Do Researchers Practice what they Preach? An Empirical Analysis of Evaluation Criteria

Ludger Voigt, Johannes Schmidt and Dietrich von der Oelsnitz
Technische Universität Braunschweig, Brunswick, Germany

l.voigt@tu-braunschweig.de
johannes.schmidt@tu-braunschweig.de
d.oelsnitz@tu-braunschweig.de

Abstract: In our project, we take part in the debate about evaluation criteria in qualitative management and organisational science (MOS) research. Since the use of quantitative research criteria to evaluate qualitative research, as well as specific theoretically derived qualitative criteria, has been criticised, we look at this topic from an empirical perspective. Based on a comprehensive analysis of 449 articles published from 2011 to 2021 in five top-tier MOS-journals (Academy of Management Journal, Administrative Science Quarterly, Journal of Management, Organization Science, Strategic Management Journal), we analyse how researchers address the issues of rigor and quality. Focusing on the most common evaluation criteria in the articles (validity and trustworthiness), our results indicate that scholars use qualitative evaluation criteria in a pragmatic way. This shows that, in contrast to the epistemological and methodological differences in theoretically derived evaluation criteria, researchers apply similar techniques to ensure validity and trustworthiness. Given that we only had access to published articles demonstrating the state of the art of using evaluation criteria, the current use of evaluation criteria during the research process remains a ‘black box.’ Therefore, to uncover such implicit evaluation criteria, we suggest a follow-up interview study to gain a deeper understanding of this relevant ‘black box’. In doing so, we aim to explore the intentions and experiences of authors who published qualitative research in top-tier journals on the question, “How to deal with evaluation criteria in research practice as well as during the review process?” As publishing articles in leading journals is very competitive, we argue that these criteria are mainly the outcome of institutional practices, such as the review process and editorial policies. Our follow-up interview study proposes a way to shed light not only on the explicitly mentioned criteria in published articles but also on the underlying implicit criteria in-use during the research process up to publication.

Keywords: qualitative research, evaluation criteria, methodology, rigor, quality

1. Evaluation criteria in qualitative research

Despite the traditional and still dominant approaches of quantitative research within management and organisational studies (Duberley, Johnson and Cassell, 2012), the number of qualitative articles published in leading international journals in business and management has grown over the past decades (Bluhm et al., 2011; Thelwall and Nevill, 2021). This indicates the acknowledgement of qualitative research methods to study relevant issues in management and organisational science (MOS) and shows that qualitative research is accepted as a valuable and valid research approach within the MOS community (Prasad and Prasad, 2002; Cassell and Symon, 2006; Bansal and Corley, 2011). There may also no longer be any differences between quantitative and qualitative research in terms of the chances of getting one’s work published (Cornelissen, 2017). However, according to Bonache (2020), there are still some challenges to overcome during the review process, especially with regard to evaluation criteria for qualitative studies.

In 1985, Lincoln and Guba (1985, p. 290) raised the question, “How can an inquirer persuade his or her audiences that the research findings of an inquiry are worth paying attention to?”. Since then, researchers have offered valuable insights regarding the quality of qualitative studies (e.g. Johnson et al., 2006; Maxwell, 1992; Tracy, 2010). Despite suggestions to use quantitative evaluation criteria for qualitative studies (e.g. Kirk and Miller 1986), due to the nature of social reality as well as the philosophical differences between quantitative (‘erklaeren’) and qualitative (‘verstehen’) inquiry (Duberley, Johnson and Cassell, 2012; Yilmaz, 2013), researchers have claimed that the traditional criteria used to evaluate a quantitative study are not applicable for qualitative research (Easterby-Smith, Golden-Biddle and Locke, 2008). Therefore, scholars such as Lincoln and Guba (1985) and Tracy (2010) proposed specific criteria to evaluate qualitative research according to its epistemological and ontological characteristics. Instead of such ‘one-size-fits-all’ solutions (Johnson et al., 2006), Johnson (2015, p. 320) calls for a “more permissive, pluralistic and reflexive approach to research evaluation”. Following this line of thought, there have been initial attempts to investigate this issue empirically.

Pratt (2008) surveyed authors who published a qualitative study in at least one top-tier North American journal (Academy of Management Journal, Administrative Science Quarterly, Organization Science, or Strategic Management Journal) to investigate which evaluative dimensions were used. Pratt’s study indicates that researchers' experiences regarding qualitative evaluation criteria focus on three areas of tension: “(a) the need
to both break away from and be firmly embedded in extant theory; (b) the need to both provide enough data so that a reader can draw conclusions, while also providing enough interpretation of that data; and (c) the need to be both detailed and transparent with regard to one’s methods, while also fitting one’s research into the format of a top-tier North American journal” (p. 485). Pratt (2008, p. 495) describes these tensions metaphorically as “fitting oval pegs into round holes”.

Based on published articles in two International Business (IB) journals (Journal of International Business Studies and Journal of World Business), three expert interviews, and sample reviews, Welch and Piekkari (2017) analysed what the IB-scholarly community considers as ‘good’ qualitative research. As an alternative to the rigid application of a single set of qualitative evaluation criteria and procedures, they argue for a more pluralistic, contextual approach that reflects the inherent characteristics of IB as a research field.

Drawing on these two studies, which examined the requirements that evaluation criteria must meet to assess rigor and quality in qualitative studies, we go a step further by exploring the use of concrete evaluation criteria. However, to the best of our knowledge, no empirical studies have investigated what qualitative evaluation criteria researchers use to ensure rigor and quality in journal articles. Therefore, our research project aims to contribute to the growing literature on empirically derived qualitative evaluation criteria, which we have briefly reviewed. Specifically, we aim to answer the question of how qualitative researchers use evaluation criteria within a peer-reviewed MOS journal article, as it is considered the most important medium for high-quality science communication (Huff, 1999).

Our research project has the potential to make several contributions to the field. First, it can help develop a more nuanced understanding about the role of evaluation criteria in qualitative management research. By examining the ways in which researchers report on such criteria, our project can shed light on the various ways in which criteria are employed and their relative importance in assessing the quality of qualitative research. This may help advance our understanding about the philosophical assumptions of qualitative research and the underlying criteria. Second, our research project potentially contributes to both the development and the use of best practices of evaluation criteria in qualitative research. This can lead to an improvement in the rigor and quality of qualitative studies. Third, our project can also help to establish qualitative methods as a rigorous and respected research method, which can contribute to its wider acceptance and integration into various research fields. Overall, our research project has the potential to make a significant contribution to the advancement of qualitative research methodology and its broader impact in the academic community.

2. Method

2.1 Sample

Our sample included articles published in five top-tier MOS-journals. Specifically, we focused on the Academy of Management Journal (AMJ), Administrative Science Quarterly (ASQ), the Journal of Management (JOM), Organization Science (ORSC), and the Strategic Management Journal (SMJ). These journals are consistently ranked as top journals in the field (Podsakoff et al., 2008) and are highly influential (Pratt, 2008; Scandura and Williams, 2000). Moreover, previous studies have used these journals to identify the critical evaluative dimensions necessary for publishing an academic article (Pratt, 2008) and to address self-reported limitations and future research directions found within the articles (Brutus, Aguinis and Wassmer, 2013). Our sampling process focused on empirical contributions using only qualitative methods and data. We analysed all articles from 2011 to 2021. Our final sample included 449 articles: 155 AMJ, 74 ASQ, 14 JOM, 156 ORSC, 50 SMJ.

2.2 Coding Procedures

In the first step, we coded each article to assess the research approach. The research approach was determined from the reported methodology, data collection methods, and procedure of analysis. Therefore, we coded each article on the following research approaches: narrative research, phenomenology, grounded theory, ethnography, and case study (Creswell and Creswell, 2018).

In the second step, to code the evaluation criteria, we conducted a narrative literature search on theoretically derived evaluation criteria. Given that case study, ethnography, and grounded theory are the most common used research approaches in the articles analysed, we focused not only on universalistic evaluation criteria but also on theoretically derived criteria specified in case study, ethnography, and grounded theory to link the
evaluation criteria with philosophical assumptions. We included evaluation criteria from the following sources in our coding scheme.

- Universalistic: Lincoln and Guba (1985); Maxwell (1992); Altheide and Johnson (1994); Whittemore, Chase and Mandle (2001); Tracy (2010); Miles, Huberman and Saldaña (2020)
- Case study: Eisenhardt (1989); Yin (1994/2018); Klein and Myers (1999)
- Ethnography: Golden-Biddle and Locke (1993); Hess (2001)
- Grounded theory: Corbin and Strauss (1990); Locke (2001); Charmaz (2006/2014)

3. Results

Validity and trustworthiness were the most commonly used evaluation criteria in the journal articles analysed. While a variety of evaluation criteria were mentioned, in the following, we focus on these two criteria to provide detailed insights.

3.1 Validity

Our analysis showed that the empirically derived understanding of validity represents an umbrella term including different techniques in data sources, findings, interpretations, and the context of generalisability. To ensure the validity of data sources, many authors used data triangulation. For example, in their case study, Helms and Patterson (2014) triangulated interviews and observations with archival data. Monin et al. (2013, p. 264) mentioned that “qualitative process analysis involves methodological challenges related to validity”. To deal with it, they had regular meetings with the companies they analysed to discuss and validate their findings. This kind of member check to ensure validity was discussed in many papers as a validity check. In another case study, Ozcan and Gurses (2018, p. 1795) “did a final validity check by conducting retrospective interviews with two prominent actors who played a fundamental role”. In addition to such member checks, Ravasi, Rindova and Stigliani (2019) discussed their findings, based on combined procedures for grounded theory and multiple-case study, with colleagues as peer debriefing to check the validity of their interpretations. Addressing validity in data analysis and interpretation is not only found in case studies, but also in ethnographic research. To increase the validity of their findings, Kremser and Blagoev (2021) conducted “an insider-outsider approach to data analysis [...]”, whereby the first author assumed the role of a neutral outsider who interrogated and critically questioned the second author’s initial interpretations of the data as well as emerging theoretical ideas”. In another ethnographic study from De Rond and Lok (2016), “the co-author played an important reflexive role by actively questioning the validity of emerging insights in ongoing discussions about data interpretation and theory development.”

3.2 Trustworthiness

Lincoln and Guba (1985) suggested four criteria to achieve trustworthiness. To ensure credibility, a commonly used technique was triangulation. While Caza, Moss and Vough (2018) used multiple sources of data to increase the trustworthiness of their findings by following a grounded theory approach, Wright, Zammuto and Liesch (2017, p. 214), who conducted a case study, “compared iteratively within and between [their] different sources of data and [their] emergent categories and challenged each other’s perspectives to arrive at the most credible interpretation of our data”. The most commonly used technique to increase trustworthiness in general, and credibility in particular, was member checks. In their case study, Ben-Menahem et al. (2016, p. 1316) “discussed [their] emerging model with several key informants” and Danneels (2011, p. 4) “exchanged emails with some of the interviewees” to confirm his interpretations of a historical case. According to Lincoln and Guba (1985), Garg and Eisenhardt (2017, p. 1835), who conducted a multiple-case study, “engaged in peer debriefing with colleagues”. Transferability was mentioned in grounded theory articles. While Harrison and Rouse (2014, p. 1278) focused “on describing how [their] context is similar to others in order to allow for transferability”, Schinoff, Ashforth and Corley (2020, p. 1418) described “the unique characteristics of [their] sample”. An illustration of a dependability audit was given by Harvey and Kou (2013, p. 355), who also followed a grounded theory approach: “the second author, who had not been involved in coding at that point, was trained in using the coding scheme and performed two reliability checks”. To ensure confirmability, in his case study, Carton (2018, p. 334) “asked six scholars who publish inductive research in top management journals to audit the data presented in the paper as well as [his] interpretations of them to check the coherence of the theoretical model”.
4. Discussion

Our study focused on criteria to evaluate qualitative research in five journals in the MOS field (AMJ, ASQ, JOM, ORSC, SMJ). All 449 articles analysed contained evaluation criteria. To gain detailed insights into best practices, we focused on the two most commonly used criteria: validity and trustworthiness. Our analysis provided two key findings.

First, our results indicate that the evaluation criteria used in the articles are to some extent linked to the philosophical assumptions of the research approach. Therefore, validity, as a positivistic criterion used to evaluate qualitative research, was mainly addressed in case studies, which is in line with the literature on case studies. Thus, Yin (2018) suggested logical tests, including validity and reliability, to judge the quality of a case study, and Eisenhardt (1989) underlined this by arguing that the aim is to build a testable and logically coherent theory. Nevertheless, techniques to ensure validity were also included in ethnographic studies. Trustworthiness was mainly found in grounded theory articles. Using trustworthiness and the specific criteria of credibility to evaluate grounded theory articles is in line with the literature. Thus, Corbin and Strauss (1990), Locke (2001) and Charmaz (2014) proposed credibility as a fundamental criterion to judge the quality of a grounded theory article. Nevertheless, researchers of ethnography and case study also referred to trustworthiness to ensure the rigor and quality of their articles.

Second, our analysis suggests that the techniques used to ensure validity and trustworthiness in qualitative research are closely related. Although Lincoln and Guba (1985) argued that positivist criteria (such as validity) are inappropriate to evaluate a qualitative study and therefore developed criteria based on a naturalistic point of view, our findings do not clearly reflect this difference between these two opposing perspectives. Rather, the techniques used to ensure validity and trustworthiness appear to be quite similar. For example, scholars mostly used triangulation, member checks, and peer debriefing in the same way to achieve either validity or trustworthiness.

Given that we only had access to published articles demonstrating the state of the art of using evaluation criteria, the underlying use of evaluation criteria during the research process remains a ‘black box’. However, both key insights mentioned in this discussion are insufficient to answer our research question. Therefore, to uncover implicit evaluation criteria, we propose publishing an academic article as a socially constructed process. To shed light on this process, it would be necessary to gather information directly from authors who have published at least one qualitative study in a top-tier journal.

“Finally, my deductive editors often like propositions, and if so, I usually provide them” (Gehman et al., 2018, p. 296). This statement by Kathleen M. Eisenhardt, made at the 2016 Academy of Management Annual Meeting, highlights the influence of reviewers on the proposition output of inductive research. In terms of evaluation criteria, researchers may start with a research design that is consistent with the epistemological and ontological positions. Nevertheless, the research and review processes are influenced at different stages. Researchers submit their study to conferences and receive feedback, interact with colleagues during the research process, and finally report to reviewers to publish their article in peer-reviewed journals. The feedback researchers receive may encourage them to modify their initial positions. As publishing articles in leading journals is highly competitive (Aguinis et al., 2020), reviewers and editors may make additional demands that require further modifications. Therefore, it is essential to avoid possible misjudgements about an article’s rigor and quality, which might be based on different philosophical understandings between authors and reviewers (Bonache, 2020).

In contrast to Pratt (2008), who used a survey design to study qualitative evaluation criteria, we suggest a qualitative interview study to provide a more detailed and thick description of this ‘black box’. In doing so, we aim to explore the intentions and experiences of authors who have published at least one qualitative study in the articles analysed by us on how to deal with evaluation criteria in research practice as well as during the review process. A first attempt was made by Welch and Piekkari (2017), who conducted three expert interviews in their study of evaluative criteria in IB. To conceptualise our follow-up interview study, we suggest the framework shown in Figure 1.

On the one hand, the theoretically derived criteria to evaluate qualitative studies are described in textbooks and methodological articles. On the other hand, researchers describe explicit evaluation criteria in their articles, as we analysed in our study. These empirically derived criteria can be seen as “espoused theory” (Argyris and Schoen, 1974) which refers to the worldview and values that an individual claims to follow. According to Maxwell (1992), we argue that these criteria are mainly the outcome of institutional practices, such as the review process.
Therefore, we assume that there is a gap between the explicitly mentioned criteria in published articles and the implicit evaluation criteria used during the research process. According to Argyris and Schoen (1974), we call these implicitly used evaluation criteria “theory-in-use” in our framework; i.e., the worldview and values reflected in the behaviours that drive individual actions.

![Diagram](image)

**Figure 1:** Framework to uncover qualitative evaluation criteria “in-use”

### 5. References


