

Alma Mater Studiorum Università di Bologna
Archivio istituzionale della ricerca

Methodological Paper What counts as “good” qualitative accounting research? Researchers' perspectives on assessing and proving research quality

This is the final peer-reviewed author's accepted manuscript (postprint) of the following publication:

Published Version:

Steccolini, I. (2023). Methodological Paper What counts as “good” qualitative accounting research? Researchers' perspectives on assessing and proving research quality. ACCOUNTING, AUDITING & ACCOUNTABILITY JOURNAL, 36(3), 1032-1057 [10.1108/AAAJ-05-2022-5808].

Availability:

This version is available at: <https://hdl.handle.net/11585/948862> since: 2025-02-05

Published:

DOI: <http://doi.org/10.1108/AAAJ-05-2022-5808>

Terms of use:

Some rights reserved. The terms and conditions for the reuse of this version of the manuscript are specified in the publishing policy. For all terms of use and more information see the publisher's website.

This item was downloaded from IRIS Università di Bologna (<https://cris.unibo.it/>).
When citing, please refer to the published version.

(Article begins on next page)

What counts as ‘Good’ Qualitative Accounting Research?

Researchers’ Perspectives on Assessing and Proving Research Quality

By Ileana Steccolini

Essex Business School

University of Essex

ileana.steccolini@essex.ac.uk

ABSTRACT

Purpose. This study explores the everyday experiences of researchers in assessing their own and others’ research, highlighting what ‘good’ qualitative accounting research is from their perspectives.

Design/methodology/approach. The analysis is based on interviews with accounting scholars from the UK, Germany, Italy, Spain and Australia, with diverse ethnic background and methodological preferences.

Findings. Interviewees pointed to a plurality of practical, and to some extent tacit, ways in which they demonstrate and assess the quality of research, concerning “contribution”, “consistency” and “confidence”, with generalizability being seen as more controversial and difficult to attain. In general, interviewees highlighted the underlying ambiguity on what constitutes good research in the qualitative accounting community, contrasting it to the perceived stronger clarity to be found in the quantitative accounting community. This was seen as potentially strengthening the positions of “gatekeepers” in the accounting communities, and encouraging conformance and “signaling” behaviors, at the risk of hampering innovation.

Originality/value. The main critical issues affecting qualitative research quality highlighted by interviewees concern the engagement with the world of practice, and with theory and literature, the importance of accounting for the analysis of qualitative data and for the messiness of the underlying process, and the implicit search for compliance with editors’ and community’s expectations and conventions. These findings suggest the need to continue debating how we assess the quality of qualitative research in our everyday activities, and reflect on how we can promote acceptance and openness to pluralism, in our communities, as well as in data collection, analysis, in the theorizing, and in connecting epistemology and methodology.

1. Introduction

What “good” qualitative accounting research is has been the subject of a lively debate. Various conceptual contributions have reflected on how qualitative accounting research should be assessed, and its quality proven (e.g., Lukka and Kasanen, 1995; Ahrens and Dent, 1998; Ahrens and Chapman, 2006; Baxter and Chua, 2008; Lukka and Modell, 2010; Parker, 2012; Messner et al., 2017; Parker and Northcott, 2017; de Villiers et al., 2019). This study complements this literature by providing an exploration of the practical ways in which scholars demonstrate and assess the quality of qualitative accounting research.

Understanding how qualitative accounting scholars understand, internalize, interpret, and translate research quality criteria is of particular importance, in light of (i) the critical relationship with the quantitative accounting approaches, (ii) the multiplicity of potential quality criteria, and (iii) the difficulties faced by qualitative researchers in processes of research assessment. “Mainstream” accounting scholarship has been generally described as embracing predominantly quantitative approaches (de Villiers et al., 2019), and relying

on established quality criteria, such as validity and reliability. In the qualitative accounting arena, similar to wider trends in social , ‘alternative’ criteria have been increasingly suggested (for example, Lukka and Modell, 2010). In spite of this, it has been pointed out that qualitative studies tend to be criticized not necessarily (or not only) for their actual shortcomings, but for being evaluated as if they were quantitative studies (for example, de Villiers et al., 2019). Even when specific criteria are adopted for qualitative studies, these tend to be less established, multiple and variable (Lukka and Modell, 2010; Parker and Northcott, 2016). As Parker and Northcott (2016: 1117) put it, in the qualitative field, “[...] demonstrating the credibility of research conducted arguably admits a greater degree of democracy than the quantitative approach to establishing validity.” Moreover, reliance on forms of rankings, and existing measures of performance are seen as penalizing interdisciplinary, interpretive and qualitative researchers (for example, Gendron, 2013; Guthrie and Parker, 2019). It has thus been suggested that qualitative researchers, having internalized a habit to compare themselves with the quantitative community, may rather need to recognize that the aims of qualitative and quantitative studies are different, and they thus cannot be evaluated on the same bases (de Villiers et al., 2019). More generally, it may be useful to provide full accounts on the practical ways in which qualitative accounting scholars prove and assess research quality, to complement the conceptual debate with experiences from the field, to further raise awareness in our communities on everyday practices of research assessment, and also to identify possible areas of improvement.

This paper focuses specifically on the experiences of scholars who engage, fully or partially, in qualitative accounting research, exploring the ways in which, as authors, they ensure and signal the quality of their research, and, as reviewers and readers, they assess it. In doing so, and through reporting the results from 18 interviews, this paper aims to provide a multi-vocal narrative on how colleagues conceive of, and assess, the quality of their own, and others’ research¹.

In the face of the proliferation and plurality of criteria suggested in handbooks, the analysis highlights that, in their everyday experiences, researchers tend to focus on three main areas for assessing qualitative research (contribution, consistency, confidence). They also identify specific critical issues common to qualitative papers, including the extent of engagement with theory and literature, the importance of accounting for the analysis of qualitative data, and for the messiness of the underlying process, and the implicit search for compliance with editors’ and community’s expectations and conventions. In general, interviewees highlight that what makes good research still remain ambiguous, and the feeling that this is in contrast with the stronger clarity characterizing the quantitative accounting community. This is seen as potentially strengthening gatekeeping practices, and encouraging conformance and “signaling” behaviors, while discouraging innovation.

The paper is structured as follows. The next section describes some of the most common criteria suggested in the literature about qualitative research. Section 3 clarifies the methods. Section 4 presents the results, which

¹ It is worth noticing that my use of the word “qualitative” refers to the “methodological” level (i.e., the use of qualitative methods to collect and analyse data), and not to ontological and epistemological choices. While it has been found that often qualitative methods are more likely adopted within an interpretive paradigm (and at times are seen as more in line with the relevant ontological and epistemological bases) and quantitative methods are more likely (and seen as more consistent) within the positivistic paradigm (Kihn and Inhantola, 2015), quantitative approaches are also used in interpretive research (sometimes jointly with qualitative), and some (methodologically) qualitative research takes a positivistic epistemological stance. Hence, when referring to qualitative approaches, I will be focusing especially on the methodological level, though some reflections may also be mirror the predominance of interpretive stances in the accounting qualitative community. It is also worth noticing that, to account for this diversity, all the interviewees have experience with qualitative research, but to different degrees (some conduct only qualitative research, others predominantly qualitative, yet others also have wide experience of quantitative research).

are further discussed in section 5. Section 6 draws conclusions and identifies possible implications and ways forward.

2. Assessing and proving the quality of qualitative research: overview of relevant literature

The quality of qualitative research, and how it should be recognized and assessed, has been the subject of significant attention and debate over the last few decades. Three main positions exist on how qualitative research should be ascertained (for example, Steinke, 2004). According to some, it should follow the same quality criteria of quantitative studies (construct-, internal-, and external- validity, and reliability). According to others, no criteria should be pre-defined to decide what is good qualitative research (for example, Bochner, 2000). Yet, an intermediate position suggests that ad-hoc criteria are to be identified for assessing qualitative research (e.g. Lincoln and Guba, 1985; Creswell, 2007; Richardson, 2000). A plurality of quality criteria for qualitative research have been proposed in the general literature. Some “classics” in particular appear to have been quite influential in qualitative accounting literature, including, among others, Miles and Huberman’s (1994) adaptation of quantitative criteria into confirmability of qualitative investigations, dependability/auditability, credibility/authenticity, transferability/fittingness, application/action orientation; Lincoln and Guba’s (1985) dependability, credibility, transferability and confirmability, Golden Biddle and Locke’s (1993) convincingness, authenticity, plausibility and criticality. In addition to those, a comprehensive list of criteria is provided by Tracy’s (2010) worthy topic, rich rigor, sincerity, credibility, resonance, significant contribution, ethical, meaningful coherence (Appendix 1 provides the relevant definitions for each author).

In conceptual contributions around the assessment of qualitative research in accounting, the concepts of credibility, plausibility, authenticity, originality and contribution, and generalizability appear to have attracted particular attention. In their chapter, for example, Messner et al (2017) specifically discuss issues of *credibility and authenticity*. Drawing on Lincoln and Guba (1985), they define credibility as the extent to which a qualitative account is convincing with reference to its findings, which will depend on the “strength” of empirical data, and the plausibility of the interpretation (p. 433). Other accounting scholars have particularly pointed to the relevance of plausibility in qualitative studies, at times drawing on Golden-Biddle and Locke (1993), referring to their attitude to “make sense”, also thanks to reliance on a “thick description” (Kihn and Inghantola, 2015; Lukka and Modell, 2010; Baxter and Chua, 2008), but also the complexity of ascertaining it (Ahrens and Chapman, 2006). Relatedly, Ahrens and Dent (1998), focusing on the richness of field research, identify three critical issues to be addressed, including the deep understanding of the field, the process of theorizing and the emergence of theoretical constructs from the field material. Lukka and Modell (2010), adopting a “pragmatist stance”, suggest further that, in ensuring plausibility of research (related to the “etic” view of research, i.e. its ability to explain and make sense), explanations cannot only rely on the identification of patterns and regularities, and that the latter are not necessary conditions for explanation. Stronger attention should then be paid to counterfactuals and contrast classes, and abduction.

In parallel, *authenticity* will require the “skillful exploitation” of the rich empirical material rather than simply providing abstract propositions (Messner et al. 2017: 437 and Baxter and Chua, 2008; both also referring to Golden-Biddle and Locke, 1993), thus ensuring proof that of the researchers’ presence in the field, and to allow them to communicate the complexity of such field, in an ‘emic’ perspective (on this, also Lukka and Modell, 2010). In addition to authenticity and plausibility, some authors highlight the importance of *criticality* (Baxter and Chua, 2008), defined by Golden-Biddle and Locke (1993) as the attitude of research to encourage readers to re-think the taken for granted assumption underlying their research work.

Contribution appear to be the most difficult criteria to define, as it “means different things to different people”, may change over time, and appears to rather reflect prior socialization and complex social processes (Gendron, 2013: 1). While this may attract critique, it has been observed that “chaos and confusion over the very nature of research contribution are consistent with the search for knowledge about a fundamentally ambiguous, contradictory and unstable reality” (*ibidem*, p. 11). A particularly contentious issue concerns its relevance for policy and/or practice, as documented by Gendron (2012). Yet, according to some, this type of contribution is seen as particularly salient in qualitative research (for example, de Villiers et al., 2019).

Generalization has occupied a peculiar place in the qualitative accounting research debate. Parker and Northcott (2016) provide a comprehensive discussion about it, identifying the multiple forms of generalizing, and classifying them into two main typologies, depending on the relevant audience. Theoretical generalizing becomes key when the audience is especially academia, and it happens when “[...]the researcher employs logical analysis and/or contextualized thick description to support a generalization argument that s/he develops and presents to the reader. This [...] primarily seeks to deliver convincing findings to an audience interested in theory development.” (*ibidem*: 1114). Conversely, analytical generalizing, including analogical, communicative, naturalistic forms of generalization, “[...] has an engaged and democratic flavor” (*ibidem*: 1114) in that “[...] researchers, we only begin the process of generalization which also rests with the actors and ultimately is in the hands of the readers. These readers need to understand the research study site and context so they can compare it to the settings with which they are familiar. The more the research study context is analogous to a reader’s own context, the more plausible and defensible is the transferability of research findings and resulting naturalistic generalization.” For this type of generalization, practical relevance, and making a difference to practice is key, and the audiences are mostly practitioners and policy makers.

Among the few studies exploring how the above quality criteria are applied in practice, Kihn and Ithantola (2015), in their analysis of 212 management accounting qualitative papers in 11 accounting journals over 2005/2015, find that about 1/3 of the papers adopted classic (validity and reliability) criteria; almost half abandoned any criteria; only 9% embraced the alternative ones, specific to qualitative research; and ca. 14% mixed classic and alternative. Interestingly, the authors also found that papers abandoning common criteria generally provided accurate description of methods, and that, while having adopted a mainstream approach will not mean a specific preference for criteria, conversely, having adopted critical or interpretive will make it more likely to report alternative criteria or abandon them. A few other studies have looked at the concrete application of quality criteria in some sample papers, generally pointing to the complex nature of validation in qualitative research (Ahrens and Dent, 1998; Ahrens and Chapman, 2006) or highlighting the risks of bias in the interpretation of data (McSweeney, 2021). It would thus be interesting to complement these desk-based analyses with explorations of the personal views and experiences of accounting scholars on what counts as good research, and thus how they feel they need to demonstrate the quality of their research, and how, in turn, they assess others’ research.

3. Methods

To explore the views and experiences of accounting scholars on how they assess and prove the quality of qualitative accounting research, I focused on authors and reviewers active in the accounting community, at different stages of their careers (from PhD students, early career researchers, through lecturers/assistant professors, senior lecturers/associate professors, and readers, to the exclusion of more senior positions – professors –). Senior colleagues often have their views shared as speakers, editors, or leaders in their institutions, whereas less may be known on the perceptions of less senior colleagues, who are deeply involved in research both as authors and reviewers. Also, several of them had not been previously interviewed, which may suggest their responses to be spontaneous, though I cannot in principle exclude they trying to “impress”

the interviewer or offer “desirable” answers. These interviewees may also be expected to be in a position where the concrete ways in which research is assessed may significantly shape their careers, while being less likely to be already in roles where they can actively and fully shape the used quality criteria, and editorial policies.

After obtaining ethical approval, I conducted 18 semi-structured interviews during July and August 2021. They all took place via Zoom, as requested by the Covid-19 pandemic context of the country where I operated. Zoom allowed recording and (in most cases) automatic transcriptions, complemented by my own notes. Interviews revolved around the motivations behind the choice of conducting qualitative research, the context where such research is conducted (institutional, organizational, scientific community, world of practice and policy), including enabling and constraining conditions, as well as the requirements of good qualitative research, both from an author’s and reviewer’s perspective. This paper focuses particularly on the last theme, but the others allowed to provide context for the responses collected. The initial/standard questions concerning this theme were kept general (e.g., “What is good qualitative accounting research for you?”), avoiding wording which may suggest reference to specific concepts (validity or reliability; consistency and credibility etc.), so as to let the specific views, experiences, and wording of colleagues to emerge. During the discussion, and depending on responses, more specific questions and probes were used (in particular referring to how researchers feel they prove the quality of their research as authors, and how they assess it as reviewers).

The interviewees were granted anonymity and will thus be identified with letters from the Greek Alphabet in the presentation of findings. Particular attention was devoted to include people of different genders, geographical areas, institutions and ethnicity. It must however be noticed that the countries covered, in terms of academic institutions, were Australia (1), Germany (1), Italy (4), Spain (1) and the UK (11). While the latter was more represented in the pool of interviewees, it allowed coverage of a plurality of ethnical origins, including people from different European (5), African (3), South American (1), and Asian countries (2). The presented findings may thus be further enriched by future studies including people currently working at institutions in different geographical areas, such as Asia, Africa, or the Americas, as well as other European countries with different traditions (such as Eastern European ones). It may thus be expected that the findings from the interviews reflect predominantly the reality of Western Europe, seen from the eyes of people with a certain plurality of ethnic backgrounds and origins. Future studies may be usefully complemented by (or contrasted with) views coming from Eastern Europe, North and South America, Asia and Africa. A further aspect to be noticed is that the pool of interviewees includes in equal parts both colleagues who only conduct qualitative research, and others who also conduct quantitative and mixed methods analyses (including people giving different weight to them in their research). Interestingly, even those conducting solely qualitative analysis tend to have some background in quantitative research methods. So, while nobody was interviewed who is only a quantitative researcher, the interviewees were diverse in terms of the research time devoted only to qualitative research.

Among the people interviewed, some appeared to take more a pragmatist view, others embraced a more interpretive perspective, yet others appeared to take a more critical stance. This allows to account for a variety of ontological and epistemological positions, while, as pointed out above, the main focus of the paper is on qualitative methods, and not on the underlying ontological and epistemological approaches. Exploring how people reflect these connections in their everyday experiences may represent a further development of this study.

The interviews were analyzed through a thematic analysis. In a first stage, all the interview transcripts and notes were read and detailed themes emerging were identified (such as examples of specific ways in which confidence in research was guaranteed). In a second stage, detailed themes were merged into more aggregate “intermediate” themes (e.g., “elements” vs “critical issues” concerning confidence or contribution). Finally,

the intermediate themes were further connected to overarching themes (e.g., confidence, contribution). These were also compared with extant literature, to identify areas where interviews confirmed it, and areas where new, in-depth or controversial reflections emerged. Intermediate findings were shared with colleagues, including in seminars, to gather feedback.

After 4 months from the initial analysis, the interviews were read again with the specific aim to identify alternative themes, different interpretations of the data, and possible missing nuances, which may have gone unnoticed in the first stage, in trying to reduce risks of simply focusing on saturation and confirmation of initial ideas, and purposely looking for diverging views, and disconfirming evidence or counter-accounts. During the first stage the focus had been to identify common themes and reduce the data complexity, and reach some degree of “saturation” to converge on a “main structure”. Conversely, the second stage was aimed at adding depth to the analysis, identify additional nuances, and contrast perspectives, along the lines of a “crystallization” exercise (Richardson, 2000; Tracy, 2010). This allowed the voices of interviewees be represented in their multiple facets, enriching the main structure previously identified. The final structure of the analysis, and emerging themes and sub-themes, are summarized in table 1. This structure is followed in the presentation of results in section 4. It is also worth pointing out that section 4 and 5 serve two distinct purposes. Section 4 teases out the main themes and subthemes emerging from the interviews, with a specific emphasis on providing a multi-faceted account on the interviewees’ views, and letting a narrative to emerge from their voices. Subsequently, in Section 5 I integrate, reflect and interpret the results, and compare and contrast them with relevant literature. In clarifying this, it is also important to recognize that Section 4 still reflects my subjective choices concerning the structure of the themes, the sequencing and organization of narratives, and the selection of quotations.

4. Researchers’ experiences with assessing and proving the quality of qualitative accounting research

This section illustrates the results emerging on the experiences of interviewees with assessing and proving the quality of qualitative accounting research in their everyday professional lives. As highlighted above, this section is specifically devoted to represent their voices, within the “structure” based on my identification of themes and sub-themes, illustrated in Table 1 and informing the narrative in this section.

Overall, interviewees appeared to highlight that standards of quality for qualitative research are much less clear and agreed upon than for quantitative studies.

“So I think sometimes there is a lack of understanding of [...] what differentiates between rigorous, well done and poorly done qualitative research. While perhaps with quantitative research [...], the standard is slightly more settled and agreed upon, perhaps.” (Epsilon).

The criteria and practical ways in which interviewees identified good research were multiple, and named and described in different ways. They can be traced back to three main concepts: contribution, consistency, and confidence. A fourth concept, generalizability, emerged as less consensual and more controversial. For each, several elements and facets, as well as critical issues, emerged, as shown in table 1. These “concepts” relevant for assessing research quality, their dimensions, and ways to assess them, as well as the related critical issues, as narrated by interviewees, are discussed in the next sub-sections.

Table 1 – The structure of the thematic analysis

Main research quality dimensions	Sub-dimensions and relevant features of research quality dimensions	Emerging critical issues concerning research quality dimensions and sub-dimensions
Contribution		Implicit and variable standards for assessing contribution
	<i>Contribution to theory</i>	Over-theorization Use of theories to signal allegiance to certain communities
	<i>Contribution to literature</i>	Need for more engagement with literature
	<i>Contribution to practice</i>	Practical contribution not relevant for all and not easy to attain
Consistency		
	<i>Facets of consistency</i>	Tradeoffs between the dimensions of consistency Lack of connections between research questions, methods, findings and conclusions
	Sampling and unit of analysis	Lack of explanation on the choice of countries, contexts, organizations, interviewees
	Saturation	Tradeoffs with representation
	Representation	Lack of coverage of a plurality of voices focusing only on “one side” of the story
	Counterfactuals	
	Contextualization	
	<i>Qualitative features of consistency</i>	
	Alignment with ambitions	
	Plausibility	Inherent subjectivity
	Persuasivity	Importance of narratives and interaction with audiences
Confidence		Honesty vs calculative accountability “A round peg into a square hole” – undue influences from quantitative research
	<i>Richness of data</i>	Thinness of data
	<i>Audit trail</i>	Lack of acknowledgement of limitations
	<i>Accounting for the analysis</i>	Lack of explanation on how the analysis was conducted
	<i>Accounting for the messiness of the process and serendipity</i>	Simplified, ex-post rationalization of research process
Generalizability		
	<i>Theoretical generalization</i>	Not always seen as necessary or in line with aims of the research
	<i>Analytical generalization</i>	Seen as difficult to translate into practice

4.1. The contribution

Contribution to theory, the literature...and practice? Interviewees pointed to one fundamental quality criterion for qualitative research to be its ability to challenge extant theories, literature, conventional ways of seeing things. However, others also highlighted the importance to shed lights on important phenomena,

making a difference in policy and practice. In describing these ideas, they referred to concepts such as “relevance”, “contribution”, “challenging what is already known” or “taken for granted”, “intriguing”, “surprising”, “learning something new” or looking at some phenomenon in new ways.

“Relevance is the step made in respect to extant literature. [...] ‘specific’ advancement to the literature, not yet ‘another’ paper on sustainability accounting or using Foucault, but a further step.” (Ksi)

“[...] you must be able to question the conventional wisdom [...] Lessons [...]if they don’t challenge the very domain assumption, than they should at least challenge the general perception or common understanding” (Beta)

While the importance of theoretical relevance was generally acknowledged, the issue of **practical relevance** appeared to remain more in the background of discussions. Yet, for some colleagues, especially taking a critical stance, this was considered a crucial aspect, and one deserving more attention:

“Sometimes we pay so much attention to the methodology that we don’t even ask ourselves ‘yeah but is this relevant?’ [...] You say ‘Okay, I would like to study if going to the supermarket using one way is shorter than using the other way’ [...]if it provides the contribution which is clear, and not just clear but actually relevant, whether the paper is actually adding something substantial” (Epsilon)

Implicit and variable standards to assess contribution? In discussing contribution, interviewees also suggested that sometimes assessing the “contribution” of a paper was more the result of an implicit process, than a deliberate one. They felt they would “recognize” the contribution as part of tacit social norms and shared ways of assessing papers. At the same time, the subjectivity and relativity of the standard of relevance was often pointed out, as qualitative research was seen as having the potential to be judged differently by different audiences, communities or journals.

(Dis-)engaging with theory. As pointed out above, “contribution” was mostly seen in relation to theories and literature, yet, several colleagues pointed to one of the typical shortcomings of qualitative papers being their ceremonial engagement with theories, and lack of engagement with the literature (i.e., with extant empirical studies providing evidence in the area of the paper), even when they widely engage with the underlying theories. Some interviewees highlighted how at times they felt uncomfortable with having to engage “at all costs” with a theory, as in some cases an excessive focus on theorizing and theories was seen as distracting from the goal of addressing some of the world’s pressing challenges, or from the same findings. More generally, ceremonial engagement with theories was rather seen at odds with the possibility to offer an original contribution. Academic excessive “virtuosity” in seeking theorization sometimes was seen as happening at the cost of lack of impact and influence on behaviours and societal change.

“[...] we’re judging based on the citations, how many people read my paper, how many people cite and like a paper. [...] My point is...will this thing significantly change the people’s lives that we’re talking about, the issue we are talking about? Will it make a significant difference that, for example, instead of using Weber I use Marx or Bourdieu in my PhD? Would it significantly change the life of the garment workers in Bangladesh [...]?” (Rho)

Others pointed to the fact that at times papers lack a “real” engagement with relevant theories, and rather appear to embrace certain theoretical perspectives in a ceremonial way, or only because a certain theory is fashionable, legitimized in a certain context.

“[...]a couple of journals if you do not use Foucault or ANT, you know, very well established kind of theoretical framing, you will struggle” (Alpha).

“What is noticed is for some journals, and in some cases, the published papers must ‘sound well’”, that is, “mimick” what is required to be part of the community, use certain words, certain keywords, a well-established theoretical perspective...” (Tau).

Apart from the negative (and demotivating) impact on individual researchers, and their careers, this situation was seen as discouraging innovation, as people *“instead of being original, need to conform to certain “ways of writing” and theorizing”* (Tau)

As such, the use of a theory, rather than being guided by research considerations, appeared to be instrumental in signalling the belonging and allegiance to a certain theoretical worldview, reinforcing the idea that people need to be part of a certain “club” to publish, and comply with the relevant expectations, rather than worrying about the novelty of their findings (which may end up being unsettling for the “established” club).

They also suggested that this may be due to risk-averse approaches taken by editors, who would prefer ‘conservative’ papers, rather than experimenting with new theories and methods: *“[...] there needs to be more openness to new ideas, new methods, new approaches. I was listening to a presentation from a journal editor [...]. Someone asked him a question ... ‘I have not seen any papers using social media data and ethnography in your journal’, [...] they want to see how other journals handle that first and see.”* (Alpha)

Colleagues generally criticized the practice of using empirics simply to “illustrate” theories, with a substantial lack of engagement in linking theories with the empirics: *“Maybe the theoretical lens was used just to label things in the empirical section, which is not shown actually in the quotes. It's more like sort of a theoretical idea that is applied to a case but it's not confirmed by the presentation of the material.”* (Kappa)

Engaging with the literature. The interviewees highlighted that at times authors were more worried about showing their coverage of relevant theories, but did not appear to engage sufficiently with the literature, i.e., with the empirical studies published in a certain area, which may be relevant to be acknowledged and reviewed to correctly problematize an issue, identify gaps so as to justify and problematize the research questions in a paper, and highlight its contribution.

A colleague provocatively observed that as a reviewer sometimes he would like to ask the authors *“There are 20 years of literature written on this topic. Where were you?”* (Ksi).

Some colleagues went further in this critique, suggesting that sometimes the qualitative accounting community does not feel it necessary to engage with quantitative literature, or even literature in other disciplines which may be relevant. This “blindness” to other approaches and streams of literature was seen as detrimental, hampering the full interdisciplinarity of research, encouraging methodological silos and focusing on the narrow contribution to a specific area of research rather than a wider understanding of “real” phenomena. More generally, an increased dialogue across methodological silos and underlying literature was seen as both needed for knowledge advancement, and enriching for the researchers:

“Maybe I would have liked a little bit of more exposure to quantitative as well [...] as opposed to just what was necessary for my research so that [...] I also need to be able to engage with other people's work, able to review and understand” (Omicron)

4.2. Consistency

“Making sense”. A series of ways in which interviewees described the quality of research can be traced back to an “umbrella” concept of consistency. Consistency emerged through descriptions as the attitude of the research to ensure “continuity” and coherence across all the aspects of the research (and the related sections of the paper), starting from the justification of the problem to be addressed, through the review of the

literature, the formulation of the research question and the development of the design, to the explanation and justification of methods, as well as the connections and alignment across literature review, questions, findings and conclusions.

"[W]hether the results are compelling and actually answer the research question people ask, because, sometimes, this is not the case." (Epsilon).

All in all, this was seen as the ability of "keeping promises" in writing a paper and was described more as an attribute of the research (and the related *output*) *per se*, different from the idea of "audit trial" (next section), which was seen as referring to the rigour in the research *process*, and the ability to convey it to stakeholders.

"The paper kind of makes sense together [...] They did what they said they were going to do and... Suppose it's about like, how does the theory and the methodology make sense for the paper and the story that they're trying to tell" (Omega)

Elements of consistency

Consistency was the *fil rouge* underlying several more detailed dimensions, including sampling and unit of analysis, representation, saturation and contextualization.

Sampling, and unit of analysis. According to some interviewees, authors need to explain *"why I am studying or researching that specific organization or that specific group of people, showing that they are the right ones to answer my research question"* (Gamma).

Other interviewees pointed to the importance, to assess consistency, to understand and correctly define the unit of analysis:

"You know, I've got two case authorities, but it's not a case study, it's the case examples [A colleague suggested] 'this is not a case study you need to justify. [...] you've got cases but it's not a case study. That's not the research unit that you're investigating'. So I think I was probably at a point where I was fairly well able to justify my methods" (Iota).

Representation. An important corollary to consistency was related to the extent to which the findings were free from "bias", or they represented the plurality of potential perspectives and instances from the field. While using the term representativeness may convey a false reference to statistical concepts, colleagues were here more worried about plural perspectives from the field to be fully represented, and avoiding selecting only some findings over others.

"How much you can stretch your results, [...] and how much you can provide evidence that your results are not isolated instances but are part of a broader picture. I mean, I'm referring to triangulation, but not only necessarily, and thinking more about providing results that people read and understand, but also that they are not episodic instances selected by you" (Epsilon)

"And sometimes you were able to construct a good relation with some set of people that work on that phenomena, but there are other relevant others that you cannot access, so of course this creates a problem. Because we're listening to just one part of them" (Gamma)

Tradeoffs: saturation, representation and counterfactuals. Interestingly, a few colleagues raised the issue of potential tradeoffs between saturation and representation. Reaching saturation at the expense of representation was not seen as satisfactory, but a strong focus on representation was seen as causing the attainment of saturation much more energy-draining:

" [...] really trying to embrace that kind of heterogeneity of my data which actually made it probably all the more difficult to argue for things like saturation" (Iota)

Interestingly, saturation and overall consistency appeared to be a more common reference point and preoccupation than the search for contradictory findings and counterfactuals. Only a few colleagues, especially working on content analysis and adopting also mixed methods, appeared to refer to the role of counter-evidence in their research:

Omega referred to the importance of contrasting results with extant literature *"Taking a kind of critical position in the sense that... have they thought of different perspectives? Have they been careful to review relevant literature?"*

Theta referred more generally to the importance of searching for contrasting explanations *"We would say we independently assign codes and we only, you know, brought the categories, where we agree with each other and we searched for alternative explanations" (Theta).*

Contextualization. Overall, according to interviewees, an important way to ensure consistency, and address the above tradeoffs was contextualization:

As Omega pointed out *"I think sometimes that the context is everything, so if someone was to have five interviews, but there was a very clear reason why there were only five interviews[...] and then that will make perfect sense to me, so I think it's about explaining the decisions that you've made in your empirical journey and explaining why. [...]quite often in the reviews that I've received [it seems like] context can be ignored and they'll just be a question of 'well why didn't you do more interviews, so why didn't you do more observations, or why didn't you do that'...[...]. In some cases it just feels like they're not paying attention to the context within which I've been writing".*

Assessing and proving consistency

An emphasis on consistency was seen as a useful and pragmatic response to evaluate research considering its idiosyncrasies, the necessary tradeoffs and compromises, relying on human judgement and understanding rather than abstract criteria. In particular, people identified three main elements as essential in assessing consistency of a piece of research, including its attitude to be (i) in line with authors' ambitions, (ii) with the expectations of the audience, and (iii) to be persuasive, through reliance on effective narratives. They also identified typical "flaws" of papers related to lack of consistency.

"A matter of ambitions". According to the interviewees, consistency was not to be evaluated in absolute terms, but rather keeping in mind its ambitions and audience. Some colleagues especially insisted that they would value better a paper with a narrow, simple focus, whose methods, research questions and findings were aligned, than a more ambitious one, but lacking the same consistency.

"A paper has certain ambitions. I will judge the paper based on that. If you wish to generalize your conclusions, I will check whether the results are consistent with this ambition. If you find an exception to the rule, which may happen in case studies, I will read the empirical analysis, and try to see that element of uniqueness [...]I see the final conclusion and go backward trying to understand the logic behind that conclusion" (Sigma)

"The paper which gets accepted is the one which is consistent from beginning to end. It does not promise to go to Mars and then stops at 10 metres from the atmosphere."(Ksi).

Plausibility and audiences. The subjectivity of the assessment of consistency was seen as dependent on its audience, shaping its overall "plausibility". Plausibility was described as an important corollary of the overall

consistency of the research piece, which takes the perspective of the readers in assessing the ability of the authors (and their piece of research) to convince them that what they have done “makes sense”.

“I think the most important thing, in my view, would be that somehow the methodology should plausibly fit to what type of conclusions they develop out of it, so that means, if it's a specific conclusion that applies to maybe just one part of the organization, it's okay to talk with people just from that part of the organization.[...] I think you can really develop interesting arguments from a small set of interviews, but you need to explain why this is sufficient. On the other hand, just by making 30 interviews or more or 60 does not make your argument stronger per se.” (Kappa)

Plausibility was described as being “in the eyes” of the recipients of the research, being a reflection of the degree of homogeneity of a scientific community.

“Plausibility will be judged differently by different audiences as it depends on the audience’s experiences” (Beta).

Persuasion. Several colleagues pointed to the importance of “persuasion” and narratives in convincing audiences about the consistency of research, while highlighting the tacit nature, and ineliminable subjectivity, of the process:

“[in qualitative research] the dialogue between research and audience is more open to a plurality of points of view, and interpretations [...] this openness has the disadvantage that it becomes more subjective whether your research is liked by someone or not...in qualitative research there is a tendency to accept research based on ‘taste’, how much you are persuaded by a certain narration. [...] Persuasion is whether I like the research you are presenting”... (Tau)

“People need a narrative that is clear, that is logical, integrated and basically you need to show that you answer, and you pose, that question in a systematic way. [...] the persuasiveness of your writing leads the reader to be more or less comfortable with the results that you are showing.” (Gamma)

“Qualitative research is about convincing people with “words” and storytelling, not with numbers” (Beta).

“Typical” critical issues with consistency. Colleagues identified typical “flaws” they felt are often found in papers, which may jeopardize the consistency of research, including the lack of explanation on the choice of countries, contexts, organizations, interviewees; the lack of coverage of a plurality of voices, and thus the focus of data only on specific views and perspectives and lack of representativeness of a plurality of voices, when this was “promised” in the declarations around the purpose of the paper; the “thinness” of data provided, and the lack of detail, or the impression that the analysis only focuses on “one side” of the story; the lack of connection between research questions, methods, findings and conclusions.

4.3. Confidence, honesty and the keeping of an “audit trail”

The third important aspect which was seen as central by colleagues in assessing the quality of qualitative research is the confidence that a piece of research can inspire, in terms of the ways in which data collection and analysis are conducted and accounted for. Important conditions for inspiring confidence in qualitative research were considered to be the (i) *richness* and depth of data, the (ii) keeping of an “audit trail”, and (iii) the *honesty* with which the *process* was accounted for. Yet, colleagues also highlighted critical issues concerning the ways in which qualitative research is accounted for, and in particular (i) the risk of it becoming a ceremonial exercise, and (ii) a general lack of focus on the analysis of data. The conditions for ensuring confidence in research, and relevant critical issues, are discussed in the following subsections.

Confidence: richness of data and audit trail

Richness of data was seen as an important element to generate confidence in audiences. This would often require the strong involvement of researchers in the collection of data, and the use of quotations to provide full evidence. Along these lines, limited quotations or shallow ones would be seen as a threat to confidence, as Kappa admitted *"[The paper is based on] quite a few interviews, maybe 60 or 100 or whatever, and just has hardly any quotes in the findings section! That would irritate me [...] So I would guess that they sort of have not actually analyzed the whole material and are just reluctant to put it in there."*

However, a balance needs to be reached between providing evidence, but also avoiding flooding the reader with less relevant details:

"At the same time, I try to fit a lot of things out of the empirical narrative because otherwise it has too many facets and gets too rich and hard to read for the reader, to direct them to the most important and interesting points in relation to the research question of that particular piece of research." (Kappa)

The 'audit trail' and the "limitations". A second element which was considered important by colleagues was a thorough description of the research process, keeping track of what it involved, in terms of sources, data collection and analysis, and providing an "audit trail", as explicit as possible about the process.

For example, a colleague highlighted the importance *"[...] to be able to describe properly what has been done. [...] I got an audit trail, I think that's the word I'm looking for. [...] you need to be able to show clearly what your sources were, how you use the sources, [...] being able to articulate properly what were your sources, to show that you do not rely on just one source"* (Alpha).

Data collection, though lengthy and complex in terms of access, was seen as relatively less problematic to account for. Yet, it was more the overall process, including the inevitable "bumps in the road" and limitations, and the analysis, that were seen as particularly critical to narrate. As Ypsilon pointed out: *"You have to pay a lot of attention while you are doing your research to be sure that you are keeping track of everything you are doing. So, I think that the most important thing is to be assessed okay about all the stages that characterize your research, and if you feel that something was wrong or that something might be, you simply have to recognize this"*.

Accounting for the analysis. A particular aspect most colleagues felt was often missing in the "audit trail" was the full account of the analytical process. Alpha pointed out that in qualitative papers, often *"[authors] tell you a lot about data gathering and, [...] what they don't talk about is the analytical process. And so, how did you move from the data to your findings? I would say we did thematic analysis, but there needs to be more."* Along similar lines Gamma suggested *"One paragraph saying, what's the name that they use, the magic 'content analysis'...' I did the content analysis'... [...] I don't think this is enough"*. This was seen as potentially different from quantitative research, as *"there is several different ways to analyze qualitative data and is not so straightforward as analyzing a quantitative data"*. (Gamma)

Interviewees pointed to possible strategies for tackling the difficulties of accounting for the analysis, through reliance on self-evident, rich analytical material, Kappa: *"I find it hard to incorporate all sort of this chaotic jumps and intuitions, and whatever also takes place in this process into this kind of methodological summary[...]. So my strategy [...] is actually to try to present the actual findings in a way that are sort of speaking for themselves"*.

The “chaotic” and messy nature of the research and analytical process were seen as contributing to the difficulties in writing about the methods adopted, but also as a central feature of qualitative research, to be recognized and accounted for.

Accounting for a the “twists and turns” of a messy process, and serendipity. Many colleagues highlighted that an important signal of research quality was whether a piece of research goes the ‘extra mile’ in accounting for a process which is expected to “messy” and far from linear, and where the research questions and the interpretation of data may be continuously revised as a consequence of exposure to the field.

For example, according to Epsilon: *“Sometimes you'll have an idea of what you are investigating, but the specific research question kind of comes out, or becomes better defined, as you go, so it's a little bit of a moving target.”*

Similarly, Ypsilon highlighted *“Qualitative research requires a continuous re-consideration of the research question. [...]the more you talk with people, the more you can realize that your interpretation or phenomena needs to change, so you have to continuously understand whether the initial research question is that one, or whether you need to change your research objectives.”*

These descriptions were seen in contrast with the rhetoric of handbooks.

“You know I read, I read the methods books and [...] then I got into my project and realized, it was like really much messier than that. It wasn't so straightforward.” (Iota)

The messiness and unpredictability of the qualitative research process were also connected by some colleagues to the serendipitous nature of qualitative research: *“[...] so you start with certain assumptions and go into the field and set up an interview and talk to people and then, if you see that an interesting topic comes up that you haven't got sort of in your mind before, then you can dive deeper into this issue [...]I think that's one of the advantages: to discover phenomena that you sort of haven't covered before, for theory development.” (Kappa).*

Accounting for the research process: honesty or calculative accountability?

Full honesty? Colleagues declared that they would expect full transparency on the actual research process and they suggested that they would not trust a qualitative paper whose methods section presents a very linear, simple process. As Beta put it, *“a messy process would be more consistent with my experiences”, as tThe story, you know, either it has a certain depth in it, or it has certain unexpected twists and turns. If this is not the case, then it means it is a superficial story which hasn't really gone deeper” (Beta).*

Some colleagues recognized that “full honesty” on the real messiness of the process may not always be “prized” positively by editors and reviewers. As a consequences, researchers will tend to simplify and provide an ex-post rationalization, rather than an actual account of what happened:

“What you write in your paper as your methodology isn't necessarily the reality of what happened. [...] so, you end up more developing an ex-post rationalization” (Omega).

“It can be really difficult for qualitative research because qualitative research can be very messy and it doesn't always just make sense in a logical order. You didn't just do X, Y, & Z. Sometimes you change. Sometimes you think ‘I'll do focus groups’, and then the focus groups don't work, and so you do interviews instead, and so that process, that journey, can be quite complex and quite messy and sometimes doesn't seem very legitimate because of that messiness. [...]I suppose I try to think through it in the most step-by-step process that I can, try to be honest, but you can't be too honest when something went wrong, because then the reviewer will think

that can be the problem. Sometimes, maybe, especially as an early career researcher, you can feel like anything that you say like that could just be pounced on, like reviewers are looking for reasons to reject” (Omega)

“Performing” good research: conventions, conforming and complying. The considerations developed in the previous section suggest that respondents highlighted that “full honesty” was seen as important. However, they also pointed to several cases where they did not see it translated into practice, as ceremonial writing of methods sections appeared to be often the case. It was indeed suggested that sometimes methods sections look like a “copy and paste” exercise to justify what was done, and that authors may end up trying to mimick the ways in which a methods section is generally written, to show that research followed conventions, and how things are “normally done”:

“I'm not re-inventing the wheel here [...]trying to understand how other people have done this, and try to do the same thing and because obviously there's various tests you could do, but, again, it is the one that is acceptable within the literature that I need, to kind that ensure that my work is very much similar”. (Omicron)

Interviewees commonly referred to typical signalling behaviours and cues adopted in writing methods sections, for example concerning the use of “magic” numbers (e.g., numbers of interviewees, hours of interviews, cases) and words (e.g., “saturation” and “triangulation”).

“You rely on other people [...] one question I'd ask is, you know how many interviews and then, like everybody says I'll do 30 interviews because it's like a magic number.” (Iota)

“ Sometimes, often you'll get the question of ‘oh, you only did 20 interviews that's not generalizable’, and you run into all of these. You weren't trying to make it generalizable but you run into these battles with reviewers. Even in very qualitative journals you end up in that same cycle”. (Omega)

Even the list of acknowledgements, and co-authoring choices were suggested to have an important signaling value:

“Sometimes you see certain accepted papers...That paper was accepted because this person has visited all the most “enclosed and exclusive” places and conferences of history [...] From the list of acknowledgements sometimes you understand [who you need to know to submit to a certain journal] [...] Or sometimes, to publish, as a second author you need the name of the ‘guru’” (Ksi)

A round peg into a square hole? The risks of “calculative accountability” and the pressures to “look like quantitative research”. Adopting the signaling tactics described above was seen as a way to replace trust among researchers, and reliance on human judgement to assess good research, with automatic “heuristics” and forms of calculative accountability:

“The risk is calculative accountability. [...] I use a number of signals that are seen as signs of validity, such as number of interviews, length of time in the field[...] I am however critical that these are seen as signals of validity, as this should be problematized much more [...]do we need this level of (calculative) transparency?...[...]I think is important in a scientific community, if you send a paper I expect what you say is based on a serious and valid analysis... when I write I expect people to trust me”. (Sigma)

In particular, it was noticed that in order to please editors’ and reviewers’ expectations of conformance to conventions, in some cases, authors would try to make qualitative research look more like quantitative research. Interviewees pointed to cases where reviewers actively encourage authors to change the ways in which research is accounted for, towards a quantitative model, for example by asking to discuss “validity and reliability” of qualitative research.

"I sometimes I see qualitative interpretive papers that they sort of adopt the language of quantitative papers [...] you are pretending to be something that you're not and you sort of trying to squeeze yourself in...it's that sort of expression in English, that he tried to put a round peg into a square hole" (Eta)

"When I send a research report or a submission to a journal, I can see that the reviewer is not aware of the qualitative way of doing research, even the language... I received some comments, saying that I need to include some variables [...] It's not the problem of the reviewer maybe is more probably of the of editor of the journal, they send qualitative research to someone who has no experience or is not interested in in the qualitative work." (Gamma)

"They often say that they're sympathetic to qualitative research, but when it comes down to it, that questioning you on things that are quantitative [...] And you end up having to just sort of make the qualitative research more quantitative" (Omega).

This all suggested that quantitative approaches may still influence the ways in which qualitative research is accounted for, through the pressure to please editors and reviewers and get papers accepted.

4.4. Generalizability?

Interviewees saw generalization as a complex exercise in qualitative research, and different from quantitative research.

"[It] is always difficult at that point where you try to generalize from a qualitative study or theorize a qualitative study, because you always have sort of a limited set of data which is usually not representative for anything [...]. But still, you can sort of theorize it in a way that might inform other similar cases, but you have to build an argument there and this this jump is always tricky". (Kappa)

Some appeared to suggest that generalization is not the main aim of qualitative research: *"They want some level of generalizability and that isn't necessarily what you're trying to do. There's always things that can always be generalized in some way and not be generalized in other ways, [...] but you can't make sweeping generalizations and in the same way like big data." (Omega)*

This was seen also as being related to qualitative research not necessarily focusing on "average" cases:

"The advances are typically made on the boundaries of any discipline or phenomenon. Some people are interested in the average. Other people are interested in the things which are extremely good or exceedingly poor. And that's where qualitative research takes us, not the average." (Delta)

Difficulties were also referred to in terms of analytical generalization, and the related practical relevance of qualitative research:

"The practical implications are much more difficult to judge [...] before any paper has any practical implications, this would require almost to write a different paper[...]sometimes "the ten- line practical implications we write at the end of our paper may make anyone smile [...]There is a joke from [name of a colleague] along the lines of 'this paper is too useful, journals will not like it'". (Ksi)

Yet, colleagues recognized contextualization of research as a possible way to overcome some of the difficulties in generalization:

"[N]ow that I am trying to publish more in international journals and giving more importance, to the context, on the same idea that the context that I find in [name of country] is to some extent similar, or can be found elsewhere too, so the reader at least can relate to his or her own environment or institution" (Gamma).

5. Discussion

5.1. Ambiguity of validation criteria “by design”: assessing research quality as a tacit, social process

Ambiguity of criteria and the tacit norms for assessing research

The interviews pointed to the ambiguous, variable and at times unsettled and unsettling nature of qualitative research validation criteria. This was not generally cast in a positive light, but rather as a matter-of-fact situation that people had to muddle through. Interestingly, interviewees would generally avoid referring to quantitative-derived terminology. Interviewees also used with parsimony the words commonly suggested by handbooks (with some exception for words such as saturation, triangulation, relevance, contribution, richness of data, honesty and plausibility), whereby they preferred to describe in detail the ways in which they assess research, or how they expect rigorous and convincing research to be developed. These findings may help explain some of the results of Kihn and Ihantola (2015), who found quite a significant number of qualitative papers in management accounting embracing validity and reliability criteria, but also an overwhelming majority not necessarily referring specifically to criteria, but rather describing the research process in depth. They are also partly at odds with some literature described in section 2, that appears to have been particularly focused on defining specific (and a multiplicity of) criteria for “validating” research, and which may have paid less attention to the assessment of research as being enacted through concrete practices. In this context, the decision on what “counts” in validating research is left to the authors and reviewers, and is continuously defined and re-defined in the social interactions happening in the research community (see also Lukka and Modell, 2010). Colleagues felt that the recognition of a “good paper” happens mostly in a tacit, implicit way, whereby people in the community can “recognize” what a good paper is, and understand what needs to be done in writing it, probably before rationalizing and making explicit what renders it “good”. As such, assessment of qualitative research appears to be first of all a tacit, implicit process, which, however, can then be explained and justified ex-post.

Signals of research quality or of belonging to tribes?

In a context characterized by high uncertainty, pressure to publish, and anxiety about the ways in which research is evaluated, colleagues pointed to several “signaling” tactics, used by authors to show they comply with editors’ and reviewers’ expectations. This is perpetrated for example by using words such as saturation, figures such as the number or durations of interviews, in picking certain theories and avoiding others to show allegiance to certain world-views, by only embracing literature coming from a selected group of authors, and ignoring other studies, by using acknowledgements to signal belonging to a certain “club”, and by adopting a style of writing which “sounds” like being expression of a certain “tribe”. At times, it may even translate into embracing terminology and ways of writing methods section from the quantitative field.

Especially Early Career Researchers appeared to be worried about the perceived ambiguity of criteria, which conveyed a sense of insecurity in their careers and lives. Some colleagues pointed to the fact that the lack of clarity on validation criteria may actually exacerbate the distinction between those who “know the implicit rules of the game”, who “belong to the club”, and to strengthen the power of editors, reviewers, and more generally gatekeepers to communities. This was also seen as potentially discouraging originality of thinking

and research, forcing people to conform to certain ways of writing and theorizing, at the risk, otherwise, of not been seen as belonging to the right network.

Yet, while some suggested that signaling allegiance to certain academic tribes may be a shortcut for publishing, this was frowned upon, and seen as a seriously dangerous to the extent it also created a sense of exclusion and anxiety for those not belonging to a certain academic “club”, and strengthened the power of the “gatekeepers”, and those who are safely in the “club”. This was in line with Lukka’s (2010: 113) depiction: “While belonging to a certain ‘tribe’ certainly has several positive consequences, it can also lead to dogmatism and dysfunctional defensive tendencies within the various paradigms. More radical new propositions, or critical self-reflections, can easily be labelled as non-issues or straw-men. This is the surest way for a ‘tribe’ to knock down new ideas. The result is that debates are very quickly at risk of becoming excessively political. Also, within each paradigm, there is usually an elite which implicitly assumes that only it has license to innovate radically[...]”.

The findings also provide additional nuances to Humphrey and Gendron’s (2014) concept of *academic docility*, which the authors connect to the ascendancy of journal rankings and the forcing of people to adopt the same ways of thinking and writing used in “top” journals. Interviewees clearly highlighted that, while qualitative researchers may at times feel discriminated within the mainstream, quantitative-oriented accounting community, the qualitative accounting community is not necessarily a fully democratic, open space. On the contrary, interviews suggested that within the “small” qualitative accounting community a “mainstream” is also identifiable, which forces new forms of docility, “theoretical” compliance, and alignment of findings and ways of writing with what is considered acceptable in that community. In the area of qualitative research, academic docility may be even exacerbated by the ambiguity of research assessment criteria.

Finally, the findings provide a strong reminder that the ambiguous, as well as social and processual nature of research assessment, and of proving its quality may be both its main strength and weakness. On the one hand, it can create a democratic space (Parker and Northcott, 2016), providing opportunities for a piece of research to be appreciated for its merits, considering alignment of methods with its specific research questions, purposes, and idiosyncrasies. It may also allow its continuous improvement through a dialogue with the relevant scientific community, and the exposure to useful, open, constructive feedback. On the other, however, embarking in “social processes” for legitimizing one’s research may also imply for the authors to navigate a narrow, club-like, space, having to “play the games” which signal being part of the community (such as adopting certain theories, quoting certain authors, or editors, showing up at specific conferences and seminars, or connecting to the “right” colleagues and institutions). Yet, interviewees suggested that the latter situation, far from desirable, may not be conducive to research innovation.

It is particularly worth noticing the contrast between the depiction of forms of “tacit” recognition of the quality of research, but also its critique as something that may be “manipulated” artfully using signaling tactics. Another inherent interesting contrast was between complaints about the ambiguity of validation criteria, but also recognizing that an excessive reliance on very specific formal criteria may not allow proper research assessment either. Colleagues are generally profoundly aware of these inherent contradictions, and try to address them through heuristics. Interestingly, these mostly relied on human judgement, and experience, as colleagues suggested the importance of looking at the overall quality of work as a matter to be assessed “in practice”, rather than through the application of automatic criteria. At the same time, they were mostly of the idea that more should be done to raise awareness on these issues, to create more opportunities for sharing views, experiences and difficulties in the qualitative accounting arena and to fight opportunistic behaviors, that seriously jeopardize the qualitative accounting community and the wellbeing of researchers. They also pointed to the important roles academic leaders, editors, and choices in the governance of communities and journals can have to address these critical issues.

5.2. What counts as “good” qualitative research: contribution, consistency, confidence and...generalization?

Conceptual contributions have often pointed to a plurality of criteria to be applied in ensuring good quality research. It is also worth noticing that some of the literature appears to have given particular prominence especially to issues related to confidence and consistency, connecting them to the concept of “validation” of research (for example, Golden Biddle and Locke, 1993; Baxter and Chua, 2008; Lukka and Modell, 2010; Messner et al., 2017). Yet others have highlighted the importance of contribution, and relevance in assessing research (e.g., “worthy topic”, Tracy, 2008; or “criticality”, Baxter and Chua, 2008). The interviewees appeared to offer a balanced and comprehensive view that all those dimensions need to be considered jointly to assess the quality of research, though some suggested that at times the evaluation of contribution, and the attainment of generalization, may be more difficult than other elements.

Importantly, colleagues tended to avoid to reduce their experience to words such as the search for validity, authenticity, reliability, or even confidence or consistency, preferring to describe, in their own words, the actual practices and ways in which they assess, ensure and signal research quality, depicting more an iterative, and adaptive process. Yet, for exposition purposes, I had to connect and bring together those experiences, identifying three main “themes” around which their views on what counts in ensuring the quality of qualitative research, including contribution, overall consistency, and confidence and honesty. A fourth, less discussed area, concerned the generalizability of research.

Contribution

Contribution was seen especially in terms of the extent to which it adds to extant theories and literature, and challenge usual, taken for granted ways of seeing things, in line with ideas of *criticality* (Golden-Biddle and Locke, 1993; Baxter and Chua, 2008) and *worthy topic* (Tracy, 2008). More controversy arose on the practical relevance of research, and whether this was a necessary feature of qualitative research, as this was seen as necessary by some colleagues (similar to de Villiers et al., 2019), but also difficult to attain for others (or even at odds with the goal of publishing in “top” journals).

Assessing and demonstrating the contribution in qualitative research was seen as a particularly critical issue, for several reasons. First, identifying the contribution was seen as highly dependent on the context and audiences, compounding the overall ambiguity of qualitative research assessment (on this, also Gendron, 2012). Second, the contribution was seen as dependent on the quality of the engagement with theories and literature. The interviews revealed that engagement with theory was seen at times as an exercise of style, ending up in theories just being used to label concepts, empirics to illustrate exhausted theories, rather than challenging them, or to give way to the exploration of new theoretical perspectives. This was in line with the reflections of Humphrey and Scapens (1996) on the need to avoid mere use of empirics to illustrate theories, and Lukka and Vinnari (2014) suggesting a clear distinction between domain and method theory. Another related issue was over-theorization, as highlighted by Guthrie and Parker (2017), referring to the problem of “theoretical engorgement”. Interviewees also pointed to conservative choices of theories, and the use of empirics to confirm them, being likely to be a reflection of tribal allegiance (as pointed out above). This was seen as being at odds with the search for innovation and new perspectives in research, and the very idea of challenging common wisdom and criticality. Interviewees also felt that an excessive focus on theories distracted from proper reviews of extant literature, i.e., empirical studies addressing the same themes and issues, even from other theoretical and methodological stances. Rather, authors appear to engage selectively with literature, providing further evidence of tribal allegiance.

The above reflections appear to suggest that qualitative scholars may need to rethink if their paper allows a satisfactory balance between practical (or policy) relevance, theoretical contribution, and contribution to literature. So far it seems that the community may have prioritized theoretical engagement (at times to the point of “engorgement”) over engagement with practice and with more general literature. This may be probably the result of pressures to comply with disciplinary (and gatekeepers’) expectations. Indeed, a strong, full engagement with the literature, including relevant literature adopting other methods or approaches to look at the same phenomenon, may be seen as a way to overcome tribal boundaries, and strengthen relevance, and inter-disciplinarity of research, ultimately empowering scholars. From a different perspective, it may also represent a dangerous venturing in the unknown, and uncontrollable by the gatekeepers, signalling a lack of discipline (and docility) by tribe members.

Consistency

Overall consistency was described as the capacity of authors to “keep promises” and ensure coherence throughout the whole research process, as reflected in the writing of the paper. In doing so, the colleagues appeared to implicitly refer to the concepts that in the literature are labelled as authenticity and plausibility (Golden Biddle and Locke, 1993; Baxter and Chua, 2008; Lukka and Modell, 2010; Messner et al., 2017).

Consistency was generally seen as attained by addressing choices and trade-offs between issues of sampling and proper representation and saturation. Representation of multiple voices, and of those voices which were consistent with the research purpose, appeared to be a particularly important preoccupation for many colleagues as it was seen at times as something often lacking in papers. This recalled the concept of “crystallization”, offering views of multiple facets through the use of multiple voices (Richardson, 2000). The search and explanation of contrary evidence did not appear to be a central preoccupation for many colleagues, suggesting that more reflection may be needed to address Lukka and Modell’s (2010) invitation to provide stronger depth when authors embrace explanation purposes.

The above considerations, and the trade-offs emerging from interviews, appear to reflect the co-existence, and the search for compromises, between the different approaches to collect and analyse qualitative data which are described in De Loo and Lowe (2017), and Alvesson (2003), the first being aimed mostly at “understanding” phenomena, and the second mostly focused on “explaining” them. The former will happen through collecting data in human encounters, and to get an understanding of the different experiences, perspectives and meanings attributed by people to phenomena; and analysing them in a way that reflects those plural perspectives and realities, including counter-accounts, offering a basis for understanding them. The latter will similarly rely on multiple sources and views, but more with the aim of gathering information and knowledge about phenomena, to identify possible counter-factual information, and to rely on triangulation and abduction (see also Lukka and Modell, 2012). Whether the aim for interviewees was more to understand, or explain, pluralism of sources, and voices, remained central, though the search for counter-accounts or counter-evidence were seen as clearly time-consuming and in potential trade-off with saturation purposes. The solution to trade-offs was identified in seeing the search for consistency more as a process, and as a matter of dialogue with audiences.

Colleagues appeared to feel that awareness of the inter-subjective nature of the evaluation of consistency and interpreting it as a dialogue between authors and their audiences (including reviewers and editors), would allow better assessment of research, but also require to listen and understand the other’s side, rather than at imposing choices and judging from afar. They especially pointed to the importance, for authors, of using contextualization to explain choices, and to account for the ineliminable compromises faced in the research process. Devoting stronger consideration to contextualization was also seen as important for reviewers, who

may wish to interpret consistency as an adaptable, pragmatic, dialogical and processual approach to assess research that often presents unique, idiosyncratic features. As such, in light of the interviews, consistency should be seen more as a “common-sense” response, strongly based on human judgement, to avoid the risks of relying on specific, quantitative derived, or one-fits-all criteria, as it allows to adapt the evaluation of research to its specific context, aims, methods.

Confidence

The third dimension of research quality assessment, concerning honesty and confidence, appears to reflect an emphasis on issues which, in the literature, have been labelled as dependability, credibility (Lincoln and Guba, 1986), or authenticity (Lukka and Modell, 2010; Messner et al., 2017). Here, however, the colleagues were especially interested in highlighting not only the importance of richness and “thickness” of data, often described in the literature, but especially the messiness of the qualitative research process, and thus the importance for it to be reflected in accounting for it. Their depictions appear to support Vaivio’s (2008: 64) description of qualitative research as “[...]a messy and time-consuming affair. Any academic who has been involved in the production of a case or field-study would probably recall the same difficulties: identifying relevant theory, formulating the research objectives, gaining access, finding the key people, getting your hands on documents, observing without disturbing, drowning in data, being puzzled by conflicting interpretations, trying to find theoretical sense, and writing an argument which is not only novel and intriguing, but also credible.”

Two main critical issues were raised around confidence. The first was the substantial lack of accountability, in many cases, on the concrete ways in which the analysis was conducted, including its serendipitous and messy nature. This may represent an important area for authors to reflect and invest more in the future. Acknowledging explicitly that qualitative research is generally based on a messy process, and that in the very process of entering the field, being part of it, interacting with subjects of the research, and other researchers, writing about the research, and changing our mind about it, we develop new ideas and ways of seeing things, will encourage us to embrace the serendipity of research more, and also recognize the power of writing it as a further important moment of discovery (on this, see also Parker, 1997; De Loo and Lowe, 2012). Second, as more and more colleagues disclose such processes, this will bring about a virtuous circle and ensure that these processes are seen as legitimated, and desirable.

The second critical issue was related, again to the ambiguity of criteria, which forced researchers to adopt certain ways of accounting for their research in a ceremonial way, just to show compliance with generally accepted ways of accounting for research. In their views, this forced researchers to “perform” in the research arena by adopting ways of writing, and explaining the research process, which reflect general expectations from the field, or may even end up forcing researchers to embrace ways of describing methods derived from the quantitative field. All this was seen as driven by pressures to publish in a world where writing according to conventions was seen at times as reducing the actual transparency and honesty of the research account.

These behaviours may suggest that some “bullshitting” (an emerging area of study in psychology, see for example see Turpin et al, 2021) may be encouraged by audiences and reviewers’ expectations or behaviours. Bullshitting has been described as a possible evolutionary response assisting individuals in navigating social systems by impressing (and deceiving) others, without regard for truth or meaning, and is more likely to take when perpetrators believe it will go undetected. Turpin et al (2019), for example, have found that indiscriminately attaching meaningless pseudo-profound bullshit titles to artworks increases their perceived profundity, suggesting bullshit can be used to gain a competitive advantage in any domain of human activities where the criteria for determining who succeeds and fails at least partially relies on impressing others. In the

field of research, this may for example happen by labelling findings with fashionable words, or by providing an account of the research which give an impression of a clean, linear process. This nascent area of investigation in psychology may suggest that researchers may seek to rely on bullshitting strategies if they think they will not be found out, and this may increase the perceived (though not the real) plausibility of their research, and thus, paradoxically, the confidence of audiences. As such, in the future it may be worth exploring, and raising awareness, on the possible roles bullshitting may play in (qualitative) research, and how this could be detected and avoided.

Generalizability

Generalizability was not an agreed upon feature of qualitative research, representing rather a controversial area for qualitative accounting researchers. According to some, generalization was not easy, feasible or even a desirable purpose of qualitative research. However, when colleagues repeatedly pointed to “generalization” not being necessarily the aim of qualitative research, they seemed in most cases to refer to statistical generalization. Parker and Northcott (2016) provide an exhaustive illustration on the multiple meanings taken on by “generalization”, both in theoretical and analytical terms, well beyond simple statistical generalization. This suggests a multiplicity of ways in which (also qualitative) research can generalize and, in light of the results of this exploration, calls for raising awareness in our community about the various forms generalization can take, and especially reject the reductionist idea that generalization is a concept to be only applied to statistical, quantitative analyses.

A further reflection is needed with specific reference to analytical generalizing, as several colleagues appeared to be quite pessimistic about the possibility of academic papers to offer useful generalizations to practitioners. These doubts about the possibility of generalizing to practice appear to be potentially at odds with Parker and Northcott’s (2016 p. 1114) optimistic descriptions of analytical generalizing, and similar considerations by de Villiers et al (2019). They also appear inconsistent with studies in psychology, public policy and media, which, comparing the effects of concrete and abstract information on decision-making (for a review, see Olsen, 2015), show that detailed and personalized information is found to affects judgment more than statistical and general information (“vividness effect”, Zillmann, 2006). The latter is seen as less concrete, being an abstract aggregation of information that cannot be otherwise seen in the real world by the observer. Along similar lines, in media research, popular examples (“vox pop”) are seen as having a stronger effect on the public than poll results (e.g., Daschmann, 2000, p. 160). In political research, episodic frames (single cases, putting “a face” on political issues) have been found to be more emotionally engaging and, in turn, more potent than thematic frames (based on statistical data) (Aarøe, 2011; Gross, 2008).

All in all, this suggests the need to continue the discussion advanced by Parker and Northcott (2017), and to explore further the potential of qualitative research for generalization beyond stereotypical, narrow views of generalization only being possible within quantitative types of research.

6. Final reflections and possible future developments

Qualitative research in accounting has attracted several conceptual contributions on how its quality should be ensured and assessed, validated. This study aimed at bringing to life the perspectives of qualitative accounting scholars, and especially those whose voices are less likely to be heard (early to mid- career colleagues) to explore how they experience processes of validation of qualitative accounting research in their everyday lives of authors and reviewers.

In the face of a multiplicity of criteria, the accounting scholars I interviewed (and especially the early career ones) tended to suggest their feeling uncomfortable, or anxious about the variety and ambiguity of criteria which are applied in practice when their research is assessed. While they were aware of handbooks and extant literature, they tended to rely more on a series of pragmatic heuristics and practical processes to adjudicate on the quality of research. In particular, they emphasized the importance for qualitative accounting papers to articulate a strong contribution, though for some this was seen more in relation to theory, others to literature, and yet others to practice; to show consistency throughout the research process and research outputs; and to provide honest evidence on the richness and quality of data, and of the messiness, complications and limitations of a process, which is rarely perfectly linear and rational, so as to generate confidence in audiences.

The participants in this study appeared to suggest that the multiplicity and vagueness of the criteria of the field will have two main types of consequences. First, proving and assessing qualitative research will not be seen as a static exercise of application of criteria, but rather as a dialogical and socially shaped process, highly dependent on the audience for the paper, and its expectations. The process is also partly “tacit”, in that members of the community will “understand” whether a research piece is “good”. But then reviewers will also make sense of their first impression by checking pragmatically whether this satisfied their expectations in terms of contribution, consistency, and confidence. Second, they felt that, parallel to these “physiological” processes of writing about research, and assessing it, other more “pathological” and undesirable processes were also in place, which encouraged a more ceremonial approach to develop papers, to theorize them, and to write the methods section. This ceremonial approach was driven by the belief that certain ways of theorizing, or writing about methods, are preferred in certain communities (or tribes), allowing to signal the quality of research better, or the belonging to certain “clubs” and “worldviews”, and ultimately rendering publication possible (or easier).

Specific critical issues were raised, representing flaws researchers may wish to redress and consider when developing their research and writing about it. In particular, there was a feeling that at times engagement with theory was excessive, or driven by a desire to please senior academics and editors, and not necessarily allowing the best forms of innovation and theorization. This may even be seen dangerously encouraging forms of confirmation bias (McSweeney, 2021) and thus deserve future attention. Conversely, engagement with literature was seen at times as being too limited, as some authors tend to avoid acknowledging relevant papers, perhaps because they belong to different disciplines, or adopt different methods. This may create potential blind spots in the research, for example not allowing proper identification of research contribution, and hampering the participation in, and contribution to in wider, inter- and multi-disciplinary debates.

A third area which, based on the findings, may require further reflection and intervention, is the consideration of inherent tradeoffs between saturation and the role of counter-evidence or counter-narratives in research, and thus the need to ensure proper representation of voices and crystallization of results, and further explore abductive approaches in research (Lukka and Modell, 2017). Colleagues especially highlighted the importance of contextualization in making those tradeoffs clearer to audiences.

A fourth area which emerged as crucial is the need for qualitative researchers to avoid mimicking quantitative research in accounting for the qualitative research process, and to embrace its ineliminable serendipitous, messy, unpredictable nature, by not being shy in narrating its “twists and turns”, the times research questions changed, findings brought about re-thinking of the whole project, or the unavoidable limitations that any research will present.

Finally, generalization was probably seen as the most controversial issue, and difficult to attain. Interestingly, this was seen as especially true for analytical generalization, at odds with psychological and public policy literature, suggesting that the vividness effect may actually make qualitative evidence particularly salient in shaping audiences’ understanding of phenomena and decisions.

The above considerations suggest several future research avenues for those interested in how we develop qualitative research in accounting. First, we may need to increase our awareness of how psychological elements may shape it, for example exploring the roles of cognitive biases (on this, see also Mc Sweeney, 2021), anxiety and other emotions, as well as “bullshitting” in the work of authors and reviewers. Second, we may need to explore further the connections between the sense of belonging to academic tribes and communities, and issues of research assessment, and publication, from both sociological and psychological perspectives.

On a more practical ground, the findings suggest that, as authors, we may need to explore how we can improve theorizing, and innovate in the theories we use. We may also need to consider if we are engaging sufficiently with practice, and with the overall literature addressing the themes of our interest, even when it embraces different disciplines and methods. In short, we should make sure we are offering a concrete contribution, and not just one that depends on how we set the boundaries of our literature review.

As reviewers (and editors), we may need to reflect on whether we are serving the advancement of our field properly, or whether we, perhaps unwillingly, encourage authors to avoid being fully honest in offering a linear account of the research process, rather than a “true” one; and if we force them into writing a paper which is the one addressing the question we like, rather than supporting them to fulfill their ambitions to write. Finally, if we force them to simplify the messages coming from their papers, because, again, we prize simplicity and saturation over crystallization or consideration of plural evidence and interpretations.

All in all, it appears that these findings invite us to reflect further on the advantages of taking a pluralistic and open stance in our research, from different perspectives. Pluralism may probably need to be strengthened in our communities, providing open fora for discussion, and debate, as a response to the risk of them being seen as clubs, where only members get access to publication. Pluralism would also be central in collecting data, as good research will often require to rely on multiple sources (documents, social media, internal and external documents, interviews, focus groups, etc.), as well as represent multiple voices (representing different perspectives, power positions, interests). Openness is also crucial in embracing the iterative and serendipitous nature of the research process, and also seeking external validation, and considering the views of the potential audiences, in it. Openness and pluralism were also highlighted as important steps in considering theories, and theorizing, as well as literature, whereby our community may need to further explore different, competing or complementary perspectives, including from different theories and disciplines (on this, also Lukka and Modell, 2017). Finally, qualitative research is not always only critical or interpretive. Increasingly, different qualitative and quantitative methodologies are embraced by researchers taking pragmatic, interpretive or (post)positivistic stances. As such, reviewers especially will need to be careful and open in considering the connections between epistemology and methodology, as they may not always be taken for granted or defined by stereotypes.

The present findings are a reflection of interviews which were conducted with early- and mid-career researchers in the UK, Italy, Australia, Spain and Germany, with a variety of ethnic origins. It would be interesting to explore how working in other countries may shape perceptions on research differently, in a world where global trends (and international journals) co-exist with local translations and practices of research evaluation, promotions and hiring of academics. Also, the perspectives represented here may be complemented by exploring the views of editors and senior colleagues.

References

- Aarøe, L. (2011). "Investigating frame strength: The case of episodic and thematic frames", *Political Communication*, 28(2), 207-226.
- Alvesson, M. (2003), "Beyond neo-positivists, romantics and localists: a reflexive approach to organizational research", *Academy of Management Review*, 28(1), 13-33.
- Ahrens, T. and Chapman, C.S. (2006), "Doing qualitative research in management accounting: positioning data to contribute to theory", *Accounting, Organizations and Society*, 31 (8), 819-841.
- Ahrens, T. and Dent, J.F. (1998), "Accounting and organizations: realizing the richness of field studies", *Journal of Management Accounting Research*, 10(1), 1-39.
- Bochner, A. (2000), "Criteria Against Ourselves", *Qualitative Inquiry*, 6(2), 266-272.
- Creswell, J. W. (2007). *Research design: Qualitative, quantitative, and mixed methods approaches* (2nd ed.). Thousand Oaks, CA: Sage.
- Baxter, J. and Chua, W.F. (2008), "The field researcher as author-writer", *Qualitative Research in Accounting and Management*, 5(2), 101-121.
- Daschmann, G. (2000). "Vox pop & polls: The impact of poll results and voter statements in the media on the perception of a climate of opinion", *International Journal of Public Opinion Research*, 12(2), 160-181.
- De Loo, I., & Lowe, A. (2012). "Authoritative interpretation in understanding accounting practice through case research", *Management Accounting Research*, 23(1), 2-16
- De Loo, I. and Lowe, A. (2017), "[T]here are known knowns ... things we know that we know": Some reflections on the nature and practice of interpretive accounting research", *Accounting, Auditing & Accountability Journal*, 30(8), 1796-1819.
- Gendron, Y. (2013). "Rethinking the Concept of Research Contribution". *Accounting Auditing Control*, 19, 133-155.
- Golden-Biddle, K. and Locke, K. (1993), "Appealing works: an investigation of how ethnographic texts convince", *Organization Science*, 4(4), 595-616.
- Gross, K. (2008), "Framing persuasive appeals: Episodic and thematic framing, emotional response, and policy opinion", *Political Psychology*, 29, 169-192.
- Guthrie, J., Parker, L.D., Dumay, J. and Milne, M.J. (2019), "What counts for quality in interdisciplinary accounting research in the next decade: A critical review and reflection", *Accounting, Auditing & Accountability Journal*, 32 (1), pp. 2-25.
- Guthrie, J., Parker, L. D., (2017), "Reflections and projections 30 years of the interdisciplinary accounting, auditing and accountability search for a fairer society", *Accounting, Auditing & Accountability Journal*, 30 (1), pp. 2 – 17.
- Humphrey, C. and Scapens, R. W. (1996), "Methodological Themes; Theories and Case Studies of Organisational Accounting Practices: Limitation or Liberation?" *Accounting, Auditing and Accountability Journal*, 9(9), 86-106.
- Humphrey, C., Gendron, Y. (2014), "What is going on? The sustainability of accounting academia", *Critical perspectives on accounting*, 26 (2), 47-66.
- Kihn, L-A., Ihanola, E-M. (2015), "Approaches to validation and evaluation in qualitative studies of management accounting", *Qualitative Research in Accounting and Management*, 12(3), 230-255.
- Kvale, S. (1995), "The Social Construction of Validity", *Qualitative Inquiry*, 1 (1), 19-40.
- Lincoln, Y. S., Guba, E. G. (1985), *Naturalistic Inquiry*, Beverly Hills, CA: Sage Publications.
- Lukka, K. and Kasanen, E. (1995), "The problem of generalizability: anecdotes and evidence in accounting research", *Accounting, Auditing & Accountability Journal*, 8(5), 71-90.
- Lukka, K., & Modell, S. (2010). "Validation in interpretive management accounting research", *Accounting, organizations and society*, 35(4), 462-477.
- Lukka, K. and Vinnari, E. (2014). "Domain theory and method theory in management accounting research", *Accounting, Auditing & Accountability Journal*, 27(8), 1308-1338.
- Lukka, K., Modell, S. (2017). "Interpretive research in accounting. Past, present, future", In Hoque, Z., Parker, L.D., Coaleski, M., Haynes, K. (eds). *The Routledge Companion to Qualitative Accounting Research Methods*. Routledge, Taylor and Francis, Oxfordshire, UK.

McSweeney, B. (2021), "Fooling ourselves and others: confirmation bias and the trustworthiness of qualitative research – Part 1 (the threats)", *Journal of Organizational Change Management*, 34(5), pp. 1063-1075.

Messner, M., Moll, J. and Strömsten, T., (2017), "Credibility and authenticity in qualitative accounting research", *The Routledge Companion to Qualitative Accounting Research Methods*, Routledge, London and New York, NY, 432-443.

Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook* (2nd ed.). Sage Publications, Inc.

Olsen, A.L. (2015), "The Numerical Psychology of Performance Information: Implications for Citizens, Managers, and Policymakers", *Public Performance & Management Review*, 39(1), 100-115.

Parker, L.D., Northcott, D. (2016) "Qualitative generalising in accounting research: concepts and strategies", *Accounting, Auditing & Accountability Journal*, 29(6), 1100-1131.

Parker, L.D. (1997), "Informing Historical Research in Accounting and Management: Traditions, Philosophies and Opportunities," *Accounting Historians Journal*, 24(2), pp. 111-149

Parker, L.D. (2012), "Qualitative management accounting research: Assessing deliverables and relevance", *Critical perspectives on accounting*, 23(1), 54-70.

Richardson, L. (2000), "Writing: A method of inquiry", In N. K. Denzin & Y. S. Lincoln (Eds.), *The Sage handbook of qualitative research* (2nd ed.), Thousand Oaks, CA: Sage, 923-948.

Seale, C. (1999), *The Quality of Qualitative Research*, London: Sage.

Steinke, I. (2004) "Quality Criteria in Qualitative Research", in Flick, U., von Kardoff, E., & Steinke, I. (Eds.). *A Companion to Qualitative Research*. SAGE, 184-190.

Tracy, S. J. (2010), "Qualitative quality: Eight big-tent criteria for excellent qualitative research", *Qualitative Inquiry*, 16(10), 837-851.

Turpin, M. H., Walker, A. C., Kara-Yakoubian, M., Gabert, N. N., Fugelsang, J. A., Stolz, J. A. (2019), "Bullshit makes the art grow profounder", *Judgment and Decision Making*, 14(6), 658–670

Turpin, M. H., Kara-Yakoubian, M., Walker, A. C., Walker, H. E. K., Fugelsang, J. A., Stolz, J. A., (2021), "Bullshit Ability as an Honest Signal of Intelligence", *Evolutionary Psychology*, April-June, 1–10.

de Villiers, C., Dumay, J., & Maroun, W. (2019), "Qualitative accounting research: dispelling myths and developing a new research agenda", *Accounting & Finance*, 59(3), 1459-1487.

Vaivio, J. (2008). Qualitative management accounting research: rationale, pitfalls and potential. *Qualitative Research in Accounting & Management*, 5(1), 64-86.

Zillmann, D. (2006). "Exemplification effects in the promotion of safety and health", *Journal of Communication*, 56(1), 221-237.

APPENDIX 1

Table 1 – Relevant examples of criteria for assessing qualitative research in general literature

References	Criteria and definitions
Lincoln and Guba (1985)	Trustworthiness requires satisfying four dimensions: dependability, confirmability, credibility and transferability. Dependability refers to a logical, traceable and carefully documented research process, relying on multiple data collection methods and sources. Credibility of research requires sufficient engagement on the field, and data to support claims, strong logical links between observations, categories and materials that allow others to be able to check the claims and decide whether they agree with them. Transferability is about the possibility of finding similarity in other research contexts. Confirmability refers to the idea that research findings and interpretations are linked to data in ways that can be understood and checked by others.
Golden Biddle & Locke (1993)	Convincingness requires authenticity, plausibility and criticality. Authenticity requires the researchers to prove their engagement with the field and the readers to be convinced about it. Plausible research “makes sense” in readers’ eyes, while offering a distinctive contribution. Criticality encourages readers to re-think the taken for granted assumption underlying their research work.
Miles and Huberman (1994)	Trying to bridge qualitative and quantitative criteria, Miles and Huberman (1994: 278–280) identify the following criteria: • Objectivity/confirmability of qualitative investigations • Reliability/dependability/auditability • Internal validity/credibility/authenticity • External validity/transferability/fittingness • Utilization/application/action orientation.

Tracy (2010)	Worthy topic. The topic of the research is “worthy” when it is counterintuitive, taken-for granted assumptions, or challenges well-accepted ideas is often worthwhile. This requires research to be relevant, timely, significant, interesting. Rich rigor. The study uses sufficient, abundant, appropriate, and complex theoretical constructs; data and time in the field; sample(s); context(s); data collection and analysis processes Sincerity requires research to rely on self-reflexivity about subjective values, biases, and inclinations of the researcher(s), and transparency about the methods and challenges. This requires vulnerability, honesty, transparency, and data auditing. Credibility refers to the trustworthiness, verisimilitude, and plausibility of the research findings. It requires thick description, triangulation or crystallization, multivocality and member reflections. Thanks to resonance, research influences, affects, or moves audiences through evocative representation, naturalistic generalizations and transferable findings The research provides a significant contribution when it extends knowledge, improve practice, generates new research, empowers. The research is ethical when it considers rocedural ethics (such as human subjects), situational and culturally specific ethics, relational ethics, exiting ethics (leaving the scene and sharing the research). Research has meaningful coherence when it achieves what it purports to be about, uses methods and procedures that fit its stated goals and interconnects literature, research questions, findings and interpretations in a meaningful way.
--------------	---