is the increasing use of qualitative methods and information for evaluation purposes in social policy and health care. The health sector in particular has seen a surge in approaches and writings on *evidence-based* procedures and evaluation research that involve or require inclusion of qualitative methods (see, e.g., Pope & Mays, 2006). Some of the reasons for this are an increased understanding and acknowledgment of the limits of experiment or questionnaire-based quantitative research on “what works” (see, e.g., Oakley, Strange, Bonell, Allen, & Stephenson, 2006; Roe & Lysaker, 2012; Lub, 2014), a growing demand for ethical considerations in evaluations (see, e.g., Parahoo, 2014), insight into clients’ and patients’ well-being (see, e.g., Haber, Carlson, & Braga, 2014), and the need for a more thorough understanding of how experimentally determined *evidence-based* interventions connect to people’s emotions, culture, experiences, and habits (see, e.g., Gibbs, Jewkes, Sikweyiya, & Willan, 2015; Lohan, Aventin, Maguire, Clarke, Linden & McDaid, 2014). In 2010, the U.K.’s Department of Health and the National Health Service Institute (NHS)—

traditionally a bastion of quantitative effect research—commissioned King’s College London to undertake research into the subject: What Matters To Patients? Developing the Evidence Base for Measuring and Improving Patient Experience. For a large part, the report draws on the input from key stakeholders who attended workshops and sets out arguments for how the NHS can improve services and patients’ experience of health care (see Robert & Cornwell, 2011).

However, the increased importance given to qualitative information in the *evidence-based* paradigm in health care and social policy requires a more precise conceptualization of validity criteria that goes beyond just academic reflection. After all, one can argue that policy verdicts that are based on qualitative information must be legitimized by valid research, just as quantitative effect research is subject to validity standards. Yet how to determine the validity (or “truth value”; Lincoln & Guba, 1985, p. 290) of such investigations is a difficult question. Although validity in qualitative research has been widely reflected upon in the methodological literature

¹ Bureau for Social Argumentation, Rotterdam, the Netherlands

² Department of Sociology, Erasmus University Rotterdam, Rotterdam, the Netherlands

Corresponding Author:

Vasco Lub, Department of Sociology, Erasmus University Rotterdam, Burgemeester Oudlaan 50, 3062 PA Rotterdam, the Netherlands.
Email: vascolub@gmail.com



Creative Commons CC-BY-NC: This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 3.0 License (<http://www.creativecommons.org/licenses/by-nc/3.0/>) which permits non-commercial use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access page (<https://us.sagepub.com/en-us/nam/open-access-at-sage>).

(and is still often subject of debate), the link with evaluation research is underexplored.

In this article, I will explore aspects of validity of qualitative research with the explicit objective of connecting them with aspects of evaluation. Given the nature of the evaluator–stakeholder relationship in evaluations (see Rossi, Lipsey, & Freeman, 2004), and the methodological properties of qualitative research in particular, qualitative information in evaluation can have three different purposes. First, it can contribute to or focus on the *instrumental effectiveness* of the program itself. Does it work? What are its main working components? Can, for example, intended effects of a support program for pregnant teenagers—such as encouraging them to remain in school—indeed be observed in the field? What are additional effects? Second, qualitative research can focus on the *meaning* of the policy or program for clients, target groups, and practitioners. How do the teenagers experience the support program? How do the trainers shape it? Third, qualitative evaluation can follow an *emancipatory approach* in which the evaluation itself can take either of the two aforementioned perspectives, but the information derived from the research simultaneously and deliberately aims to empower or educate those involved in the program (see e.g., the many forms of participatory action research). Staying with our hypothetical example, research questions in such an evaluation could read: Did the teenage pregnant mothers themselves benefit from the information gathered in the evaluation? How did it empower them and generate solutions to practical problems (Meyer, 2006)?

I will argue that the different purposes of qualitative evaluation in social policy and health care can be linked with different scientific paradigms and perspectives and aligned with relevant validity procedures. Such a conceptualization transcends unproductive paradigmatic divisions and provides a framework for researchers and reviewers of qualitative evaluations. The framework presented *can* serve as a checklist for qualitative evaluations. But its main value is as a theoretical reference point. It aims to sensitize the reader's own paradigmatic assumptions about evaluation research and the application of qualitative information within those evaluations. It is, however, with a more general discussion of qualitative inquiry and validity that my exploration must begin.

Qualitative Research and the Question of Validity

Since roughly the 1970s increasing criticism of the reliability and objectivity of qualitative research has resulted in a growing interest in establishing more rigorous criteria and methodological standards. This attention has somewhat shifted from standards for the implementation of the study by the researcher to verification strategies for evaluating the credibility of qualitative findings by external reviewers (Morse, Barrett, Mayan, Olson, & Spiers, 2002). Validity is a key concept in this discussion. In the positivistic, rational tradition of science methodology, “validity” can be defined as the degree to which the indicators or variables of a research concept are made

measurable, accurately represent that concept. Does, for example, a response scale that measures interactions with members of other ethnic groups indeed refer to intercultural tolerance? Obviously, this rational definition of validity does not work well in qualitative naturalistic research—which does not focus on variables on interval or ratio level. As a result, in the qualitative methodological literature, “validity” has been labeled with alternative terms such as authenticity, adequacy, plausibility, and neutrality (see, e.g., Lincoln & Guba, 1985; Maxwell, 1996; Merriam, 1998). Nevertheless, within the academic community, the idea seems to be dominant that qualitative researchers must demonstrate in one way or another that their research results are valid. Several authors have therefore sought to develop specific research procedures and criteria aimed at increasing the validity of qualitative outcomes.

Probably, the most influential is the work of Guba and Lincoln (see Guba & Lincoln, 1981; Lincoln & Guba, 1985). Guba and Lincoln were one of the first to develop specific criteria for qualitative research. They started from the premise that although all research must possess high truth value, the properties of knowledge within the “rational” (or quantitative) paradigm is different from the properties of knowledge within the “naturalistic” (or qualitative) paradigm (as cited in Morse, Barrett, Mayan, Olson & Spiers, 2002). According to Guba and Lincoln, each paradigm requires specific criteria to determine the veracity of the research. Within the rational paradigm, criteria can be formulated in terms of internal validity, external validity, reliability, and objectivity. Within the naturalistic paradigm, one is better to speak of criteria such as “credibility,” “fittingness,” and “confirmability.” Later Lincoln and Guba (1985) redefined these concepts to credibility, “transferability,” and “dependability.” Guba and Lincoln subsequently formulated several procedures aimed to increase the credibility of qualitative research.

Popular procedures originally conceptualized by Guba and Lincoln are *negative case selection*, *peer debriefing*, *prolonged engagement and observation* in the field, *audit trails*, and *member checks*. Negative case selection is the process of data analysis through which the interpretation of the data is stretched by consciously seeking out and explaining outliers (negative cases) in the data (see also Miles & Huberman, 1994). Peer debriefing is a form of external evaluation of the qualitative research process. Lincoln and Guba (1985, p. 308) describe the role of the peer reviewer as the “devil’s advocate.” It is a person who asks difficult questions about the procedures, meanings, interpretations, and conclusions of the investigation. Prolonged engagement implies that the investigator performs the study for a considerable period. That is to say, a period long enough to adequately represent the subject under investigation (see also Glesne & Peshkin, 1992). An audit trail—also called decision trail—means that researchers document the research process and the choices during that process meticulously and chronologically, for example, through logs and memos. Halpern (1983) identified several classes of record keeping: raw data (e.g., audio files and written notes), data analysis products (e.g., field notes, summaries, and theoretical notes), process

notes (e.g., notes on methodological choices), materials related to the researchers' intentions and dispositions (e.g., research proposal and expectations), and instrument development information (e.g., preliminary schedules and observation formats). This documentation trail allows external evaluators to check the following questions: Can the findings be supported by the data? Are the conclusions logical? Can methodological choices be justified? Member checking involves systematic feedback obtained from informants or participants on the collected data, set categories, interpretations, and conclusions of the study. In member checking, the participants are given the opportunity to assess the credibility of the authors' account (Stake, 1995). Its aim is to minimize the risk of misinterpretations by the researchers.

Many qualitative researchers still regard these criteria as methodological standards. In the wake of Guba and Lincoln, many authors supplemented or perfected their criteria, or suggested alternative terminology for similar procedures. Around the turn of the last century, Morse et al. (2002, p. 15) concluded that this had resulted in a "plethora" of terms and criteria that often brought more confusion than clarity in establishing the validity of qualitative research. Today, still, methodological textbooks on this point show a lot of overlap and most criteria are directly obtained from the themes first conceptualized by Guba and Lincoln.

Critique on Validity Standards in Qualitative Research

Despite efforts to advance the debate on validity, some authors reject the desirability of predetermined criteria for qualitative research altogether. Sandelowski and Barroso (2002), for example, distance themselves from the search for general criteria for qualitative research because in their view the epistemological range of qualitative methods is too broad to be represented by a uniform set of criteria. Instead, they argue for a more rhetorical approach in which the quality of each project must be determined separately for every study. Sandelowski and Barroso (2002) write: "The only site for evaluating research studies—whether they are qualitative or quantitative—is the report itself" (p. 8). In the same vein, Rolfe (2006) points out that qualitative research cannot fall back on a single scientific paradigm. Any attempt to reach consensus on qualitative criteria, according to Rolfe, therefore has little chance. There simply is no common understanding of the field of qualitative theory or methodology which can collectively be described as "qualitative research" (unlike quantitative research, perhaps, that despite the diversity in applications is based on similar mathematical laws). Rolfe argues his case by showing contradictions and paradoxes of common validity checks. Member checking and peer debriefing, for instance, are problematic because if it is assumed that there is no universal truth but only different and additionally constructed truths to which every individual provides his or her own meaning (in effect the premise of much qualitative research), then we cannot expect that the respondents or external evaluators of

qualitative studies will come to corresponding categories and conclusions (cf. Sandelowski, 1993, p. 3).

Hammersley (2007) is also critical of the attempt to formulate uniform criteria of qualitative research. He points out that there are several qualitative approaches that explicitly reject the idea that the production of knowledge should be the only immediate goal of research, and instead insist on political "action." Proponents of this approach believe that qualitative research is a part of the education and social advancement of people and that this function is rendered useless when education is separated from research (see, e.g., J. Elliott, 1988). Related approaches call for a political function of qualitative research by requiring that they should be focused on bringing change of one kind or another: for example, by challenging capitalism, racism, homophobia, or social disadvantage. In addition to traditional epistemological considerations, Hammersley emphasizes that it is important to point out that these approaches produce alternative considerations in assessing the quality of research. Such alternative criteria should be much more formulated in terms of education, politics, ethics, aesthetics, or even economics (e.g., does the study offer value for money?).

Like Rolfe and Sandelowski, Hammersley ultimately rejects the idea that a final set of universal criteria can be formulated. The obstacles to this not only originate out of political "action" objectives but also out of differences in value assumptions. He illustrates this with the example of the growing research on the impact of gender differences in educational achievement of children (see Hammersley, 2007, pp. 294–295). To accept this as a relevant research topic, argues Hammersley, it is vital that one believes in the equality of the sexes (which may not be shared by certain religious groups or sociobiologists). One also has to share the assumption that certain disparities in the classroom affect educational performance, defined in terms of exam success. However, there are people who see gender differences as a predominantly social construct, and there are those who deny that school exams provide a sound indication of educational performance. What Hammersley shows with this example is that research in the social domain is framed by a series of value assumptions which can produce serious differences. The fewer underlying assumptions of a particular research field are shared, the more difficult it is to defend the relevance of the research and the more difficult it is to reach consensus on the validity criteria of that research. Hammersley (2007) nevertheless believes that certain criteria, in the form of "guidelines," can play a role for a more rigorous assessment of qualitative research, though he does not clarify what these guidelines should be.

My conclusion is that guidelines for qualitative research are desirable [.] However, the barriers to our being able to produce any set of common guidelines, even among qualitative researchers, are formidable. At the same time, we should not simply accept at face value methodological pluralism, reinforcing it by treating each qualitative approach as having its own unique set of quality criteria. Dialogue on this issue across

Table 1. Validity Procedures Within Qualitative Lens and Paradigm Assumptions.

Paradigm Assumption/Lens	Postpositivist Paradigm	Constructivist Paradigm	Critical Paradigm
Lens of the researcher	Triangulation	Disconfirming evidence	Researcher reflexivity
Lens of study participants	Member checking	Prolonged engagement in the field	Collaboration
Lens of people external to the study (reviewers and readers)	Audit trail	Thick description	Peer debriefing

Note. Adapted from Creswell and Miller (2000, p. 126).

different approaches, and indeed across the qualitative—quantitative divide is essential for the future of social and educational research. (p. 301)

Finally, according to some authors, the debate on validity criteria has little attention for the ethics of qualitative research. One of the defining characteristics of qualitative methods is that they—more than quantitative methods—provide a participatory function to the researcher. Qualitative research requires that the researcher talks to people and observes them up close and captures their behaviors and experiences accurately. The social interaction with the respondent thus requires tact and sensitivity of the researcher. Davies and Dodd (2002) argue that because of this, the quest for greater rigor cannot be separated from the interaction with the research subject and the ethics that the researcher should take into account (see, e.g., Grol, 2001, on building bridges between professional pride, payer profit, and patient satisfaction). In their eyes, qualitative research should certainly be transparent and accountable but not at the expense of the interests of the respondent and its context. Davies and Dodd therefore argue that the validity of the research should also be formulated in terms of attentiveness, empathy, carefulness, sensitivity, respect, reflection, conscientiousness, engagement, awareness, and openness on the part of the investigator(s).

A Model for Validity in Qualitative Evaluation: Linking Purposes, Paradigms, and Perspectives

One's stance on the question of validity in qualitative research, then, primarily depends on which scientific paradigm is supported, leading some authors to reject the desirability of predetermined criteria for qualitative research altogether. Yet one could equally argue that *different* paradigms require *different* criteria and this line of reasoning also has implications for determining validity standards in qualitative evaluations (to which I will come in a moment). Creswell and Miller (2000) argue that general discussions about validity in qualitative research provide little guidance as to why one procedure might be selected for use by researchers over other procedures. They suggest that this choice is essentially governed by two perspectives: the researchers' paradigm assumptions and the lens researchers use to validate their studies. In order to advance this idea, Creswell and Miller constructed a two-dimensional framework that can help researchers identify appropriate

validity procedures (see Table 1). In the framework, three traditionally competing paradigms are central, derived from Guba and Lincoln (1994), that can shape one's epistemological position toward qualitative research: postpositivism, constructivism, and the critical paradigm (see Creswell & Miller, 2000, pp. 125–126). The postpositivist researcher assumes that qualitative research—like quantitative research—must be systematic and consist of rigorous methods. Within this paradigm, one in fact is looking for the qualitative equivalent of the rigid methodological protocols in the quantitative research community (see e.g., Maxwell, 1996). The constructivist researcher assumes more pluralistic, interpretive, and contextualized perspectives of reality (i.e., sensitive to time, place, and situation). The procedures within this paradigm hence look for an alternative vocabulary for validity labels, for example, transferability instead of “external validity.” The third paradigm assumption involves the critical perspective. This perspective emerged as a critique of alleged structural inequalities of modern society and power structures and was embraced among qualitative researchers who committed to the empowerment of marginalized groups, for instance, through action research (see e.g., Barnes & Cotterell, 2011; Reason & Bradbury, 2001). The implication for validity checking within this paradigm is that the validity of the study should constantly be criticized and negotiated with stakeholders and participants and that researchers should be reflexive and transparent about the kind of knowledge they disclose.

Based on the three paradigm assumptions, Creswell and Miller identify nine different types of validity procedures (see Table 1). Besides the paradigm assumptions, the procedures are arranged to different perspectives—Creswell and Miller call these “lenses”—by which the validity of qualitative research can be assessed (see vertical axis of the table). These lenses constitute the researchers' own perspective, that of the participants in the research or that of external reviewers or readers.

Member checking, audit trail, prolonged engagement, peer debriefing, and disconfirming evidence (negative case selection) are criteria discussed earlier from the work of Guba and Lincoln. Triangulation is a validity procedure where researchers base their categories and/or conclusions on different sources of information (see Denzin, 1978). The researcher might look, for example, whether conclusions derived from interviews are consistent with findings from document analysis and observations. The more the categories and conclusions are confirmed by different data sources, the more valid the results. Reflexivity of the researcher refers to the extent to which

Table 2. Validity Procedures of Qualitative Evaluation Aligned to Purposes, Paradigms, and Perspectives.

Purpose Evaluation > Perspective	Instrumental Effectiveness Policy/Program (Postpositivist Paradigm)	Meaning Policy/Program for Target Group and Practitioners (Constructivist Paradigm)	Empowerment Clients/Target Group/Practitioners (Critical Paradigm)
Evaluator perspective	Triangulation (contrasting)	Disconfirming evidence (fair dealing)	Researcher reflexivity
Evaluation participant perspective	Member checking	Prolonged engagement in the field	Collaboration
External reader/reviewer Perspective	Audit trail	Thick description	Peer debriefing

researchers make their personal values and beliefs explicit in the research report, in such a way that is clear to what extent they might have influenced the results. This can be done in the form of a methodological paragraph or comments throughout the report. Thick description involves the detailed description of the setting, the participants, and the themes of the study. The purpose of thick description is that it creates “probability,” that is, a statement of affairs that takes readers as much as possible into the studied world and its main characters. Detail is the key word here. Researchers should describe, for instance, interactions with informants, personal experiences, or provide a detailed description of the emotions of the respondents. Collaboration is a criterion that is particularly associated with the critical paradigm, meaning that participants should be involved in the study as coresearchers, or in less formal relationships.

Creswell and Millers’ work advances the debate on validity in qualitative research in several ways. It elegantly unites different worldviews or paradigms within qualitative research with key perspectives by which the validity of qualitative research can be assessed: that of the researcher, the respondent, and the external reader. It further explicates the criteria that are essential for each respective paradigm and/or perspective.

The framework of Creswell and Miller provides a basis for a new model for validity in qualitative evaluation. As argued in the introduction of this article, qualitative evaluation can have three different purposes. It can contribute to or focus on the *instrumental effectiveness* of the policy itself (does it work? what are its main working components (process evaluation)?), on the *meaning* of the policy or program for clients, target groups, and practitioners (how do clients and practitioners experience it? How do practitioners shape it?), and it can follow an *emancipatory approach* in which the research itself aims to empower or educate those involved in the program (see, e.g., the many forms of participatory action research). Given their properties and focal points, these evaluation purposes can be linked with the paradigm assumptions Creswell and Miller distinguish. Instrumental effectiveness corresponds to postpositivism. Within the postpositivist worldview, a particular social program or policy is primarily seen as a separate entity—as an “instrument”—whose independent effect can be evaluated accordingly. Postpositivists also tend to believe there is a single reality, whereas constructivists believe that there are multiple, constructed realities (Lincoln & Guba, 1985). The second purpose of qualitative evaluation, uncovering the significance/meaning of the intervention

for clients and target groups, thus corresponds to constructivism, which aims to expose the multiple realities about the implementation and functioning of the policy or program constructed by those involved in the policy or program. Finally, the emancipatory function can be linked to the critical paradigm, which underlines the educational and social advancement of clients and target groups and cooperation between researchers and respondents involved in the evaluation (see also Fetterman, Kaftarian, & Wandersman, 1996). By linking these purposes and paradigms, we can create a new model with relevant validity criteria, specifically for qualitative evaluation (see Table 2).

Naturally, as is the case with Creswell and Miller’s original model, the assessment procedures are partly interchangeable. Member checking and peer debriefing, for example, can be applied in all three paradigms. In this sense, one must keep in mind that the framework is an ideal type. But the model nonetheless poses priority in which procedures are especially important for what paradigm and evaluation purpose. Each procedure in effect serves as a counterweight for inherent methodological weaknesses of the respective evaluation purposes.

In case of a qualitative evaluation that primarily focuses on the instrumental effectiveness of a particular policy or program (does it work? what are its working components?), the criteria triangulation, member checking, and conducting an audit trail are essential. These criteria are most appropriate to avoid or detect spurious (causal) inferences and possible biases, which in itself are significant potential distortions when assessing the instrumental effectiveness of a program or policy. Triangulation, in particular, reduces chance associations and biases due to specific methods used, allowing for greater confidence in interpretations (Fielding & Fielding, 1986; Maxwell, 1992). This is crucial when evaluating the effectiveness of any method or policy. Oliver, Aicken, and Arai (2013) used this procedure to help policy makers make better decisions on childhood obesity. By triangulating user involvement data with a mapping study of interventions aimed at reducing child obesity, the investigators concluded that enhancing mental well-being should be a policy objective, and greater involvement of peers and parents in the delivery of obesity interventions would be beneficial.

If the goal is to uncover the meaning of the intervention for clients and target groups, then the research should acknowledge disconfirming evidence (or negative case selection), there must be prolonged engagement in the field (not a snapshot study) and external readers should be able to identify the experiences

of respondents adequately through thick description. These criteria counterbalance a too-one-sided report of the experiences of particular individuals (disconfirming evidence) or circumstances (prolonged engagement) and allow for a thorough understanding of the experiences of respondents (thick description). Washington, Demiris, Oliver, Wittenberg-Lyles, and Crumb (2012) used the procedure of prolonged engagement to conduct an analysis of informal hospice caregivers who had participated in a structured problem-solving intervention (using open-ended exit interviews). During their prolonged participation in the program, they reported how caregivers actively reflected on caregiving, structured problem-solving efforts, partnered with interventionists, resolved problems, and gained confidence and control. The study thereby provided depth to the understanding of problem-solving interventions for informal hospice caregivers which can be used to enhance existing support services.

If the evaluation has an emancipatory intent (empowerment), then reflexivity of the researcher in the study becomes particularly important. It should become clear how personal beliefs or dispositions might have influenced the investigation as most empowerment-based evaluations (e.g., participatory action research) require a strong involvement of the researcher with his or her research subjects and the theme under study (with the possible risk of “going native”). Elliott, Fischer, and Rennie (1999, p. 221) argue for “owning one’s perspective,” whereby authors specify their theoretical orientations and personal anticipations, both as known in advance and as they become apparent during the research (see also Choudhuri, Glauser, & Perego, 2004; Morrow, 2005). As a hypothetical example of poor practice, Elliot et al. present a case of authors who report an investigation of the process of recovering from childhood sexual abuse, but give no indication of who they are and what they brought to the research. The reader is thereby forced to read between the lines in order to detect the authors’ presuppositions. To illustrate a good practice, Elliott et al. argue that the authors should have described their theoretical, methodological, or personal orientations as relevant to the research (e.g., feminist, symbolic interactionist, and heterosexual); their personal experiences or training relevant to the subject matter (e.g., therapist who works with sexual abuse survivors), and their initial (or emerging) beliefs about the phenomenon they are studying (e.g., that recovery from abuse requires forgiveness). From the perspective of the participants, finally, empowerment evaluations must also employ collaboration, which means that participants should be involved in the evaluation as coresearchers, or in less formal relationships.

Let me further illustrate the model with the hypothetical example I presented in the introduction (support program for pregnant teenagers). Suppose a qualitative case study is performed which aims to investigate the working components of the program. In the case study, interviews, observations, and documentation analysis are conducted. Given its main purpose—evaluating the effectiveness of the program itself—it is essential that from the evaluator’s perspective, triangulation is performed (do findings from interviews with teenagers, observations of the execution of the program by practitioners

and document analysis overlap?) and that from the participant perspective there is member checking (do participant teenagers endorse certain conclusions/interpretations made by the evaluators?), and an audit trail is conducted so that external reviewers can verify if presented findings can be supported by the data and (causal) inferences about the workings of the program are grounded (e.g., are intended effects—such as engaging the teenagers to remain in school—indeed achieved? On what data are these conclusions based? On what grounds are arguments made?). The same steps can be followed with the other evaluation purposes (meaning and empowerment) “checking” procedures from the columns down and linking them with the perspectives from the rows.

Note that in the new evaluation model, Creswell and Millers’ original criteria are completed with other relevant procedures. Triangulation can be enhanced by *contrasting* outcomes with findings from other types of research or previous research outcomes (see Onwuegbuzie & Leech, 2007). For instance via so-called multisite studies whereby observations from different evaluative situations in different research locations are compared, or through the systematic comparison of results of qualitative research with insights from the (scientific) literature. The principle of *fair dealing* (Dingwall, 1992; see also Mays & Pope, 2000) is a logical addition to disconfirming evidence and is particularly relevant when the evaluation aims to uncover multiple realities. Fair dealing ensures that many different perspectives are covered in the evaluation (not only that of the policy’s or program’s target group) so that the viewpoint of one particular group cannot be presented as an overall representation of the program. Staying with our example, this criterion ensures that not only the pregnant teenagers are interviewed but also the relevant groups around the teenagers such as social professionals, trainers, family members, and so on. After all, in such an evaluation one does not investigate “pregnant teenagers” but the program designed for their education and support and to assess it adequately, therefore, several perspectives are needed.

Conclusions and Discussion

It is important to note that the framework presented in this article *can* serve as a checklist for qualitative evaluations. But its main value is as a (theoretical) reference point. Its most important feature is that it avoids “taking sides” in a paradigmatic and epistemological sense. Instead, it accommodates a more pragmatic approach when taking the different purposes, paradigms, and perspectives of qualitative evaluation into account. What criteria are preferred for “good” qualitative research will always be dependent on one’s scientific world view, and these preferences can change over different time periods (Lewis, 2009). It can be expected, therefore, that practitioners and policy makers will continue to make use of different types of qualitative evaluations—emphasizing different purposes and starting from different paradigms—to evaluate their specific programs and policies. All the more reason not to assess the truth value of such investigations on one particular monolithic world view, but instead to let qualitative evaluation

criteria correspond with the paradigms and lenses through which it can be assessed and the different functions that qualitative information can have (instrumental, meaning, and empowerment). In this way, the model not only combines flexibility with rigor, it also answers the call for some “common guidelines” (Hammersley, 2007) while at the same time respecting paradigmatic differences.¹

However, we must keep in mind that the actual application of validity procedures of qualitative inquiry takes time and energy. Whether it concerns member checks, keeping an audit trail, or thick description of the data, respecting validity criteria for qualitative research is easier said than done (causing some researchers to present a “procedural charade” in their reports, see Whittemore, Chase, & Mandle, 2001). In the realm of policy and program evaluation, in particular, it can be difficult to maintain certain standards. A PhD student working with a time frame of several years will generally have the patience and opportunity to apply validity procedures adequately. But for an evaluator or policy researcher who has to make an assessment of the impact of a social measure in, say, 2 months because the political situation calls for it, the situation is different. For him or her, the temptation will be greater to cut corners in the analysis. It is therefore important that funders of qualitative evaluations create the time and space for evaluators to implement validity criteria in earnest.

Finally, apart from the methodological and practical considerations, it would be fruitful to take a step back and study the social, cultural, and institutional aspects of some of these issues (see also Strassheim & Kettunen, 2014). It would be interesting, for instance, to discern why certain preferences for particular paradigms and purposes for evaluation seem to correspond with different time periods and sectors. In health research, the personal experience and realities clients and target groups provide to a particular program (constructivist paradigm) certainly has become more important over the last two decades or so, and this development has served as a supplement (or perhaps counterweight) to the dominance of postpositivist investigations focused on the instrumental effectiveness of programs. In community and social work, it seems the reverse is at work. Historically, highly influenced by postmodern and constructivist schools of thought, programs in community and social work are now increasingly fitted into experimental or “quantized” research models reminiscent of the old *modus operandi* in the health sector. Moreover, in several European countries, institutions and government agencies that are active in community and social work have set up databases of “effective interventions” analogous to the already established *evidence-based* databases in the health sectors (treating interventions as “independent instruments”: postpositivist paradigm, see, e.g., the Cochrane reviews and Campbell Collaboration). The emancipatory function of evaluation (critical paradigm), prevalent in the 1970s and 1980s, is today again visible in research projects commissioned by the European Union (EU). Most EU Research Calls demand involvement of practitioners and negotiations with stakeholders and require that proposals elaborate on how such “end users” will benefit from the undertaken research.

Sociological and sociohistorical research not only can shed light on how and why such sectorial paradigm shifts occur, it could also investigate how these shifts influence ideas on what counts as “evidence” of particular social programs or policies, and if or how this in turn influences ideas on the role of qualitative information in evaluation.

Author's Note

The author thanks Movisie, Netherlands Institute for Social Development; RIVM National Institute for Public Health and the Environment; Netherlands Institute for Sport & Physical Activity.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: Financial support for the study underlying this article was received from Movisie, Netherlands Institute for Social Development; RIVM National Institute for Public Health and the Environment; Netherlands Institute for Sport and Physical Activity.

Note

1. One can pose the question if the model should be supplemented with ethical criteria, such as those formulated by Davies and Dodd (2002). I propose not to. Not because ethical considerations are not important, but because the objective of upholding certain ethical standards in evaluation research is to some extent naïve. In practice, it is impossible to be fully transparent about the research outcomes to respondents or to fully protect their interests and contexts. Evaluating a program or policy requires a critical stance and it goes without saying that some research results might affect respondents negatively. For instance, because it turns out that practitioners of a social program do not perform well or that their particular approach is actually counterproductive. Should the qualitative researcher be transparent about such possible outcomes in advance, then it would be unlikely he or she could count on much cooperation from respondents in the field.

References

- Barnes, M., & Cotterell, P. (2011). *Critical perspectives on user involvement*. Bristol, England: The Policy Press.
- Choudhuri, D., Glauser, A., & Peregoy, J. (2004). Guidelines for writing a qualitative manuscript for the Journal of Counseling & Development. *Journal of Counseling & Development, 82*, 443–446.
- Creswell, J. W., & Miller, D. L. (2000). Determining validity in qualitative inquiry. *Theory into Practice, 39*, 124–130.
- Davies, D., & Dodd, J. (2002). Qualitative research and the question of Rigor. *Qualitative Health Research, 12*, 279–289.
- Denzin, N. K. (1978). *The research act: A theoretical introduction to sociological methods*. New York, NY: Praeger.
- Denzin, N. K., & Lincoln, Y. S. (Eds.) (2005). *The Sage handbook of qualitative research* (3rd ed.). Thousand Oaks, CA: Sage.
- Dingwall, R. (1992). Don't mind him—He's from Barcelona: Qualitative methods in health studies. In J. Daly, I. McDonald, & E.

- Willis (Eds.), *Researching health care* (pp. 161–175). London, England: Tavistock/Routledge.
- Elliott, J. (1988). Response to Patricia Broadfoot's presidential address. *British Educational Research Journal*, 14, 191–194.
- Elliott, R., Fischer, C. T., & Rennie, D. L. (1999). Evolving guidelines for publication of qualitative research studies in psychology and related fields. *British Journal of Clinical Psychology*, 38, 215–229.
- Fetterman, D. M., Kaftarian, S. J., & Wandersman, A. (1996). *Empowerment evaluation: Knowledge and tools for self-assessment and accountability*. Thousand Oaks, CA: Sage.
- Fielding, N., & Fielding, J. (1986). *Linking data*. Beverly Hills, CA: Sage.
- Gibbs, A., Jewkes, R., Sikweyiya, Y., & Willan, S. (2015). Reconstructing masculinity? A qualitative evaluation of the Stepping Stones and Creating Futures interventions in urban informal settlements in South Africa. *Culture, Health & Sexuality: An International Journal for Research, Intervention and Care*, 17, 208–222.
- Glesne, C., & Peshkin, A. (1992). *Becoming qualitative researchers: An introduction*. White Plains, NY: Longman.
- Grol, R. (2001). Improving the quality of medical care: Building bridges among professional pride, payer profit, and patient satisfaction. *Journal of the American Medical Association*, 286, 2578–2585.
- Guba, E. G., & 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., & Lincoln, Y. S. (1994). Competing paradigms in qualitative research. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (pp. 105–117). Thousand Oaks, CA: Sage.
- Haber, R., Carlson, R. G., & Braga, C. (2014). Use of an anecdotal client feedback note in family therapy. *Family Process*, 53, 307–317.
- Halpern, E. S. (1983). *Auditing naturalistic inquiries: The development and application of a Model* (Doctoral dissertation). Indiana University, Bloomington.
- Hammersley, M. (2007). The issue of quality in qualitative research. *International Journal of Research & Method in Education*, 30, 287–305.
- Lewis, J. (2009). Redefining qualitative methods: Believability in the fifth moment. *International Journal of Qualitative Methods*, 8, 1–14.
- Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry*. Thousand Oaks, CA: Sage.
- Lohan, M., Aventin, A., Maguire, L., Clarke, M., Linden, M., & McDaid, L. (2014). Feasibility trial of a film-based educational intervention for increasing boys' and girls' intentions to avoid teenage pregnancy: Study protocol. *International Journal of Educational Research*, 68, 35–45.
- Lub, V. (2014). *The Plausibility of Policy. Case studies from the social domain*. The Hague: Eleven International Publishing.
- Maxwell, J. A. (1992). Understanding and validity in qualitative research. *Harvard Educational Review*, 62, 279–299.
- Maxwell, J. A. (1996). *Qualitative research design: An interactive approach*. Thousand Oaks, CA: Sage.
- Mays, N., & Pope, C. (2000). Assessing quality in qualitative research. *British Medical Journal*, 320, 50–52.
- Merriam, S.B. (1998). *Qualitative research and case study applications in education*. San Francisco: Jossey-Bass.
- Meyer, J. (2006). Action research. In C. Pope & N. Mays (Eds.), *Qualitative research in health care* (pp. 121–131). Oxford, England: Blackwell.
- Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook* (2nd ed.). Thousand Oaks, CA: Sage.
- Morrow, S. L. (2005). Quality and trustworthiness in qualitative research in counseling psychology. *Journal of Counselling Psychology*, 52, 250–260.
- Morse, J. M., Barrett, M., Mayan, M., Olson, K., & Spiers, J. (2002). Verification strategies for establishing reliability and validity in qualitative research. *International Journal of Qualitative Methods*, 1, 13–22.
- Oakley, A., Strange, V., Bonell, C., Allen, E., & Stephenson, J. (2006). Health services research: Process evaluation in randomised controlled trials of complex interventions. *British Medical Journal*, 332, 413–416.
- Oliver, K., Aicken, C., & Arai, L. (2013). Making the most of obesity research: Developing research and policy objectives through evidence triangulation. *Evidence & Policy*, 9, 207–223.
- Onwuegbuzie, A. J., & Leech, N. (2007). Validity and qualitative research: An oxymoron? *Quality & Quantity*, 41, 233–249.
- Parahoo, K. (2014). *Nursing research. Principles, process and issues* (3rd ed.). Basingstoke, England: Palgrave Macmillan.
- Pope, C., & Mays, N. (2006). *Qualitative research in health care*. Oxford, England: Blackwell.
- Reason, P., & Bradbury, H. (2001). *Handbook of action research. Participative inquiry and practice*. London, England: Sage.
- Robert, G., & Cornwell, J. (2011). *What matters to patients? Developing the evidence base for measuring and improving patient experience*. Department of Health/NHS Institute for Innovation and Improvement. Coventry: NHS Institute for Innovation and Improvement.
- Roe, D., & Lysaker, P. H. (2012). Concerns and issues that have emerged with the evolution of evidence-based practice. *Journal of Mental Health*, 21, 427–429.
- Rolfe, G. (2006). Validity, trustworthiness and rigour: Quality and the idea of qualitative research. *Journal of Advanced Nursing*, 53, 304–310.
- Rossi, P. H., Lipsey, M. W., & Freeman, H. E. (2004). *Evaluation. A systematic approach* (7th ed.). Thousand Oaks, CA: Sage.
- Sandelowski, M. (1993). Rigor or rigor mortis: The problem of rigor in qualitative research revisited. *Advances in Nursing Science*, 16, 1–8.
- Sandelowski, M., & Barroso, J. (2002). Reading qualitative studies. *International Journal of Qualitative Methods*, 1, 1.
- Stake, R. (1995). *The art of case study research*. Thousand Oaks, CA: Sage.
- Strassheim, H., & Kettunen, P. (2014). When does evidence-based policy turn into policy-based evidence? Configurations, contexts and mechanisms. *Evidence & Policy*, 10, 259–277.
- Washington, K. T., Demiris, G., Oliver, D. P., Wittenberg-Lyles, E., & Crumb, E. (2012). Qualitative evaluation of a problem-solving intervention for informal hospice caregivers. *Palliative Medicine*, 26, 1018–1024.
- Whittemore, R., Chase, S. K., & Mandle, C. L. (2001). Validity in qualitative research. *Qualitative Health Research*, 11, 117–132.
- Yin, R. K. (1994). *Case study research, design and methods*. Thousand Oaks, CA: Sage.

Copyright of International Journal of Qualitative Methods is the property of International Institute for Qualitative Methodology and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.