

Approaches to Validation and Evaluation in Qualitative Studies of Management Accounting¹

Lili-Anne Kihn & Eeva-Mari Ihantola,
School of Management, University of Tampere, Finland.

Published in *Qualitative Research in Accounting and Management*, Vol. 12, No.3, 230-255.

Abstract

Purpose: This paper addresses the reporting of validation and evaluation criteria in qualitative management accounting studies, which is a topic of critical debate in qualitative social science research. The objective of this study is to investigate the ways researchers have reported the use of evaluation criteria in qualitative management accounting studies and whether they are associated with certain paradigmatic affiliations.

Design: Building on the work of Eriksson and Kovalainen [Eriksson, P. & A. Kovalainen (2008) *Qualitative Methods in Business Research*. London, Sage], we examine the following three approaches: the adoption of classic concepts of validity and reliability, the use of alternative concepts, and the abandonment of general evaluation criteria. Content analysis of 212 case and field studies published during 2006-February 2015 was conducted to be able to offer an analysis of the most recent frontiers of knowledge.

Findings: The key empirical results of this study provide partial support for the theoretical expectations. They specify and refine Eriksson and Kovalainen's (2008) classification system first by identifying a new approach to evaluation and validation and second by showing mixed results on the paradigmatic consistency in the use of evaluation criteria.

Research limitations: This paper is not necessarily exhaustive or representative of all the evaluation criteria developed and used, we focused on the reporting of explicit criteria only and the findings cannot be generalized. Somewhat different results might have been obtained if other journals, other fields of research, or a longer period of time were considered.

Practical implications: The findings of this study enhance the knowledge of alternative approaches and criteria to validation and evaluation. The findings can aid both in the evaluation of management accounting research and in the selection of appropriate evaluation approaches and criteria.

Originality/value: This paper presents a synthesis of the literature (Table 1) and new empirical findings that are potentially useful for both academic scholars and practitioners.

Keywords: Evaluation criteria, Methodology, Management accounting research, Qualitative research, Reliability, Validity, Validation

JEL Classification: B410 Economic Methodology

Article Classification: Research paper

¹ **Acknowledgments:** We thank the participants of the following conferences for their valuable comments: the 38th Annual Congress of the European Accounting Association, Glasgow; the Nordic Accounting Conference, Copenhagen Business School; the 8th EIASM Conference on New Directions in Management Accounting, Brussels, Belgium; the European Network for Research in Organizational and Accounting Change Conference, Jyväskylä, Finland; the 34th Annual Conference of the European Accounting Association, Rome, Italy; the International Academy of Business and Public Administration Disciplines Winter Conference, Orlando, Florida; and the Nordic Academy of Management Conference, Stockholm, Sweden. We are also grateful for the research assistance of Lauri Lepistö and Pyry Mäkinen. The corresponding author gratefully acknowledges the financial support of the Academy of Finland and the Foundation of Economic Education. Finally, we gratefully thank Editor Deryl Northcott and the two reviewers for many helpful comments and questions.

1. Introduction

Some social scientists have had critical debates about validation and evaluation (hereafter evaluation) in qualitative research (e.g., Garratt & Hodkinson, 1998; Schwandt, 1996; Smith & Hodkinson, 2005). It has even been argued that interpretive research is in a “crisis of validity” as there is a lack of clarity or consensus about how it should be evaluated and about its paradigmatic consistency (see Lukka & Modell, 2010, 463 for a review). According to Lukka and Modell (2010) the crisis of validation in interpretive management accounting research consists of three elements: 1) an acceptance that the interpretive research cannot be validated based on traditional (positivist) criteria; 2) that validity criteria keep proliferating; and 3) that this often leads to criteria being used in a way that is inconsistent with the strongly subjectivist philosophical premises of interpretive research. Some have been concerned about “the fallacy of pursuing stable validity criteria” and “the obsession with criteriology” and have rather seen validation as an ongoing process that is “far from reducible to the technicalities of following pre-specified criteria presumably minimizing various observational biases” (see Lukka & Modell, 2010). Others argue that the use of explicit evaluation criteria throughout the research process can help to guide the research in a direction that ensures good quality, increases the transparency of the research, and provides the means to highlight its strengths and limitations (see Eriksson & Kovalainen, 2008, 290).

While many and in part overlapping evaluation criteria have been developed, with a few exceptions, there has been relatively little discussion about whether to use them and which ones to use in qualitative management accounting research (Ahrens & Chapman, 2006; Baxter & Chua, 2008; Lillis, 2006; Lukka & Modell, 2010; McKinnon, 1988; Parker, 2012; Ryan et al., 2002). There is even less knowledge concerning the current state of evaluation in the qualitative studies of management accounting (cf., Lukka & Modell, 2010). For example, we do not know much about how evaluations are made. Furthermore, we do not have systematic evidence about how differences in paradigmatic affiliations affect the evaluation of management accounting studies. As the use of qualitative research methods has increased in management accounting research (Merchant & Van der Stede, 2006), the analysis of the above-mentioned questions has become increasingly important.

The objective of this study is to investigate the ways that researchers have reported the use of evaluation criteria in qualitative management accounting studies and whether they are associated with certain paradigmatic affiliations. Hence, our study combines different approaches to evaluation and different traditions of research to improve the understanding of the validation and evaluation of qualitative research. We examine these questions empirically using a recent sample of peer-reviewed articles published in some of the top accounting journals.

In the remainder of this paper, the literature review of the evaluation approaches and criteria is presented in Section 2, the research methods and dataset are in Section 3, the empirical findings are in Section 4, and the discussion and conclusions are in Section 5.

2. Approaches and criteria for qualitative research evaluations

In this study we investigate the ways researchers have reported the use of evaluation criteria in qualitative management accounting studies. We discuss below how evaluation criteria can be used and reported in light of earlier social science and business research. Qualitative research methods are defined to include single and multiple case studies and cross-sectional field studies (Lillis & Mundy, 2005, 132). In the following Table 1 and in the text that follows it, we distinguish between the following three approaches, as outlined by Eriksson and Kovalainen (2008, 291), for using evaluation criteria in qualitative business research (see also Baxter & Chua, 2008, 102; Lukka & Modell, 2010, 463):

1. adopting the classic criteria of good quality research, which originally stem from quantitative research
2. adopting alternative but common criteria for qualitative research
3. abandoning the idea of common evaluation criteria.

Place Table 1 here.

We also investigate whether the different approaches to evaluation are associated with certain paradigmatic affiliations. We classify the main methodological approaches of qualitative accounting research according to Hopper and Powell's (1985) taxonomy of accounting research as mainstream, interpretive, and critical (see also Chua, 1986; Ryan et al., 2002; Pihlanto, 2003). Mainstream accounting research is primarily concerned with the functioning

of accounting (Ryan et al. 2002, 41). Such work is known as functionalism in Burrell and Morgan's (1979) framework. According to the assumptions of the mainstream approach, theory and observations are independent of each other and observations may be used to verify or disprove a theory. A hypothetic-deductive form of scientific explanation is accepted. Empirical reality is assumed to be objective and external to the subject, and human beings are seen as passive objects, not as makers of social reality. Individuals and firms are assumed to be rational utility maximizers. Accounting is seen to specify means, not ends, and extant institutional structures are accepted. (Chua, 1986, 611)

Interpretive accounting research is concerned with understanding the social world and includes work that seeks to understand the social nature of, and changes in, accounting practices (Ryan et al., 2002, 42). The interpretive methodology attempts to describe, understand, and interpret the human intentions. Ethnographic work, case studies, and participant observation are encouraged and actors are studied in their everyday world. Social reality is assumed as emergent, subjectively created, and objectified through human interaction. (Chua, 1986, 615)

Critical accounting research seeks to provide a social critique and promote radical change (Ryan et al., 2002, 42). The critical (or radical) perspective assumes that empirical reality is characterized by objective, real relations, which are transformed and reproduced through subjective interpretation. Fundamental conflict arises because subjective interpretation is obscured by injustice and ideology in the social, economic, and political domains. The concern of the critical research is to construct an understanding of the social and economic world with a critical imperative: the identification and removal of domination and ideological practices. Critical research commonly uses historical and ethnographic case studies. (Chua, 1986, 622)

2.1 Adopting the classic criteria

The first approach to validation and evaluation uses the classic concepts of validity, reliability, and generalizability, which originally stem from quantitative research (see the first row of Table 1). *Validity* generally refers to "the extent to which conclusions drawn in research give an accurate description or explanation of what happened." However, in qualitative research, it is used in a rather differently defined way; it means that "the aim is to provide research with a

guarantee that the report or description is correct” (Eriksson & Kovalainen, 2008, 292). This means the reports and descriptions are true and accurate representations of the phenomenon referred to and are certain (i.e., backed by evidence) (Schwandt, 2001 in Eriksson & Kovalainen, 2008, 292). *Reliability* generally refers to “the extent to which a measure, procedure, or instrument yields the same results on repeated trials” (Eriksson & Kovalainen, 2008, 292). That is, it has to do with the minimization of error and bias (Eriksson & Kovalainen, 2008, 292). *Generalizability and generalization* refer to making “a broad statement or belief based on a limited number of facts, examples, or statistics.” Theories, for example, are generalizations that imply that an idea has general application (Eriksson & Kovalainen, 2008, 306).

In the context of cross-sectional management accounting field studies, Abernethy et al. (1999) and Lillis (2006) refer to the following forms of classic validity: internal, construct, and external. In qualitative research these concepts, which originated in quantitative research, are typically used with rather differently defined meanings. They are adapted to fit qualitative methodology with different connotations and meanings than apply in statistical research (Eriksson & Kovalainen, 2008, 291-292). This is because it is not possible to verify the validity or reliability of qualitative research by re-running the statistical tests on the quantitative data collected (Parker, 2014, 24). Therefore, *internal validity* generally stands for accurate descriptions of all the relevant phases of research; that is, whether we can draw valid conclusions from a study (Ryan et al., 2002, 141). *Construct validity* refers to how theoretical constructs are operationalized (Hair et al., 2006, 707; Lillis, 2006). The classic criterion *external validity* used in qualitative research does not refer to statistical generalizability but to generalizability to theory and practice (Lillis, 2006).

Referring to Brewer (2000), Eriksson and Kovalainen (2008) conclude that one can find these evaluation criteria originally used in quantitative studies suitable for one’s study when one’s focus is on objective indicators rather than subjective meanings of data. According to Hopper and Powell’s (1985) taxonomy of accounting research, a focus on objectivism and functionalism refers primarily to the mainstream accounting research and partly to the critical accounting research. The above literature review leads to the theoretical expectation of reporting classic evaluation criteria primarily in mainstream studies of management accounting.

2.2 Adopting alternative criteria

According to Baxter and Chua (2008, 102), much qualitative management accounting field research does not ascribe to the dominant assumptions of mainstream research. Eriksson and Kovalainen (2008, 294) recommend replacement of the traditional notions of validity, reliability, and generalizability with alternative evaluation criteria developed for qualitative research if the research relies on relativist ontology (emphasizing that there are multiple realities) and subjectivist epistemology (suggesting that the researcher and the participant jointly create understandings). According to Hopper and Powell's (1985) taxonomy of accounting research, such an approach refers primarily to interpretive research and partly to critical accounting research. Here, the researchers and their procedures during the research process become the objects of evaluation (Creswell, 1994, 158). The second row of Table 1 aims to synthesize the alternative criteria discussed in social science and business research. Each of them is elaborated upon below.

In his pioneering research, Näsi (1979) developed the following new concepts to evaluate qualitative research: doctrinal relevance, philosophical and methodological validity, internal logic, and empirical applicability. According to Näsi (1979), a crucial question in *doctrinal relevance* is how the theory used has been built on the basis of earlier studies. The reader's knowledge of the existing literature is important in imposing a disciplinary context that checks for the plausibility of the relationships from the fieldwork (Campbell, 1988 in Ahrens & Chapman, 2006, 836). *Philosophical and methodological validity* is concerned with the appropriateness of the ontology, epistemology, and methodology for the research objectives. The methods used need to be scientifically useful and justifiable (Näsi, 1979, 292; Ahrens & Chapman, 2006, 822). *Internal logic* relates to whether the argumentation is clear and logical and congruent with the interpretation. (Näsi, 1979; Lillis, 2006, 466-467). It entails accurate descriptions of all the relevant phases of research and asks whether any contradictions occur in the analysis or logic (see also Golden-Biddle & Locke, 1993). The *empirical applicability* of findings has to do with whether theoretical claims are substantiated in the real world (Näsi, 1979, 292).

Lincoln and Guba (1985) substituted validity and reliability with the *trustworthiness* criteria, subsuming the following four dimensions: *dependability*, *conformability*, *credibility*, and *transferability*. *Dependability* refers to the logical, traceable, and carefully documented

research process. *Conformability* refers to the idea that research findings and interpretations are linked to data in ways easily understood by others. *Credibility* of research requires authors to demonstrate familiarity with the topic and have sufficient data to make claims, strong logical links between observations and categories, and materials that allow others to decide whether they agree with the claims. *Transferability* is a way of re-thinking the notion of generalizability in Lincoln and Guba's (1985) framework. It is about whether some sort of similarity could be found in other research contexts (Eriksson & Kovalainen, 2008, 294).

In their finance and accounting book, Ryan et al. (1992 and 2002, 155-156) replace the traditional criterion of internal validity with the notion of contextual validity, they replace statistical (external) generalizability with transferability (see also Lincoln & Guba, 1985), and they replace reliability with procedural reliability. *Contextual validity* refers to the credibility of case study evidence and the conclusions drawn (Ryan et al., 2002, 155-156). According to Ryan et al. (2002, 149-150, 155) transferability refers to theoretical generalizability. They identify two forms of *theoretical generalizations* in interpretive research. First, case studies in new or different contexts can generalize and extend the theory to a wider set of contexts. As case studies seek to apply theories in new contexts, the theory is likely to be refined and/or modified, and through this process the theory is generalized. Second, the theory can extend the applicability of the case study findings to other contexts. The theory thus argues that the findings of a case study can be applied in other contexts.

Procedural reliability means that the researcher has adopted appropriate and reliable research methods and procedures and this has been demonstrated. Meticulous documenting and reporting allows the reader to assess how the researcher has collected, produced, and interpreted the data. Procedural reliability also requires a good case study design, including unambiguous research questions, a comprehensive research plan, a coherent set of field notes on all evidence, and a fully-documented case analysis (Ryan et al., 2002, 155).

Traditional notions of generalizability have also been replaced with criteria such as whether the results can be *extended to a wider context* (Eriksson & Kovalainen, 2008, 293-204), have *practical usefulness*² (Arbnor & Bjerke, 1977), have *contextual* and/or *constructive*

² Arbnor and Bjerke (1977) argue that although this kind of validation is of uncertain value, it can occasionally be useful and should, therefore, not be rejected.

*generalizability*³ (Lukka & Kasanen, 1995), and *analytical generalizability* (Yin, 2009).⁴ It has also been noted that case studies can be *useful* in, for example, understanding practice (Golden-Biddle & Locke, 1993; Lillis, 2006; Mäkinen, 1980). Therefore, we address questions such as whether the results are related to the actual knowledge and thus benefit the scientific field in question and whether they benefit organizations.

According to Golden-Biddle and Locke (1993), ethnographic texts need to be *convincing* and have authenticity, plausibility, and criticality. Through *authenticity*, readers accept that the researcher was indeed present in the field and grasped how the members understood their world. Through *plausibility*, readers accept that the findings make a distinctive contribution to issues of common concern. Through *criticality*, readers endeavor to re-examine the taken-for-granted assumptions that underlie their work. Some of the methodologically-oriented management accounting studies have also analyzed these criteria. Ahrens and Dent (1998, 15), for example, discussed whether studies were *convincing* and demonstrated a depth of knowledge about their respective organizations, showed a sensitivity to prior theoretical constructs, and provided theoretical insights suggestive of theoretical development (see Ahrens & Dent, 1998, 15). Ahrens and Chapman (2006) discussed the *plausibility* of qualitative field studies and the plausibility of the relationships developed from the fieldwork. They also discussed whether a researcher's key theoretical points are *convincingly* presented, whether an argument is entirely convincing, and whether a researcher has developed a convincing account of the ways in which meanings and purposes relate to patterns of activity. In their paper Baxter and Chua (2008, 113) followed Golden-Biddle and Locke's (1993) criteria closely. They drew

³ Lukka and Kasanen (1995, 83, 85), for example, distinguish between contextual and constructive generalizations. In a *contextual generalization*, a case study provides the possibility of widening the results beyond the primary observations by efficient triangulation of the data. This generalization is contingent upon an understanding of the historical and institutional context of the case. *Constructive generalization* arises when a real business problem is solved in one or more case organizations by developing a novel solution that also makes a scientific contribution (Lukka & Kasanen, 1995, 83). This generalization relates to constructive case research and relies on linking the solution concept to the existing literature and the diffusion of innovation. Consistency typically means that another person should be able to examine the work and come to similar conclusions (Douglas, 1971; Eriksson & Kovalainen, 2008, 292; Ryan et al., 2002, 155).

⁴ According to Yin (2009, 15): "...case studies are generalizable to theoretical propositions, not to populations or to universes. ... [The] goal will be to expand and generalize theories (analytical generalizability) ..."

attention to the aesthetics associated with how non-positivistic qualitative management accounting field research convinces the readership and argued for a more aesthetically informed consideration of the “goodness” of non-positivist qualitative management accounting field research. In their view different forms of literary styling contribute to qualitative research with a sense of authenticity (“been there”), plausibility (“it makes sense”), and criticality (“the imaginative possibilities that field research may provoke”) (Baxter & Chua, 2008, 104). Literary convincingness is expected to be achieved by a hybrid accomplishment in which both scientific and literary forms of authority are textually instantiated. Finally, the focus of Lukka and Modell’s (2010) study was on the manifestations of the *authenticity* of research findings due to the rich descriptions and the *plausibility* or credibility of the explanations establishing validity. Authenticity referred to whether researchers provide an account that is genuine to their field experience such that readers are convinced that the researchers have “been there.” Plausibility had to do with whether an explanation “makes sense” and whether it can be intersubjectively accepted as a likely one. (Lukka & Modell, 2010, 469).

According to Eriksson and Kovalainen (2008, 295), the following concepts also often replace the concepts of validity and reliability in qualitative research: *coherence*, *consistency*⁵, and *usefulness*. According to Parker (2012, 59), different alternative labels are more appropriate for management accounting studies contributing to the understanding and critiquing of management and accounting processes. Based on the above literature review, the theoretical expectation of this study is to report alternative criteria in interpretive and critical studies of management accounting.

2.3 Abandoning common evaluation criteria

In the third approach, authors do not report the use of any general evaluation criteria (see the third row of Table 1). Instead, they may, or may not, have validated and assessed their research in some other way, such as providing sufficient details on the conduct of the study to allow others to evaluate the validity and reliability of the findings. This is important because the methodological exposition in a qualitative paper is the very source of its overall authenticity and credibility (Parker, 2014, 24).

⁵ Consistency typically means that another person should be able to examine the work and come to similar conclusions (Douglas, 1971; Eriksson & Kovalainen, 2008, 292; Ryan et al., 2002, 155).

According to Eriksson and Kovalainen (2008, 295), it is assumed in this approach that each study should be assessed from a position of its own instead of using any universal criteria for evaluation. Hence, researchers are encouraged to employ plurality and indeterminacy in evaluating their studies. According to Eriksson and Kovalainen (2008, 295), this refers to the philosophical position where researchers are part of the knowledge production and reproduction; therefore, no universal and stable criteria for assessing the goodness or badness of the research project should be possible. According to them, this is the case especially within post-structural and cultural studies, where the importance of locality, textual reading of materials, discourse analysis, and feminist analysis methods are considered strategic devices and resources for understanding.

In a strongly subjectivist position management accounting scholars have stated that validity criteria can be seen as only temporarily and socially negotiated constructs that are potentially in flux (Lukka & Model, 2010, 474). Validation is seen as “a process not easily separated from the ongoing efforts of researchers to develop explanations as research projects unfold,” the purpose of which is to convince readers about the plausibility of explanations and the authenticity of findings (see Lukka & Modell, 2010, 462). Rather than following pre-specified stable criteria to minimize various biases, “in a very broad sense, validation refers to the ways through which the credibility of a piece of research is developed and legitimized in front of relevant audiences” (Lather, 1993; Lincoln & Guba, 2000; Polkinghorne, 2007; Silverman, 2000 in Lukka & Modell, 2010, 463). Based on the above discussion, the theoretical expectation of this study is that common evaluation criteria can be abandoned in interpretive and critical studies of management accounting.

In conclusion, the above literature review can be summarized as follows:

- Classic evaluation criteria are likely to be reported when the focus of the research is on objective indicators rather than the subjective meanings of data. According to Hopper and Powell’s (1985) taxonomy, this focus on objectivism and functionalism refers primarily to the mainstream accounting research.
- When the research relies on relativist ontology and subjectivist epistemology, that is, it is interpretive accounting research according to Hopper and Powell’s (1985) taxonomy, alternative evaluation criteria developed for qualitative research will be reported or common evaluation criteria will be abandoned.

- When the research has critical or radical perspectives and assumes that empirical reality is characterized by objective, real relations, which are transformed and reproduced through subjective interpretation, alternative evaluation criteria developed for qualitative research will be reported or common evaluation criteria will be abandoned.

The following research questions will be empirically explored:

- Are all the three approaches and all the different criteria (Table 1) to evaluation reported?
- Is there a high emphasis on classic evaluation criteria in mainstream research?
- Is there a high emphasis on alternative evaluation criteria or the abandonment of common evaluation criteria in interpretive and critical research?

With this pre-understanding about the different approaches to validating and evaluating qualitative research (see also Table 1), we analyse qualitative management accounting studies over the past over nine years (2006 - February 2015) to acquire new knowledge about the state of the art of validations and evaluations in qualitative management accounting research.

3. Research methods and the dataset

Ascertaining how scholars currently report the use of evaluation criteria necessitates content analysis of management accounting studies. Our analysis involves articles gathered from 11 journals in which management accounting case and field studies had been previously published (see Merchant & Van der Stede, 2006), namely *Accounting, Auditing & Accountability Journal* (AAAJ), *Accounting, Organizations and Society* (AOS), *Behavioral Research in Accounting* (BRIA), *Contemporary Accounting Research* (CAR), *European Accounting Review* (EAR), *Journal of Accounting and Economics* (JAE), *Journal of Accounting Research* (JAR), *Journal of Management Accounting Research* (JMAR), *Management Accounting Research* (MAR), *Qualitative Research in Accounting and Management* (QRAM), and *The Accounting Review* (TAR). The selection of these journals aligns well with previous management accounting literature reviews.⁶ Although these journals do not comprise a complete list of double-blind

⁶ Merchant and Van der Stede (2006) analyzed all of these except QRAM. Note that although Zawawi and Hoque (2010) analyzed 22 leading accounting journals, others have mostly analyzed six to ten journals and these were mostly the same journals as were used for this study. Shields (1997) also included AOS, TAR, CAR, JAE, JAR, and JMAR in his review of management accounting research. Chenhall and Smith (2011) also focused on TAR, JAR, AOS, CAR, JAE, AAAJ, MAR, JMAR, and BRIA. In addition, they included

reviewed international accounting journals, they are a manageable selection of journals that provide an overview of management accounting research.

We reviewed journals from 2006 through February 2015 in order to be able to offer an analysis of the recent frontiers of knowledge. As our intent was to accumulate a body of current management accounting case and field studies that was as complete as possible, we reviewed the journals in the relevant ScienceDirect (Elsevier), Business Source Elite (EBSCOhost), and Emerald databases. We first identified and counted the total number of articles published. We included research articles and excluded announcements and notes on matters (such as editorial board changes, brief editorial notes, forewords, introductory notes, calls for papers, and obituaries). Thereafter, we reviewed each research article in the journals and classified it as either qualitative or not-qualitative management accounting research as follows.

First, management accounting studies that had used data collected by interviews, from observations, and/or from documents were included in our study. They had typically been described as “a case study” or “a field study.” However, case studies using multiple (mixed) methods and emphasizing quantitative methods were excluded because we expected their methods and evaluation criteria to be reported in somewhat different ways. Purely methodological case studies and a study with case illustrations were also excluded.

Second, in identifying management accounting articles, we included all the research articles published in JMAR and MAR as both specialize in management accounting research. In selecting articles from the other nine journals, we focused on management accounting research themes such as cost accounting, management control, and others (including accounting information systems, benchmarking, total quality management, just-in-time, strategic management, and transfer pricing) (see also Hesford et al., 2007; Shields, 1997).

Our research method involved both authors of this study first selecting and analyzing the articles independently and, thereafter, meeting to draw joint conclusions about which articles

Accounting and Finance (A & F). Hesford et al. (2007) also analyzed AOS, BRIA, CAR, JAE, JAR, JMAR, MAR, and TAR. In addition to these they included Journal of Accounting Literature (JAL) and Review of Accounting Studies (RAS). Merchant and Van der Stede's (2006) literature analysis on field-based research in accounting also focused on AAAJ, AOS, BRIA, CAR, EAR, JAE, JAR, TAR, JMAR, and MAR. In addition, they analyzed RAS, Auditing: A Journal of Practice & Theory (AudJPT), International Tax Journal (ITJ), and the Journal of the American Taxation Association (JATA).

to include and how to classify them. Agreement was reached on all articles as to whether they were: 1) research articles; and 2) qualitative management accounting research. As a result of this process, 212 articles were selected for the analysis in this study, which is 6.5% of all the 3238 research articles published in these journals during 2006 - 2/2015. Table 2 presents the frequencies of the qualitative management accounting studies during this period by journal and the percentage they represent of the total number of research articles published in each journal. In essence, Table 2 suggests a rather narrow range of publication outlets for qualitative management accounting studies and a low share of the total output.

Place Table 2 here.

Once an article was deemed to be qualitative management accounting research related, the paper was printed, carefully read, and analyzed. The following information was first collected and analyzed for each qualitative paper: title, authors, year of publication, and journal information. We then examined whether there is an accurate description of the field research process. The following key methodological elements were analyzed (Suddaby, 2006): 1) the research strategy (i.e., whether there is an epistemological link between the research question and the case or field method selected); 2) the identification and selection of interview subjects (i.e. whether the selection of case organizations and interviewees is presented); 3) the data analysis methods; 4) the sources of data (such as interviews, documents and/or observations, and a description of the case company); and 5) the use of technical language (evaluation criteria).

When an article was found to report evaluation criteria, we analyzed the reporting of evaluation criteria and the approach to evaluation (i.e., classic, alternative, or both) based on the previous literature and Table 1. All those concepts that included the terms validity, to validate, validation, generalization, or reliability were included in the category “adopting classic criteria of good quality research.” The only exceptions were the terms theoretical generalizability (Lincoln & Guba; Ryan et al., 1992), contextual validity, procedural reliability (Ryan et al., 1992), contextual/ constructive generalization (Lukka & Kasanen (1995), and analytical generalizability (Yin, 2009), as they have been defined as alternative criteria in previous literature. Other evaluation criteria were considered to be alternative criteria.

The paradigmatic affiliations were examined based on the works of Chua (1986), Hopper and Powell (1985), Ryan et al. (2002), and Pihlanto (2003) on the traditions of accounting research (i.e., mainstream accounting research, interpretive research, and critical accounting research). Few articles reported the research tradition so we mostly had to interpret it by studying the research objectives, theoretical underpinnings, empirical analysis, and the findings and conclusions of the papers. The paper was classified as a mainstream research:

- if the research objective was to describe and/or explain the functioning of accounting (Hopper & Powell, 1985; Ryan et al., 2002, 41);
- if the research was founded on some economics-based theory (such as neoclassical economics, information economics, and agency theory) or such behavioral, organizational, and social theories that extend the discipline base of the subject but do not represent a major methodological shift (e.g., contingency theory, systems theory, and organizational and behavioral decision theory) (Ryan et al., 2002, 68-93);
- if the empirical analysis was, or seemed to be, based on the fundamental assumptions that theory and observation are independent of each other, that empirical reality is objective and external to the subject, and that human actors are passive and rationally pursue their assumed goals (Chua, 1986, 611); and/or
- if the findings and the conclusions were facts or (causal) relationships between variables rather than interpretations.

The paper was classified as interpretive research:

- if the research objective was to describe, understand, and interpret the human intentions, social nature of, and changes in accounting practices (Ryan et al., 2002, 42)
- if the research was influenced, for example, by behavioral, organizational, and institutional perspectives (Ryan et al., 2002, 68-93)
- if the empirical analysis was or seemed to be based on fundamental assumptions that social reality is emergent, subjectively created, and objectified through human interaction (Chua, 1986, 615) and/or
- if the findings and the conclusions were interpretations of the actors' everyday world (Chua 1986, 615).

The paper was classified as critical research:

- if the research objective was to provide a social critique and promote radical change in accounting practices (Ryan et al., 2002, 42);
- if the research was influenced by social theories such as the labor process perspective, Giddens' structuration theory, and the theories of Foucault, Harbermas and Latour. (Ryan et al., 2002, 68-93);
- if the empirical analysis was, or seemed to be, based on assumptions about empirical reality characterized by objective, real relations that are transformed and reproduced through subjective interpretation and that fundamental conflict arises because subjective interpretation is obscured by injustice and ideology in the social, economic, and political domains (Chua, 1986, 622); and/or
- if the findings and the conclusions were about understanding the social and economic world with a critical imperative (Chua, 1986, 622).

If the paper was not clearly mainstream, interpretive or critical but had features of both mainstream and interpretive research, it was classified as “mainstream, interpretive”, and if the paper had features of being both interpretive and slightly critical, it was classified as “interpretive, critical.” There were no papers that had features incorporating both mainstream and critical approaches.

The classification of Ryan et al (2002) was used as a starting point to aid in the analysis of the theoretical underpinnings, but it was noted that theories can also be used in different ways depending on the research objectives. The final phase of data collection resulted in the production of a dataset listing the above information. The dataset was analyzed resulting in Tables 3-7 and Appendices 1-2.⁷

⁷ We also studied the chronological order of the data to establish whether there have been trends in the reporting of various evaluation criteria but to a large extent did not find clear tendencies. Using binominal logistic regression analysis, we also tried to control for the influence of a number of variables on the selection of the evaluation approach. These variables included the frequency of the criteria reported, the age of the article, the origin of the article (Europe/outside Europe), the use of the theory (yes/no), the extent to which key elements of the methodology had been reported, and the research tradition (mainstream/other, interpretive/other, critical/other). These results are not reported as there were no statistically significant associations. The only exception was that the frequency of evaluation

Throughout the research process, we attempted to enhance the validity and reliability of the study in many ways, such as by selecting and using appropriate theoretical and methodological underpinnings, by cross-checking the data and analysis, by providing descriptions of the research process that are as accurate as possible, and by linking our findings to previous research.

4. Empirical analysis

4.1 Evaluation and validation approaches

To investigate the ways in which the use of evaluation criteria is being reported in qualitative studies in management accounting, our analysis first focused on the existence of different approaches in the reporting of evaluation criteria. Based on Eriksson and Kovalainen's (2008) classification system (first column of Table 1), we expected to find three different approaches. As Table 3 shows, we found the following four approaches in the 212 studies analyzed: the adoption of classic criteria (ca. 28%), the use of alternative criteria (ca. 9%), the use of both classic and alternative criteria (ca. 14%), and abandonment of general evaluation criteria (ca. 48%). As just over one-half of the studies (51.9%) had adopted one or more evaluation criteria, and just under one-half of the studies (48.1%) had not adopted evaluation criteria, the two main approaches to evaluation were equally likely to be chosen (Chi-Square Goodness-of-Fit Test $\chi^2 = 0.302$, n.s.). These results provide empirical support for Eriksson and Kovalainen's (2008) three approaches to evaluate and validate qualitative research but they also identify a fourth approach at the level of single studies that simultaneously combine the use of classic and alternative concepts.

Place Table 3 here.

To investigate whether the choice of paradigmatic affiliation has led to a certain evaluation approach, we cross-tabulated the evaluation approaches by traditions of research. As Table 4

criteria was positively and statistically significantly associated with the reporting of the classic approach (B=0.175, p=0.038), the alternative approach (B=0.197, p=0.096), and the blended approach (B=1.017, p=0.000).

shows, 126 (nearly 60%) of the studies were primarily interpretive, 37 (nearly one-fifth) were mainstream, and only 9 (4%) of the studies were critical. The remaining studies combined different traditions of research in that they had elements of both mainstream and interpretive traditions or of both interpretive and critical traditions.

In line with the theoretical expectations, over one-half of the interpretive studies (see the first row of Table 4) abandoned common evaluation criteria and 11% reported alternative criteria, but one fourth reported classic criteria and ca. 9% blended classic and alternative criteria. In line with the theoretical expectations, the mainstream studies (see the second row of Table 4) have most often reported classic criteria (in ca. 38% of the cases, see the second row of Table 4), as have the studies combining elements of mainstream and interpretive research (in 39.3% of the cases, see the third row of Table 4). However, contrary to the theoretical expectations, most (ca. 62%) of the mainstream studies have either blended classic and alternative criteria, abandoned common evaluation criteria, or reported alternative criteria. Contrary to the theoretical expectations, many of the studies combining mainstream and interpretive elements have also abandoned common evaluation criteria or blended classic and alternative criteria. Critical studies and the studies combining interpretive and critical elements have mostly abandoned common evaluation criteria, which is in line with the theoretical expectations. A few have adopted classic or alternative criteria (see the fourth and fifth rows of Table 4). In conclusion, as expected, critical studies have mostly abandoned common evaluation criteria but a significant proportion of the studies falling into the interpretive and mainstream categories seem to have reported validity criteria that are, strictly speaking, inconsistent with the paradigmatic affiliation of the studies concerned and these have also blended and/or abandoned criteria to a large extent.

Overall, the Chi-Square statistics suggest that whether researchers chose the classic approach to evaluation or one of the other approaches was not dependent on whether they used the mainstream approach or one of the other approaches (Chi-Square $\chi^2 = 2,008$, n.s.). However, whether researchers chose to report alternative criteria or to abandon criteria rather than use the other approaches to evaluation was contingent on the paradigmatic affiliation (interpretive or critical versus other approaches) in a statistically significant way (Chi-Square Contingency Table $\chi^2 = 24,45$, $p < 0.005$). In conclusion, the choice of the validation and evaluation depends in part on whether an interpretive or critical tradition of research is selected. This result is in line with the theoretical expectations of previous research.

Place Table 4 here.

Table 5 presents an additional analysis by tabulating evaluation and validation approaches by journal. The findings suggest quite different patterns in the reporting of criteria across journals in that the majority of management accounting articles published in QRAM (61.8%), MAR (55.8%), and AOS (54.1%) adopted common evaluation criteria, as did half of the articles published in JMAR, but most of the studies published in the AAAJ (63.9%), CAR (60%), and EAR (52.6%) abandoned common evaluation criteria. Of the articles that reported common evaluation criteria, the articles published in AAAJ, MAR, and QRAM most often adopted classic criteria. The articles published in AOS reported classic and alternative criteria to the same degree. MAR and QRAM articles had a higher tendency to report alternative criteria. The studies published in EAR most often blended classic and alternative criteria.

Place Table 5 about here.

4.2 Adopting the classic criteria of validity and reliability

Based on our literature analysis (summarized in the second column of Table 1), Appendices 1 and 2 show that a total of 276 evaluation criteria were mentioned in the 110 studies that reported evaluation criteria (when each evaluation criterion was counted only once per study). On average, this is about 2.5 criteria per study. Over two-thirds of the references were made to classic evaluation criteria and of those the classic term validity was applied most often. Although the quantities are quite small, the findings indicate a somewhat increasing trend in the reporting of classic criteria until 2012. In 2013 there was a great decline in using classic criteria although the number of published qualitative management accounting articles increased by over 20% in comparison to 2012 (from 25 articles to 32 articles). There was a great rise in the reporting of alternative evaluation criteria in 2013, but in 2014 there was a steep decrease in using all evaluation criteria. This can be explained by the fact that the number of published qualitative management accounting articles was the lowest during the whole research period.

As expected (see Eriksson & Kovalainen, 2008), the classic concepts of validity, generalizability, and reliability have also been reported with rather differently defined

meanings in the qualitative studies than in statistical research. In this study we found researchers to have reported a very wide range of expressions (see Appendix 1).

As Table 4 shows, all but the primarily critical studies applied the classic criteria of validity and reliability. That is, a large proportion of studies combining elements of mainstream and interpretive research (39.3%), mainstream studies (37.8%), interpretive studies (25.4%), and studies combining elements of interpretive and critical research (25%) utilized classic criteria.

4.3 Adopting and blending the alternative evaluation criteria

As Tables 3 and 4 show, 20 articles reported alternative evaluation criteria only. They were mostly interpretive studies. The alternative evaluation criteria were used a total of 85 times in the 110 studies analyzed (with a mean of 0.77).⁸ The articles that reported alternative criteria reported the following terms most often: theoretical generalizability, credibility, plausibility, trustworthiness, and analytical generalizability (in this order, see Appendix 1). When researchers report the use of alternative criteria, they often describe it in great detail. The detailed reporting of alternative evaluation criteria may signal that the alternative criteria have not yet reached the same kind of status in management accounting research as classic criteria. It may also explain why as many as 30 articles blend classic and alternative criteria when evaluating their own research. In addition to reporting alternative criteria, nine of the 30 articles discussed validity, reliability or generalizability of the study by using a classic criterion such as “validity/validate”, “not generalizable” or “reliability”. The remaining 21 articles used either more specified classic criterion (e.g. internal validity, context validity) or many different classic criteria.

The following alternative validation criteria identified in our classification (Table 1) were not explicitly reported: doctrinal relevance, methodological and method validity, internal logic, and empirical applicability (Näsi, 1979); usefulness (Arbnor & Bjerke, 1977; Eriksson & Kovalainen, 2008; Golden-Biddle & Locke, 1993; Lillis, 2006; Mäkinen, 1980); authenticity

⁸ Please note that the numbers are quite small here and do not indicate clear trends in the use of alternative evaluation criteria.

and criticality (Golden-Biddle & Locke, 1993); and constructive generalizability (Lukka & Kasanen, 1995).

4.4 Abandoning the idea of common evaluation criteria

Nearly half (48.1%) of the articles analyzed did not report the use of any common evaluation criteria.⁹ Perhaps surprisingly, this appears to be quite common across both interpretive and mainstream studies. As Table 4 shows, most of the critical studies (88.9%), studies combining elements of interpretive and critical research (58.3%), and interpretive studies (54.8%) abandoned the reporting of common evaluation criteria.

Nevertheless, an additional analysis shows that the vast majority of the studies provided accurate descriptions of their research strategy, selection of interview subjects and case companies, and data analysis methods and sources of data, thereby contributing to the internal validity of the study, which helps readers to evaluate the validity of the conclusions (see Ryan et al., 2002, 141). These findings not only apply to all the articles (see Table 6) and to all those articles using common evaluation criteria (see the first column of Table 7) but also to the studies abandoning evaluation criteria (see the second column of Table 7), although it is noteworthy that the share of articles is slightly smaller there in all the areas investigated. These findings suggest that the articles abandoning the reporting of common evaluation criteria also have a slightly lower tendency to report about certain other key aspects of their research methodology.

Place Tables 6 and 7 here.

5. Discussion and conclusions

The key findings

This study contributes to the limited body of knowledge on the reporting of management accounting field research. Following Lukka and Modell (2010), it is one of the first to empirically analyze validation and evaluation in management accounting field research. It investigated the ways qualitative management accounting researchers have reported the use of evaluation criteria in research, which is a topic of critical debate in qualitative social science

⁹ Neither did these studies report reasons for not reporting evaluation criteria.

research (see Lukka & Modell, 2010), and whether they are associated with certain paradigmatic affiliations.

Theoretically, this paper contributes to the literature on the validation and evaluation of management accounting field research based on the framework of Eriksson and Kovalainen (2008). Our data, which were collected from 212 qualitative management accounting articles published in some of the top accounting journals, provide partial support for the theoretical expectations as follows. First, in line with Eriksson and Kovalainen's (2008) classification system, the findings reported here also support the following three approaches to validation and evaluation: the adoption of classic concepts, the use of alternative concepts, and the abandonment of common evaluation criteria. In addition, the empirical findings reveal a new fourth approach to validation and evaluation; that is, simultaneously blending classic and alternative evaluation criteria. As about one-half of the studies had or had not adopted evaluation criteria, it seems that these two main approaches to evaluation were equally likely to be chosen.

Second, the findings of this study show partial support for the paradigmatic consistency in the reporting of evaluation criteria in relation to previous literature. Overall, whether researchers chose to report alternative criteria or to abandon criteria rather than use the classic or blended approaches to evaluation was contingent on the paradigmatic affiliation in a statistically significant way. The choice of such validation and evaluation approaches depended in part on whether an interpretive or critical tradition of research was selected. This latter result is in line with the theoretical expectations of previous research (Ahrens & Chapman, 2006; Baxter & Chua, 2008; Eriksson & Kovalainen, 2008; Lukka & Modell, 2010; Parker, 2012; Ryan et al., 2002). However, whether researchers chose the classic approach to evaluation rather than one of the other approaches was not dependent on the paradigmatic affiliation (cf. Eriksson & Kovalainen, 2008). Perhaps surprisingly, strictly speaking inconsistent (cf. Eriksson & Kovalainen, 2008) and blended reporting of criteria were found in several studies falling within both the interpretive and mainstream categories. Furthermore, the abandonment of validity criteria seemed to be surprisingly common across both the interpretive and mainstream research. However, we found strong empirical evidence about (internal) validation and evaluation as an ongoing process with accurate descriptions of the key methodological phases and of the use of less established alternative criteria.

Third, the empirical findings support the conclusion that most of the evaluation criteria in our theoretical frame of reference (Table 1) have been used in the management accounting studies examined. This concerns not only the classic evaluation criteria or the alternative evaluation criteria developed by (management) accounting researchers (Lukka & Kasanen, 1995; Ryan et al., 1992) but also the alternative evaluation criteria developed by social scientists (Golden-Biddle & Locke, 1993; Lincoln & Guba, 1985; Yin, 2009). In line with Eriksson and Kovalainen (2008), we also found that the classic concepts of validity, generalizability, and reliability have been used with rather differently defined meanings in the qualitative studies of management accounting than in statistical research. More detailed reporting was found in the context of some of the alternative evaluation criteria, perhaps also suggesting that alternative criteria have not yet reached as high a status as classic criteria and that their use has not yet become as established as the use of classic criteria in the field of management accounting research. The results suggest a somewhat increasing trend in using classic criteria until 2012 and in using alternative criteria during 2013; they also suggest a much higher emphasis on classic criteria than alternative criteria. The articles published in the AAAJ, CAR, and EAR had a slightly higher tendency to abandon common evaluation criteria whereas the articles published in QRAM, MAR, and AOS had the highest tendency to adopt common evaluation criteria.

Limitations, implications and future research topics

This study is subject to the following limitations. First, the review in this paper is not necessarily exhaustive or representative of all the evaluation criteria developed and used. There may also be other criteria of which we have not heard. Second, when analyzing the validation and evaluation of qualitative studies of management accounting, we focused on the reporting of explicit criteria only. As validation and evaluation are complex cognitive processes, it is quite likely that scholars consider and use such insights and criteria during the research process that do not get reported. This may in part be due to the space limitations of their articles. This limits the possibilities for them to conduct an analysis such as ours and results in a partial picture of reality. Third, the findings cannot be generalized. The findings further our understanding of only the convenience sample investigated. Somewhat different results might have been obtained if other journals or fields of research were considered. Fourth, although we initially intended to analyze a longer period of time, we focused on analyzing more recent

material in order to offer an analysis of the recent frontiers of knowledge. An analysis of a longer period of time might yield different results and developments.

Nevertheless, the findings of this study have the following implications. First, they enhance the knowledge of alternative approaches and criteria used in validation and evaluation. The findings can aid both in the evaluation of management accounting research and in the selection of appropriate evaluation approaches and criteria. Second, the findings of this study imply that research practice does not fully reflect theoretical recommendations on the reporting of evaluation approaches and on the reporting of evaluation criteria across paradigms (Eriksson & Kovalainen, 2008). Whether it should reflect them is beyond the scope of this explorative and descriptive study. However, we suggest that this finding is worthy of reflection by researchers. Do some of the findings on strictly speaking inconsistent and blended reporting of criteria and/or the abandonment of criteria in both interpretive and mainstream studies (cf. Erikson & Kovalainen, 2008) suggest the existence of a “crisis of validity” to the extent understood by Lukka and Modell (2010)? Is the “crisis of validity” a general phenomenon in qualitative management accounting research extending beyond interpretive research? If so, what explains this phenomenon? Is it potentially detrimental to the standing of qualitative management accounting research? What should be done about it in future research? What reasons do authors give for not reporting common validity criteria? Future research could also be directed at a more detailed analysis of how specific evaluation criteria have really been used and reported. A more detailed analysis of the articles abandoning the idea of common evaluation criteria could give a better understanding of how authors validate and evaluate their study. A longer period of time could be examined in future studies to analyze developments.

References

- Abernethy, M.A., Chua, W. F., Lockett, P.F. & Selto, F. H. (1999) Research in managerial accounting: Learning from others' experiences. *Accounting and Finance*, 39(1), 1- 27.
- Ahrens, T. & 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. & Dent, J.F. (1998) Accounting and Organizations: Realizing the Richness of Field Studies. *Journal of Management Accounting Research*, 10, 1-39.
- Arbnor, I. & Bjerke, B. (1977). *Företagsekonomisk metodlära*. Lund: Studentlitteratur.
- Baxter, J. & Chua, W. F. (2008) The field researcher as author-writer. *Qualitative Research in Accounting and Management*, 5, 2, 101-121.
- Brewer, J. D. (2000). *Ethnography*, Buckingham: Open University Press.

- Burrell, G. & Morgan, G. (1979). *Sociological Paradigms and Organizational Analysis*, Heinemann Educational Books Ltd: London.
- Campbell, D. (1988). 'Degrees of freedom' and the case study (1975). In E.S. Overman (Ed.), *Methodology and Epistemology for Social Science - Donald T. Campbell* (pp. 377-388). Chicago, IL: University of Chicago Press.
- Chenhall, R. H. & Smith, D. (2011) A review of Australian management accounting research: 1980-2009. *Accounting & Finance*, 51, 173-206.
- Chua, W.F. (1986) Radical developments in accounting thought. *The Accounting Review*, 61, 4, 601-632.
- Creswell, J. W. (1994). *Research Design. Qualitative & Quantitative Approaches*. Sage Publications, Thousand Oaks.
- Douglas, J. D. (1971). *Understanding Everyday Life*. London: Routledge.
- Eriksson, P. & Kovalainen, A. (2008). *Qualitative Methods in Business Research*. London: Sage.
- Garrat, D. & Hodkingson, P. (1998). Can there be criteria for selecting research criteria? – Hermeneutical analysis of an inescapable dilemma. *Qualitative Inquiry*, 4, 515-539.
- Golden-Biddle, K. & Locke, K. (1993). Appealing works: an investigation of how ethnographic texts convince. *Organization Science*, 4, 595-616.
- Grafton, J., Lillis, A.M. & Mahama, H. (2011). Mixed methods research in accounting, *Qualitative Research in Accounting & Management*, 8, 1, 5 – 21.
- Hair, J.F., Black, W.C., Babin, B.J., Anderson, R.E. and Tatham, R.L. (2006). *Multivariate Data Analysis*, Sixth edition, Pearson Prentice Hall, Upper Saddle River, New Jersey.
- Hesford, J. W., Lee S. H., Van der Stede, W. A., and Young, S. M. (2007) Management Accounting: A Bibliographic Study. In Chapman, Hopwood and Shields (Eds.) *Handbook of Management Accounting Research* Volume 1, Elsevier, Amsterdam, The Netherlands.
- Hopper, T. & Powell, A. (1985) Making sense of research into organizational and social aspects of management accounting: a review of its underlying assumptions. *Journal of Management Studies*, 22, 5, 429-465.
- Hoque, Z. (2008) *Methodological Issues in Accounting Research: Theories and Methods*. Spiramus: London.
- Ihantola, E.M. & Kihn, L.A. (2011). Threats to validity and reliability in mixed methods accounting research, *Qualitative Research in Accounting & Management*, 8, 1, 39 – 58.
- Kakkuri-Knuuttila, M.L., Lukka, K. & Kuorikoski, J. (2008). Straddling between paradigms: a naturalistic philosophical case study on interpretive research in management accounting. *Accounting, Organizations and Society*, 33(2-3), 267-291.
- Lather, P. (1993). Fertile obsession: Validity after poststructuralism. *Sociological Quarterly*, 34, 673-693.
- Lillis, A. (2006) Reliability and validity in field study research. In Hoque, Z. (Ed.) *Methodological Issues in Accounting Research: Theories and Methods* (pp. 461-475). London: Spiramus.
- Lillis, A., & Mundy, J. (2005). Cross-Sectional Field Studies in Management Accounting Research – Closing the Gaps between Surveys and Case Studies, *Journal of Management Accounting Research*, 17, pp. 119-141.
- Lincoln, Y. S. & Guba, E. G. (1985). *Naturalistic Inquiry*. Beverly Hills, CA: Sage.
- Lincoln, Y. S. & Guba, E. G. (2000). Paradigmatic controversies, contradictions, and emerging confluences. In N. K. Denzin & Y.s. Lincoln (Eds.), *Handbook of Qualitative Research* (2nd ed.). Thousand Oaks: Sage.
- Lukka, K. & 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, 462-477.
- McKinnon, J. (1988). Reliability and Validity in Field Research: Some Strategies and Tactics, *Accounting, Auditing & Accountability Journal*, 1, 1, 34 – 54.
- Merchant, K.A. & Van der Stede, W. A. (2006) Field-based research in accounting: accomplishments and prospects. *Behavioral Research in Accounting*, 18, 117-134.
- Modell, S. (2005). Triangulation between Case Study and Survey Methods in Management Accounting Research: An Assessment of Validity Implications. *Management Accounting Research*, 16, 231-254.
- Mäkinen, V. (1980). *Yrityksen toiminnan tutkimisen lähestymistavoista. Toiminta-analyttisen tutkimusstrategian kehittelyä.* Tampereen yliopisto, Yrityksen taloustieteen ja yksityisoikeuden laitoksen julkaisuja A:1-17.
- Näsi, J. (1979). *Yrityksen suunnittelun perusteet. Käsitteellismetodologiset rakenteet ja tieteenfilosofinen tausta.* [English Summary: The Basis of Corporate Planning]. Tampereen Yliopisto, Yrityksen taloustieteen ja yksityisoikeuden laitoksen julkaisuja Series A:1-15.
- Parker, L. D. (2012) Qualitative management accounting research: Assessing deliverables and relevance. *Critical Perspectives on Accounting*, 23, 54-70.
- Parker, L. D. (2014) Qualitative perspectives: through a methodological lens. *Qualitative Research in Accounting and Management*, 11, 1, 13-28.
- Patton, M. (1985). *Qualitative evaluation and research methods.* Thousand Oaks, CA: Sage Publications.
- Pihlanto, P. (2003) The role of individual actor in different accounting research perspectives. The holistic individual image as a tool for analysis. *Scandinavian Journal of Management*, 19, 153-172.
- Polkinghorne, D. E. (2007) Validity issues in narrative research. *Qualitative Inquiry*, 13(4), 471-486.
- Ryan, B., Scapens, R.W. & Theobald, M. (2002). *Research Method and Methodology in Finance and Accounting*, 2nd ed. London: Thomson.
- Schwandt, T. A. (1996). Farewell to criteriology. *Qualitative Inquiry*, 2, 58-72.
- Schwandt, T. A. (2001). *Dictionary of Qualitative Inquiry.* London: Sage.
- Shields, M. D. (1997) Research in Management Accounting by North Americans in the 1990s. *Journal of Management Accounting Research*, 9, 3-61.
- Silverman, D. (2000). *Doing qualitative research: A practical handbook.* London: Sage.
- Smith, J. K. & Hodgkinson, P. (2005). Relativism, criteria and politics. In N. K. Denzin & Y. S. Lincoln (Eds.), *The Sage handbook of qualitative research* (3rd ed.) Thousand Oaks: Sage.
- Suddaby, R. (2006) From the editors: What grounded theory is not. *Academy of Management Journal*, 49(4), 633-642.
- Yin, R.K. (2009) *Case Study Research. Design and Methods*, fourth edition. Sage Publications, Thousand Oaks, 4th ed. 2009.
- Zawawi, N. H. M. & Hoque, Z. (2010) Research in management accounting innovations. An overview of its recent development. *Qualitative Research in Accounting & Management*, 7, 4, 505-568.

Table 1. Approaches to qualitative research evaluations and criteria

| | Approach to evaluation | Examples of evaluation criteria |
|----|--|---|
| 1. | Adopting the classic criteria of validity and reliability and using them in a way suitable to qualitative research | <i>Validity</i> <i>Construct validity</i> <i>Internal validity</i> <i>External validity/Generalizability</i> <i>Reliability</i> |
| 2. | Adopting alternative criteria | <i>Doctrinal relevance</i> <i>Philosophical and methodological validity</i> <i>Internal logic</i> <i>Empirical applicability</i> (Näsi, 1979) <i>Trustworthiness:</i> <i>Dependability</i> <i>Conformability</i> <i>Credibility</i> <i>Transferability, theoretical generalisability</i> (Lincoln & Guba, 1985) <i>Contextual validity</i> <i>Transferability</i> <i>Procedural reliability</i> (Ryan et al., 2002) <i>Convincing:</i> <i>Authenticity</i> <i>Plausibility</i> <i>Criticality</i> (Golden-Biddle & Locke, 1993) <i>(Practical) usefulness</i> (Arbnor & Bjerke, 1977; Golden-Biddle & Locke, 1993; Lillis, 2006; Mäkinen, 1980) <i>Contextual generalizability</i> <i>Constructive generalizability</i> (Lukka & Kasanen, 1995) <i>Analytical generalizability</i> (Yin 2009) <i>Coherence</i> <i>Consistency</i> <i>Usefulness</i> <i>Results extended to a wider context</i> (Eriksson & Kovalainen, 2008) |
| 3. | Abandoning the idea of common evaluation criteria | - |

Table 2. Management accounting field and case studies published

| Journal | <i>Frequency of MA field and case studies</i> | | | | | | | | | | 1-2/ 2015 | 2006- 1-2/ 2015 | <i>Frequency of all articles 2006- 1-2/2015</i> | <i>Percent MA case studies of all articles</i> |
|--------------|---|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|----------------|--------------|-----------------------|---|--|
| | 2006 | 2007 | 2008 | 2009 | 2010 | 2011 | 2012 | 2013 | 2014 | | | | | |
| AAAJ | 1 | 1 | 3 | 4 | 5 | 5 | 6 | 7 | 4 | 0 | 36 | 395 | 9.1 % | |
| AOS | 2 | 3 | 4 | 1 | 6 | 10 | 6 | 4 | 1 | 0 ^a | 37 | 346 | 10.7 % | |
| RIA | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | N/A | 0 | 132 | 0 % | |
| CAR | 0 | 3 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | N/A | 5 | 407 | 1.2 % | |
| EAR | 3 | 5 | 0 | 2 | 2 | 3 | 2 | 1 | 1 | N/A | 19 | 246 | 7.7 % | |
| JAE | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 331 | 0 % | |
| JAR | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | N/A | 0 | 336 | 0 % | |
| JMAR | 0 | 0 | 0 | 2 | 0 | 0 | 2 | 0 | 0 | N/A | 4 | 116 | 3.4 % | |
| MAR | 9 | 8 | 9 | 7 | 8 | 14 | 5 | 13 | 2 | 2 | 77 | 187 | 41.2 % | |
| GRAM | 6 | 1 | 3 | 1 | 3 | 4 | 4 | 7 | 5 | N/A | 34 | 154 | 22.1 % | |
| TAR | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 588 | 0 % | |
| Total | 21 | 21 | 19 | 17 | 25 | 36 | 25 | 32 | 14 | 2 | 212 | 3238 | 6.5 % | |

^a January 2015 AOS was included.

Table 3. Use of approaches to evaluations – Frequency of articles in each year (percentages in italics)

| Approach to evaluation | 2006 | 2007 | 2008 | 2009 | 2010 | 2011 | 2012 | 2013 | 2014 | 1-2/ 2015 | Total |
|---|--------------------|--------------------|-------------------|-------------------|--------------------|--------------------|--------------------|--------------------|-------------------|----------------------|---------------------|
| Adopting the classic criteria of validity and reliability | 6 <i>28.6%</i> | 4 <i>19.0%</i> | 5 <i>26.3%</i> | 5 <i>29.4%</i> | 9 <i>36.0%</i> | 11 <i>30.6%</i> | 8 <i>32.0%</i> | 10 <i>31.3%</i> | 2 <i>14.3%</i> | 0 | 60 <i>28.3%</i> |
| Adopting alternative criteria | 0 | 4 <i>19.0%</i> | 4 <i>21.1%</i> | 1 <i>5.9%</i> | 2 <i>8.0%</i> | 3 <i>8.3%</i> | 0 | 5 <i>15.6%</i> | 1 <i>7.1%</i> | 0 | 20 <i>9.4%</i> |
| Adopting both classic and alternative criteria | 1 <i>4.8%</i> | 3 <i>14.3%</i> | 3 <i>15.8%</i> | 4 <i>23.5%</i> | 4 <i>16.0%</i> | 3 <i>8.3%</i> | 6 <i>24.0%</i> | 4 <i>12.5%</i> | 2 <i>14.3%</i> | 0 | 30 <i>14.2%</i> |
| Abandoning the idea of common evaluation criteria | 14 <i>66.7%</i> | 10 <i>47.6%</i> | 7 <i>36.8%</i> | 7 <i>41.2%</i> | 10 <i>40.0%</i> | 19 <i>52.8%</i> | 11 <i>44.0%</i> | 13 <i>40.6%</i> | 9 <i>64.3%</i> | 2 <i>100%</i> | 102 <i>48.1%</i> |
| Total | 21 <i>100%</i> | 21 <i>100%</i> | 19 <i>100%</i> | 17 <i>100%</i> | 25 <i>100%</i> | 36 <i>100%</i> | 25 <i>100%</i> | 32 <i>100%</i> | 14 <i>100%</i> | 2 <i>100%</i> | 212 <i>100%</i> |

Table 4. Validation and evaluation approaches by paradigmatic affiliation, frequencies and percentages

| Tradition | Approach to evaluation | | | | Total |
|--------------------------|--|----------------------|---------------------------------------|---------------------------------------|-------------|
| | Classic criteria of validity and reliability | Alternative criteria | Both classic and alternative criteria | Abandoning common evaluation criteria | |
| Interpretive | 32 ^a (25.4) ^b | 14 (11.1) | 11 (8.7)) | 69 (54.8) | 126 (100.0) |
| Mainstream | 14 (37.8) | 3 (8.1) | 12 (32.4) | 8 (21.6) | 37 (100.0) |
| Mainstream, Interpretive | 11 (39.3) | 0 (0.0) | 7 (25.0) | 10 (35.7) | 28 (100.0) |
| Interpretive, Critical | 3 (25.0) | 2 (16.7) | 0 (0.0) | 7 (58.3) | 12 (100.0) |
| Critical | 0 (0.0) | 1 (11.1) | 0 (0.0) | 8 (88.9) | 9 (100.0) |
| Total | 60 (28.3) | 20 (9.4) | 30 (14.2) | 102 (48.1) | 212 (100.0) |

^a Number of articles^b Row percentage, thus indicating the coverage of evaluation criteria by paradigmatic affiliation.**Table 5. Validation and evaluation approaches by journal, frequencies and percentages**

| Journal | Approach to evaluation | | | | Total |
|---------|--|----------------------|---------------------------------------|---------------------------------------|-------------|
| | Classic criteria of validity and reliability | Alternative criteria | Both classic and alternative criteria | Abandoning common evaluation criteria | |
| AAAJ | 9 ^a (25.0) ^b | 0 (0.0) | 4 (11.1) | 23 (63.9) | 36 (100.0) |
| AOS | 8 (21.6) | 8 (21.6) | 4 (10.8) | 17 (45.9) | 37 (100.0) |
| CAR | 0 (0.0) | 1 (20.0) | 1 (20.0) | 3 (60.0) | 5 (100.0) |
| EAR | 2 (10.5) | 1 (5.3) | 6 (31.6) | 10 (52.6) | 19 (100.0) |
| JMAR | 1 (25.0) | 0 (0.0) | 1 (25.0) | 2 (50.0) | 4 (100.0) |
| MAR | 29 (37.7) | 5 (6.5) | 9 (11.7) | 34 (44.2) | 77 (100.0) |
| QRAM | 11 (32.4) | 5 (14.7) | 5 (14.7) | 13 (38.2) | 34 (100.0) |
| Total | 60 (28.3) | 20 (9.4) | 30 (14.2) | 102 (48.1) | 212 (100.0) |

^a Number of articles^b Row percentage, thus indicating the coverage of evaluation criteria by journal.

Table 6. Reporting field research methods (N=212)

| Reporting of methodological characteristic | All the articles | | | Total |
|--|--------------------------------------|----------|-----------|-----------|
| | Yes | In part | No | |
| Research strategy | 180 ^a (84.9) ^b | 2 (0.9) | 30 (14.2) | 212 (100) |
| Selection of data | 198 (93.4) | 9 (4.2) | 5 (2.4) | 212 (100) |
| Data analysis method | 176 (83.0) | 7 (3.3) | 29 (13.7) | 212 (100) |
| Description of data | 200 (94.3) | 10 (4.7) | 2 (0.9) | 212 (100) |

^a Number of articles

^b Row percentage, thus indicating the coverage of field research methodology.

Table 7. Reporting field research methods in articles using common evaluation criteria (N=110) and in articles abandoning the idea of common evaluation criteria (N=102), frequencies and percentages

| Methodological characteristic | Articles using common evaluation criteria (N=110) | | | Articles abandoning the idea of common evaluation criteria (N=102) | | | Chi-Square ^c |
|-------------------------------|---|---------|----------|--|---------|-----------|-------------------------|
| | Yes | In part | No | Yes | In part | No | |
| Research strategy | 99 ^a (90.0) ^b | 1 (0.9) | 10 (9.1) | 81 (79.4) | 1 (0.9) | 20 (19.6) | 4.838* |
| Selection of data | 105 (95.4) | 1 (0.9) | 4 (3.6) | 93 (91.2) | 8 (7.8) | 1 (0.9) | 7.681** |
| Data analysis method | 98 (89.1) | 2 (1.8) | 10 (9.1) | 78 (76.5) | 5 (4.9) | 19 (18.6) | 6.058** |
| Description of data | 105 (95.5) | 0 (0.0) | 5 (4.5) | 95 (93.1) | 5 (4.9) | 2 (2.0) | 2.201, n.s. |

^a Number of articles

^b Row percentage, thus indicating the coverage of field research methodology by the approach to evaluation and validation.

^c Asymp. Sig. (2-sided) * p<0.10, ** p<0.05

Appendix 1. Use of evaluation criteria: concepts used in detail. (n=110 articles)

| <i>Use of evaluation criteria language</i> | <i>Frequency</i> |
|---|------------------|
| Classic criteria to validity | |
| <i>validity/ validation of the study/research</i> | 27 |
| <i>validity of the data</i> | 13 |
| <i>internal validity</i> | 12 |
| <i>validity of interpretation of the data/interpretative validity</i> | 9 |
| <i>construct validity</i> | 8 |
| <i>cross-validate</i> | 5 |
| <i>validation (validity)of findings/valid findings</i> | 5 |
| <i>convergent validity</i> | 1 |
| <i>post-meeting validation</i> | 1 |
| <i>validating of the case descriptions</i> | 1 |
| <i>validation of the case insights</i> | 1 |
| <i>validating of the data analysis</i> | 1 |
| <i>validate the theoretical propositions</i> | 1 |
| <i>validity of models</i> | 1 |
| <i>validity of the model and evidence</i> | 1 |
| <i>empirical validation</i> | 1 |
| <i>validity of predictions</i> | 1 |
| <i>invalidate</i> | 1 |
| <i>validate transcripts</i> | 1 |
| <i>can be applied validly</i> | 1 |
| <i>validate responses</i> | 1 |
| Total | 93 |
| Classic criteria to external validity/generalizability | |
| <i>generalizability/generalizable</i> | 13 |
| <i>external validity</i> | 10 |
| <i>generalization</i> | 8 |
| <i>generalize results/observations/findings</i> | 7 |
| <i>no generalizability in a statistical sense</i> | 4 |
| <i>limit the generalization of the results/limited generalizability</i> | 4 |
| <i>have to be cautious /to exercise caution in generalizing</i> | 3 |
| <i>cannot be generalized/not generalizable</i> | 3 |
| <i>generalisation problems/difficult to generalize</i> | 2 |
| <i>to be dangerous to generalise</i> | 2 |
| <i>tentative generalization</i> | 1 |
| <i>external generalizability</i> | 1 |
| Total | 58 |
| Classic criteria to reliability | |
| <i>reliability</i> | 12 |
| <i>reliability of data analysis/reliable data analysis</i> | 6 |
| <i>reliability of data/empirical material</i> | 6 |
| <i>reliability of the results</i> | 5 |
| <i>reliability of the study</i> | 5 |
| <i>reliability of findings/reliable findings</i> | 3 |
| <i>reliability of the evidence</i> | 1 |
| <i>coding reliability</i> | 1 |
| <i>reliability of transcripts</i> | 1 |
| Total | 40 |

| | |
|---|------------|
| Alternative criteria | |
| <i>theoretical generalizability/generalize theory</i> | 14 |
| <i>credibility/credible</i> | 12 |
| <i>plausibility/plausible</i> | 9 |
| <i>trustworthiness</i> | 8 |
| <i>analytical generalizability</i> | 6 |
| <i>convincing</i> | 5 |
| <i>consistency</i> | 5 |
| <i>transferability (of findings)</i> | 5 |
| <i>conformability/confirm facts/confirm findings</i> | 4 |
| <i>contextual generalization/ contextual rhetoric of generalization</i> | 3 |
| <i>generalizable at a theoretical level/theoretical generalization</i> | 3 |
| <i>coherence</i> | 2 |
| <i>contextualize findings/contextualize the interviews</i> | 2 |
| <i>dependability</i> | 2 |
| <i>contextual validity</i> | 1 |
| <i>procedural reliability</i> | 1 |
| <i>scientific generalization</i> | 1 |
| <i>authenticity</i> | 1 |
| <i>transferable to other contexts</i> | 1 |
| Total | 85 |
| Evaluation criteria in total | 276 |

Appendix 2. Use of evaluation criteria – frequencies and rank-order of concepts used each year (n=110 articles)

| <i>Use of evaluation criteria</i> | 2006 | 2007 | 2008 | 2009 | 2010 | 2011 | 2012 | 2013 | 2014 | 1-2/2015 | Total | Rank order |
|-------------------------------------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|----------|------------|------------|
| Classic criteria | | | | | | | | | | | | |
| Validity | 6 | 7 | 9 | 12 | 13 | 9 | 18 | 12 | 7 | 0 | 93 | 1 |
| External validity/ | 6 | 3 | 2 | 6 | 4 | 9 | 17 | 7 | 4 | 0 | 58 | 3 |
| generalizability | 4 | 5 | 2 | 3 | 8 | 4 | 4 | 8 | 2 | 0 | 40 | 4 |
| Reliability | 16 | 15 | 13 | 21 | 25 | 22 | 39 | 27 | 13 | 0 | 191 | |
| <i>Total</i> | | | | | | | | | | | | |
| Alternative criteria | 2 | 16 | 10 | 6 | 12 | 10 | 9 | 17 | 3 | 0 | 85 | 2 |
| Evaluation criteria in total | 18 | 31 | 23 | 27 | 37 | 32 | 48 | 44 | 16 | 0 | 276 | |