Revista de Educação e Pesquisa em Contabilidade
Journal of Education and Research in Accounting

Periódico Trimestral, digital e gratuito publicado pela Academia Brasileira de Ciências Contábeis | Available online at www.repec.org.br

REPeC, Brasília, v. 13, n. 4, art. 6, p. 457-477, Oct./Dec. 2019 | DOI: http://dx.doi.org/10.17524/repec.v13i4.2480 | ISSN 1981-8610

An approach to assess the quality of the research process in Accounting

José Renato Sena Oliveira
http://orcid.org/0000-0001-9820-4682 | *E-mail: jrsenna@uefs.br*

Gilberto de Andrade Martins
http://orcid.org/0000-0001-5144-5286 | *E-mail: martins@usp.br*

Abstract

Objective: To associate research quality attributes with those perceived in the accounting research process.
Method: The Modified Delphi Technique was used. There were two rounds with online questionnaires and personalized access. The panel was composed of 41 professors affiliated to 19 Brazilian graduate programs in Accounting. A guidance matrix with 53 attributes/relationships linked to general quality criteria and nine key-characteristics was developed and assessed by the respondents according to their adherence to research practices.
Results: Approximately ¾ of the items obtained a strong level of agreement. The attributes that obtained low or moderate levels of agreement, however, include items that may compromise the quality and integrity of research such as those related to ethical principles, errors and biases and also related to the impact the research team may have on a study's results.
Contributions: This study is expected to support researchers to self-assess their studies, identifying deficiencies and limitations, which in turn promotes greater acceptance of papers and shortens the process of submission to periodicals. As a consequence, financing opportunities and international partnerships may be created, as well as improve the relevance of scientific studies in the field.
Keywords: accounting – research, scientific production, graduate school, research quality, quality criteria.

Published in Portuguese and English. Original Version in Portuguese.

Received in 11/25/2019. Ask to Revise on 11/29/2019. Resubmitted on 12/4/2019. Accepted on 12/9/2019 by Bruna Camargos Avelino (Assistant Editor) and by Dr. Gerlando Augusto Sampaio Franco de Lima (Editor). Published on 12/20/2019. Organization responsible for the journal: Abracicon.

Copyright © 2019 REPEC. All rights, even translation, are reserved. It is allowed to quote part of articles without prior permission if the source is identified.

1. Introduction

Research is an important activity undertaken in educational institutions and is capable to promote the development and improvement of work techniques and strategies and courses of action in various fields of knowledge. In addition, it is essential to advance knowledge and develop solutions for social problems. Scientific research is an activity intended to test hypotheses, draw conclusions and contribute to generalizable knowledge, expressed in theories, principles, and stated relationships (National Institutes of Health, 1979).

The institutional research environment in Brazil establishes basic productivity standards for graduate programs (Master's and doctoral programs). These standards are consistent with the requirements of government research funding agencies and institutional research infrastructure, being established by the Coordination for the Improvement of High Education Personnel (CAPES). If, however, on the one hand, these standards promote scientific production in various fields, on the other hand, on the eagerness to publish research results, researchers may overlook the relevance of studies and disregard scientific rigor. Therefore, the institutional environment justifies the need for studies addressing the quality of scientific research, that is, verifying how the scientific process has been implemented.

International studies evaluating research in the accounting field address various aspects such as productivity and quality, involving university business departments (Chan, Chang, Tong, & Zhang, 2012; Jones, Brinn, & Pendlebury, 1996; Lowe & Locke, 2005). Other studies address research quality criteria such as the impact of studies (Brown & Gardner, 1985; Carmona, 2006); their relevance (Reiter & Williams, 2002); rigor (Evans, Feng, Hoffman, Moser, & Van der Stede, 2015; Williams, 2014); and validity (Libby, Bloomfield, & Nelson, 2002). Finally, other studies analyze citations (Brown & Gardner, 1985; Dunbar & Weber, 2014) and quality perceptions of accounting periodicals (Ballas & Theoharakis, 2003; Brinn, Jones, & Pendlebury, 2001; Brown & Huefner, 1994; Lowe & Locke, 2005; Lowensohn & Samelson, 2006; Taylor, 2011). Even though these studies do not focus on the quality attributes of good research, they explore productivity and assess the quality of what has been published in scientific periodicals in the field.

Studies addressing accounting research conducted in the Brazilian context have adopted various approaches. Some present a bibliometric perspective and analyze aspects such as methods used in theses/dissertations and publications in periodicals and congresses (Mendonça Riccio, & Sakata, 2009; Miranda, Azevedo, & Martins, 2011), research ethical issues (Antunes, Mendonça, Oyadomari & Okimura, 2011), and analysis of citations in periodicals (Aragão, Oliveira & Lima, 2014). One study, however, analyzes academic productivity (Martins & Lucena, 2014) and specifically, the productivity of professors affiliated to Brazilian graduate programs by investigating the profiles and the main practices of these Programs' scientific production, reporting multiple studies giving account of partial results of a single study, that is, the so-called salami publication, and papers that are rejected by higher impact journals and eventually published when submitted to less prestigious journals.

Bibliometric characterization, however, based on reports and publications, is inadequate to reveal failures in the Accounting research process, which should take into account from the choice of the subject up to the submission of its report for scientific dissemination. Reflecting upon the elements that contribute to the scientific nature of research and its improved quality is key. In this sense, good research practices are defined as rules researchers are supposed to follow to ensure the quality, objectivity, and integrity of results, with reasonable and explicit choices (Denscombe, 2010; Shamoo & Resnik, 2003).

These elements suggest that good research requires appropriate scientific practices to obtain the answer that is the most appropriate to the problem proposed. Given this context, this study's objective was to associate research quality attributes with those perceived in the development of scientific production in Accounting. The scientific research process encompasses key-characteristics or different stages (Mays & Pope, 2006; Spencer, Ritchie, Lewis, & Dillon, 2003) such as planning, implementation, and assessment of results (Brinberg & McGrath, 1985). This study's hypothesis is that certain practices currently adopted in the process of scientific research in accounting in Brazil do not meet the research quality attributes described in the literature.

The findings presented here show that approximately ¾ of the items concerning 53 attributes/quality relationships analyzed obtained a strong level of agreement on the part of researchers in the accounting field. Six attributes, however, obtained moderate agreement and 3 obtained a low level of agreement, including items that may compromise the quality and integrity of research, such as those related to ethical principles, errors and biases, and the impact of the researcher team on studies' results. Additionally, 5 attributes associated with reliability, integrity and internal validity criteria did not obtain a consensus, suggesting that failures may be found in certain stages of the research process that concern these criteria

Assessment of scientific production focused on results is limited to evidencing potential failures in the research process of graduate programs and researchers. Studies assessing research practices may have an impact on the field, as it can support researchers to self assess their studies, identifying deficiencies, limitations, and aspects that need to be taken into account to improve methodological choices. Greater acceptance of papers may result as well as the period to submit papers to periodicals may be shortened, as fewer reviews will be required. Better quality studies that present clear contribution and impact are considered in the allocation of resources and recruiting of professors, among other factors (Chan *et al.*, 2012). An addition result expected is that improved quality scientific research conducted by graduate programs in accounting improves the chance of publishing studies in relevant periodicals in addition to promoting new opportunities of financing, collaboration in international partnerships, and greater relevance of scientific production in the field.

2. Literature Review

The literature review is divided into three subtopics: the first presents an overview of quality and good research practices. The second highlights the stages of the research process based on the Validity Network Schema (Brinberg & McGrath, 1985), as well as on the Research Evidence Assessment Framework (Mays & Pope, 2006; Spencer *et al.,* 2003). Finally, it presents evidence concerning quality criteria and indicators used in evaluations of the research process.

2.1 Quality Research and Good Research Practices

To understand this study's objective, we need to make clear the meaning of its central elements: quality, attribute, and research process. In a generic sense, quality is described as an attribute, natural condition, or property by which something or someone is individualized and distinguished from others. It is also a way of being, essence or nature, or degree of perfection, precision, and conformity with a given standard (Michaelis, 2012). Valentine (2009) notes that, in scientific research, the answer to the question "what are the characteristics of a high-quality study?" in part depends on why such question is being asked, considering that different individuals assign different meanings to it. The aforementioned author defends that quality refers to an appropriate link between the study's objectives, design and purpose involved in its implementation. In terms of research in the accounting field, Clarkson (2012) argues that quality is associated with three fundamental factors: (i) contribution - the importance of the study's focus and its level of innovation; (ii) the rigor with which a study was performed (scientific credibility); and the (iii) ability of a document in reporting the study in a transparent and accessible way (communication).

The concept of attribute also has a variety of meanings, including (1) what is inherent and peculiar to someone or something; (2) a condition, property or quality of something. Grammatically, an attribute modifies a noun expressing its quality or strength (Michaelis, 2012). The research process, in turn, is defined as the identification, combination, and use of elements and relationships of conceptual, methodological and substantive domains, divided into three stages and with different validity perspectives (Brinberg & McGrath, 1985). Additionally, the research process may be described according to the following steps (or stages): findings, design, sample, data collection, analysis, and report (Mays & Pope, 2006; Spencer *et al.*, 2003). Given these definitions, the quality attributes of the research process described in this study include a set of characteristics that show the connection between the focus of interest, evidence, theory and researchers' methodological choices. Quality is present in the various stages of study planning and implementation, permeated by the tradition of the underlying research.

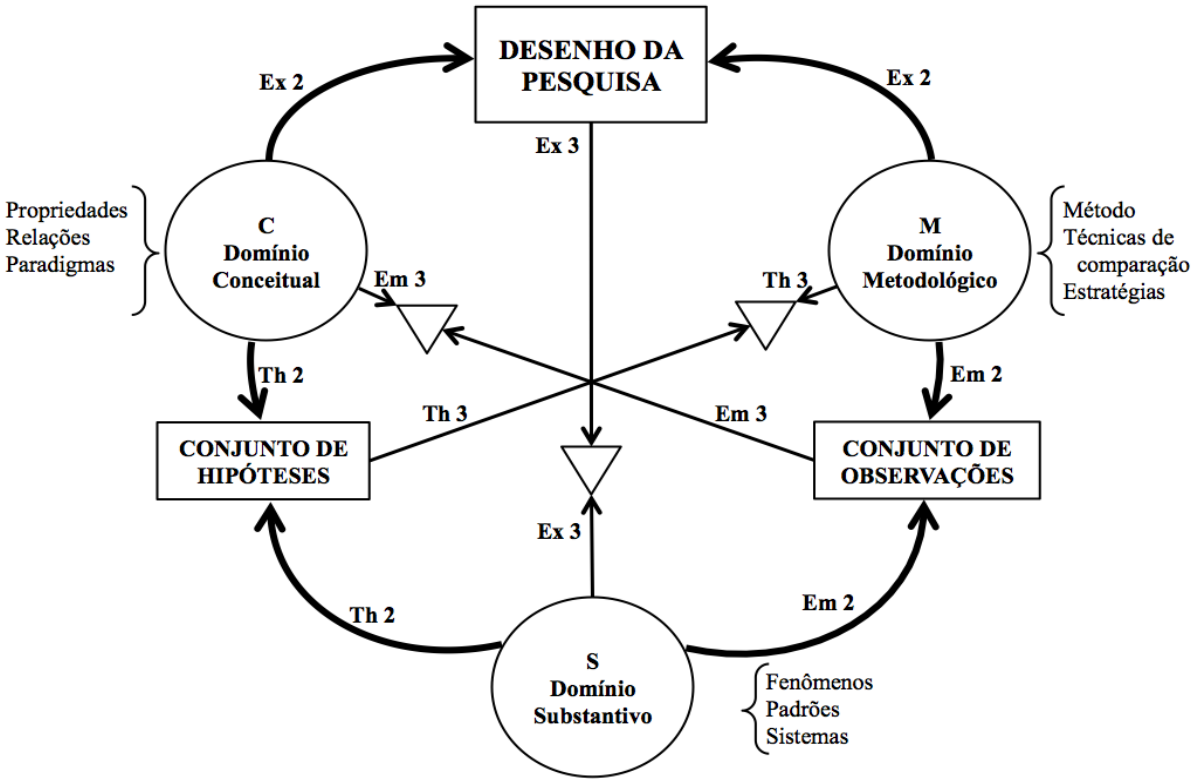
Given its subjective and evaluative nature, it is not an easy, or consensual task, to establish what good research is. Those in leading positions may play a key role in the development of a culture in which ethical behavior and good research practices prevail. If the most prominent researchers, managers, companies, and governmental agencies display and tolerate anti-ethical behavior and poor research practices, such behavior and practices will prevail (Shamoo & Resnik, 2003). Good research practices are rules researchers can follow in order to promote and ensure the quality, objectivity, and integrity of data.

These aspects can guide the assessment of a study quality because they represent the general attributes of good research. Similar elements are reported by one study (Webster & Watson, 2002) that ask questions associated to each of the following: (a) contribution ("what is new?"); (b) impact ("so what?"); (c) logic ("why?"); and (d) rigor ("was it well done?"). The two first aspects refer to the "substance" of a study, that is, its research problem, foundations, gap identified, justification, and potential implications for the field. The logic of a study is not only associated with the development of a theoretical framework. It is also associated with study design while the latter refers to the rigor with which a given study is conducted, as well as to its ethical aspects.

In short, the notion of good research is associated with rules, which if followed, can help researchers to ensure the quality, objectivity, and integrity of data. For that, one needs to make reasonable and explicit choices in regard to every component of a study and in this context, rigor, and quality are strongly related, as one cannot conduct good quality research if the process is poorly implemented. That is why each of the tasks included in the phases (stages) of a research process needs to be carefully considered.

2.2 Research Process Stages: the Validity Network Schema and Research Evidence Assessment Framework

Research is basically a study of relationships, that is, it always takes into account the relationship between units (Brinberg & McGrath, 1985). These authors studied the validity of the research process and argue this is not a commodity that be can be acquired by merely applying techniques, but an ideal state, a concept to be pursued and which is applied in the different stages of a research process. As a result, the authors present the Validity Network Schema (VNS) in which they assume that research involves three inter-related, though analytically distinct, domains: conceptual, methodological, and substantive. While the phenomenon is a central object in the perspective of the substantive domain, the conceptual domain is concerned with the identification of the concepts and attributes that can help to explain the phenomenon under study, that is, a theory (or theoretical lens) that sustains the analysis of information collected. The methodological domain, in turn, is concerned with presenting and describing the means used to conduct a study (Figure 1).



Translation:
STUDY DESIGN - C Conceptual Domain - Properties Relations Paradigms - M Methodological Domain - Method Techniques to compare strategies - S Substantive Domain - Phenomena - Patterns - Systems - SET OF OBSERVATIONS - SET OF HYPOTHESES

Experimental Path (Ex)	→	Building a project and implementing it using a set of substantive events
Theoretical Path (Th)	→	Building a set of hypotheses, testing and assessing them with an appropriate set of methods
Empirical Path (Em)	→	Building a set of observations, explaining and interpreting them based on a set of meaningful concepts

Figure 1. The VNS: domains, levels, and paths.
Adapted from Brinberg and McGrath (1985, p. 22).

Brinberg and McGrath (1985) consider that the entire research process involves three large stages (steps): Stage 1 – Validity as Value, called Pre-Study or Preparatory Stage, concerns research planning, when criteria to assess the relationships of the domains are established, that is, the merit of a research proposal is judged; Stage 2 – Validity as Correspondence, also called Central Stage, refers to the study's implementation phase, in which the use of elements and relationships of the three domains to produce a set of empirical findings using different paths are established; and Stage 3 – Validity as Robustness, called Stage 2 Findings Monitoring, a stage in which the results are assessed and one verifies the extension and outlining of findings in order to identify the study's boundaries and contributions. There are three alternative paths to perform the central stage within a research process: Experimental (Ex), Theoretical (Th) and Empirical (Em), which reflect different styles of doing research and finding different validity questions. In essence, you have a link between object of interest (represented by the substantive domain) and the conceptual relationships denoted by its properties (theory) support the establishment of a set of hypotheses that may explain a given phenomenon. On the other hand, the connection between the focus and the methodological domain is related to strategies used to systematically collect a set of observations that support evidence and findings.

Another perspective to assess research processes is the Research Evidence Assessment framework (Mays & Pope, 2006; Spencer *et al.*, 2003), intended to assess evidence of nine main characteristics and processes:

- (1) Findings – characteristics associated with the assessment of elements such as credibility of findings, how findings broaden existing knowledge, the scope of inference, and whether the original objectives and purpose were met
- (2) Design – justification of why a given design was chosen (methodological choices)
- (3) Sample – assessment of criteria used to conceptualize and selecting a sample, as well as inclusion and exclusion criteria.
- (4) Data Collection – assessment of how data were collected.
- (5) Analysis – depth and complexity of data, the approach that was chosen, and analysis of formulation, data source context, and diversity of perspectives.
- (6) Report – assessment of the connection between data, interpretations, and conclusions, as well as the coherence of the global report.
- (7) Reflectivity and neutrality – clarity regarding the assumptions, theoretical perspectives, and values that guide the study as well as consideration of errors and biases.
- (8) Ethics – assessment of how well researchers and the research team dealt with ethical issues.
- (9) Auditability – verification of formal procedures and documentation of the study process for future inspections.

2.3 Quality criteria and indicators in research process evaluation

According to the definition of the Joint Committee on Standards Committee for Educational Assessment (Yarbrough, Shulha, Hopson, & Caruthers, 2011), a criterion is a standard through which something is judged. Another definition involves the notion of merit, that is, whether something is valuable or not, or whether something is good or not (Davidson, 2005). A third concept describes criteria as a set of standards that establish acceptability (Brinberg & McGrath, 1985). The judgment of each evaluator, however, tends to be different. Evaluators use different indicators in assessment processes, and even though the word "assessment" implies judgment from a broadened perspective, it essentially focuses on merit. In essence, an indicator is described as a gauge (a scale) or measure of a variable (Weiss, 1997). The framework used to assess the quality of evidence produced by qualitative research (Mays & Pope, 2006; Spencer *et al.*, 2003) stresses that it is possible to use "(...) a series of quality indicators that indicate the type of information necessary to asses whether a given quality attribute was met or not".

There is no consensus in the literature on which criteria are best to assess the quality of research. For instance, some authors argue that it is impossible to have only one set of criteria to assess qualitative and quantitative studies due to the differences involved in these two types of research (Whittemore, Chase & Mandle, 2001). Other authors, however, defend the possibility of using common criteria to assess both qualitative and quantitative research (LeCompte & Goetz, 1982).

The assessment of research stages is a task involving various stakeholders. On the one hand, there are funding agencies, interested in the results of the planning phase (study project), relying on criteria such as relevance, impact, and feasibility to decide which projects will be funded. On the other hand, researchers are concerned with quality attributes that allow for a systematic and rigorous study, that is, the implementation of a study that produces the best quality of evidence. Finally, other stakeholders consider the quality of the "research product", that is, scientific publication. The academic community has basically used two ways to identify the quality of periodicals: surveys are conducted with members of the scientific community (referees, editors, researchers, etc.) to identify their perceptions of quality (Ballas & Theoharakis, 2003; Brinn *et al.*, 2001; Lowe & Locke, 2005; Lowensohn & Samelson, 2006; Northcott & Linacre, 2010; Van der Stede, Young, & Chen, 2005) and measures based on studies' citations (citation impact) (Aragão *et al.*, 2014; Doyle & Arthurs, 1995).

Considering the focus of this study is related to the study of practice, a combination of different sources was used to identify a set of general criteria that are possible to apply, as summarized in Table 1. Thus, a set of more general criteria was chosen based on the literature and the frameworks previously mentioned (Brinberg & McGrath, 1985; Mays & Pope, 2006; Spencer *et al.*, 2003), to guide an analysis of the research process in the Accounting field.

3. Method

This study comprises 4 stages. The first stage refers to a bibliographic survey, intended to identify the attributes of good research described in the literature. After establishing a list of general criteria, and considering the VNS domains as well as key-characteristics presented by the Research Evidence Assessment Framework, a matrix was developed to guide the development of the instruments to collect data.

The second stage refers to the application of the Delphi Technique, based on the type of consensus formation. The modified Delphi technique was used, in which the propositions of the first stage are based on the literature rather than proposed by the participants (Kelbaugh, 2003). There were two rounds with professors affiliated to academic graduate programs in accounting that are recognized by CAPES. Overall, 318 professors were invited, 41 participated in the first round and 37 in the second. Nineteen out of 23 graduate programs recognized by CAPES and still active during the period of data collection were represented.

Using a 10-point numeric scale, an instrument was developed to collect data based on a clinical study (Elwyn *et al.*, 2006). It is composed of 53 items listing attributes/relationships concerning the quality of the research process and level of adherence on the part of respondents to such attributes. In order to improve internal validity, a pre-test was performed with 5 evaluators: 3 doctoral students, 1 Ph.D. in Accounting and 1 Ph.D. in Education. Data were collected between October 2015 and February 2016 using the SurveyMonkey online platform, each using custom ID and password that were sent through an email inviting the panelists. The result of the first round presented a Cronbach's alpha equal to 92.2% while consensus was not obtained for 28 items, which were reassessed in the second round. Because only 5 items (less than 10% of the total) did not obtain consensus after the second round and also because standard deviation varied little between rounds, we opted for dispensing the third round. Data analysis considered location and dispersion measures in addition to criteria used to obtain consensus.

Having identified the attributes (Stage 1) and obtained the Delphi's results concerning the consensus (Stage 2), the third stage consisted of comparing evidence with classification according to levels of agreement and identifying potential patterns and/or variations. Criteria to established consensus were: (a) 75% or more assigned 1-3 or 8-10 scores (strong agreement); (b) 25% or less assigned 1-3 scores (disagreement); and (c) standard deviation variation ≤25% between rounds. The classification of items according to the level of agreement is detailed in Table 2.

Table 1
Summary of the general scientific criteria and their characteristics

Criterion	Meaning/Characteristics	Source
Contribution/quality of the theoretical perspective	The extent to which the target-audience can directly use the findings; the extension to which the study advances knowledge or brings broader understanding about a study object. It is characterized by the identification of knowledge gaps and may provide solutions to audience problems.	(Mays & Pope, 2006; Meyrick, 2006; Spencer *et al*., 2003; Webster & Watson, 2002)
External validity or generalization/ opportunity	Application of results in other contexts, generalization to other populations. In a qualitative approach, it is an opportunity to verify how well hypotheses would fit in a context different from the one addressed. Factor: applicability.	(Brinberg & McGrath, 1985; Spencer *et al*., 2003; Valentine, 2009; Yarbrough *et al*., 2011)
Feasibility	It is "the extent to which resources and other factors allow for satisfactory assessment to be performed" (Yarbrough *et al*., 2011). To verify whether a study is feasible, one should consider access to data, resources available, the research team's skills, and time restriction, among others, as well as cost-benefit.	(Davidson, 2005; Rossi, Lipsey, & Freeman, 2004; Weiss, 1997; Yarbrough *et al*., 2011)
Impact	Impact, when assessing a program, can be defined as "a change in the target population or social conditions that have been brought about by the program, that is, a change that would not have occurred in the program had not happened." (Rossi *et al*., 2004). Similarly, impact in the case of scientific research refers to changes that take place due to the results of a study or which were induced by such results.	(Carmona, 2006; Rossi *et al*., 2004; Webster & Watson, 2002; Weiss, 1997)

Criterion	Meaning/Characteristics	Source
Integrity	It refers to the rigor of research questions, design, conduct, and theorization, revealing principles and rules of conduct or codes of practice people and organizations involved in scientific research are supposed to follow: (a) honest communication; (b) reliable research; (c) Objectivity; (d) impartiality and independence; (e) openness and accessibility; (f) duty of diligence; (g) fairness in providing reference and making citations; and (h) being accountable with future scientists and researchers (European Science Foundation, 2011). Failure in meeting the criterion of integrity may be evidence of research misconduct.	(Antunes *et al*., 2011; European Science Foundation, 2011; FAPESP, 2012; OADS, 2012)
Internal validity/ credibility or defensibility	Measures the significance of the independent variable with regard to the dependent variable, or how well and faithfully the phenomenon is represented. In qualitative studies, it is credibility or "how vivid and faithful the description is to the experience lived." (Beck, 1993). In general, validity is defined as "the extent to which it measures what it is intended to measure" (Rossi *et al*., 2004). Factor: Truth value.	(Beck, 1993; Brinberg & McGrath, 1985; Libby *et al*., 2002; Mays & Pope, 2006; Spencer *et al*., 2003)
Relevance	The ability of research to help a group of researchers to solve problems. The relevance of a study depends on its potential use for scientific, educational or applied purposes. Schwartzman (1988) argues that the assessment of the scientific relevance of a study depends on the scientist themselves, notably by peer-review; other forms of relevance required the participation of other stakeholders.	(Mays & Pope, 2006; Reiter & Williams, 2002; Schwartzman, 1988; Weiss, 1997; Yarbrough *et al*., 2011)
Reliability/ auditability	An instrument's measure of consistency in obtaining similar results or the ability of another researcher to follow the audit trail. In other words, the "extent to which a measure produces the same results when repeatedly used to measure the same thing" (Rossi *et al*., 2004). Factor: consistency.	(Beck, 1993; LeCompte & Goetz, 1982; Mays & Pope, 2006; Rossi *et al*., 2004; Yarbrough *et al*., 2011)
Rigor/thoroughness	Complete and reliable recording. Rich, detailed and complex data. Rigor includes good practices to collect and analyze data along with transparence. The perception of rigor, however, is complex and may depend on the research method adopted.	(Denscombe, 2010; Evans *et al*., 2015; Webster & Watson, 2002; Williams, 2014)
Adequacy	The choice of a strategy that is likely to be successful in the achievement of a study's objectives, which is clearly and explicitly justified. According to Denscombe (2010), it is not a matter of whether a definition is good or bad – or correct or wrong -, but rather whether it is useful and appropriate to solve a given research problem.	(Denscombe, 2010)

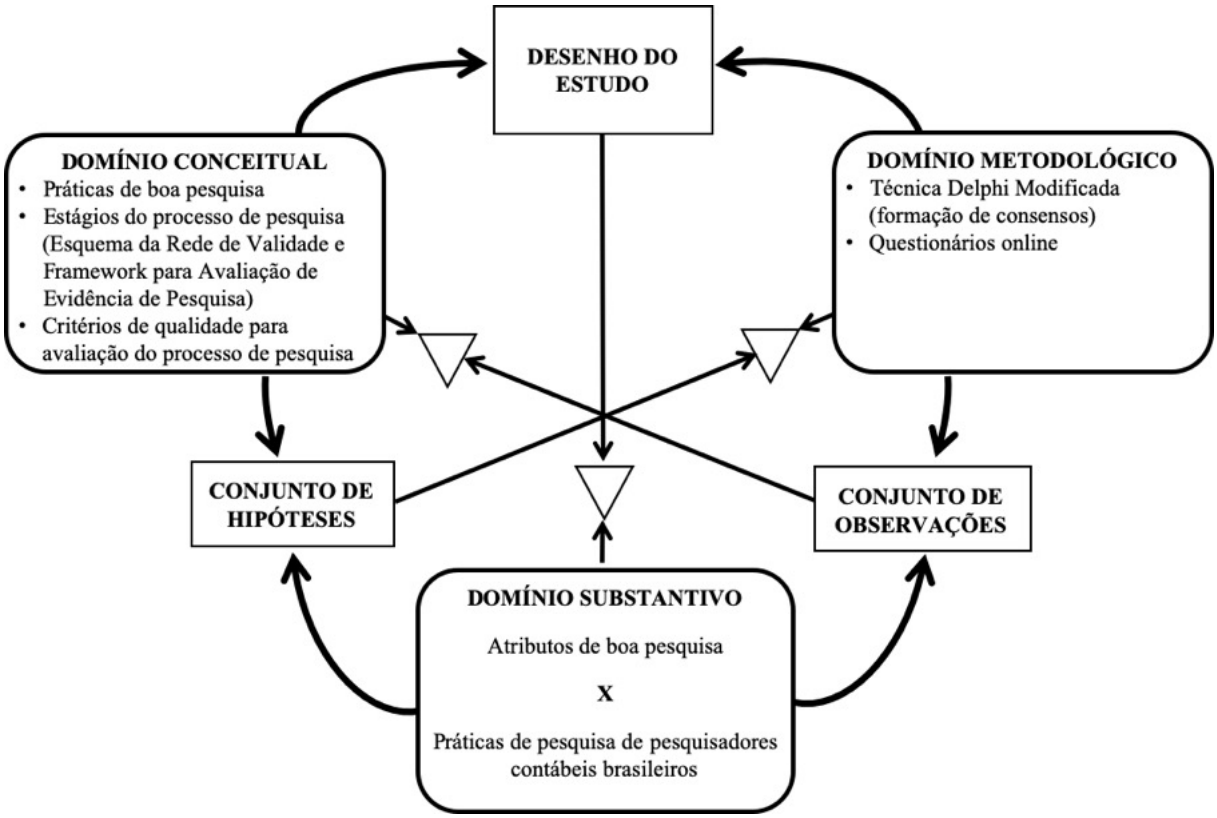
Table 2

Criteria to classify the items according to levels of agreement

Rodríguez-Mañas *et al*. (2013)	Adjusted model	Level of agreement
>80% of the answers ≥8 or ≤3	>75% of the answers ≥8 or ≤3	Strong
70% – 80%	65% – 75%	Moderate
50% – 70%	50% – 65%	Low
<50%	<50%	No consensus was obtained

Based on the propositions and the use of a logic model, the fourth stage refers to the development of an approach to assess the research process. A logic model is a technique used to assess programs and projects, both focusing on the process as well as on the outcomes (Kellog Foundation, 2014; Ladd & Jernigan, 2006). The logic model in this study was structured by adapting examples studied during the Introduction to Evaluation Theory course administered by Professor Thomas A. Schwandt from the University of Illinois at Urbana-Champaign (UIUC), USA in September 2014. The attributes/relationships identified in the literature and assessed by experts in the Delphi's panel were considered. Additionally, the potential implications of not meeting criteria in the VNS's substantive, conceptual and methodological domains were presented.

The study design is summarized according to the scientific research domains presented in the Validity Network Schema (Figure 2).



Translation:
STUDY DESIGN - Conceptual Domain - Good research practice - Stages of the research process (Validity network scheme and Assessment framework for research evidence) - Quality criterion for the assessment of the research process - Methodological Domain - Modified Delphi Technique (reaching consensuses) - Online questionnaires - Substantive Domain - Attributes of good research - Research practices of Brazilian accounting researchers - SET OF OBSERVATIONS - SET OF HYPOTHESES

Figure 2. Summary of the VNS-based research design (Brinberg & McGrath, 1985)

This study was submitted to and approved by the Institutional Review Board (IRB) and is registered in the Brasil Platform.

4. Results and Discussion

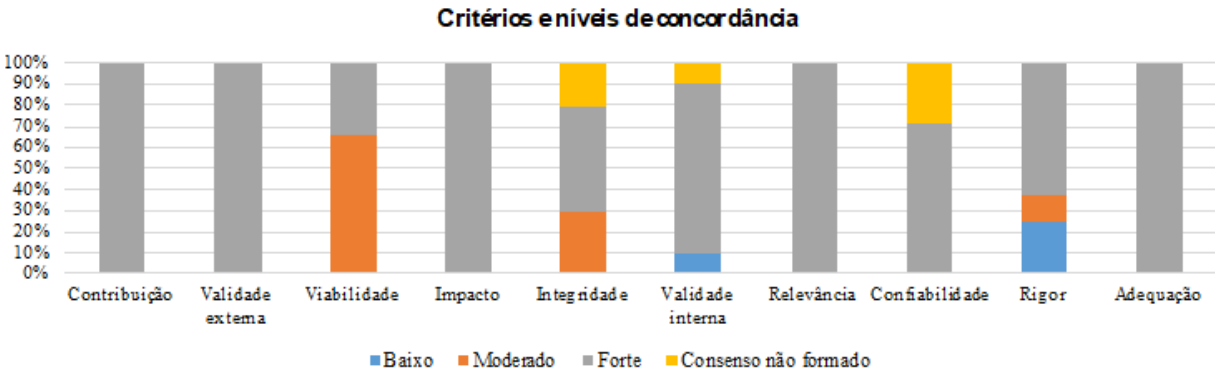
The Delphi's expert panel was composed of 41 individuals in the first found and 37 in the second round. In order to characterize the panel, the respondents were asked about their professional background and experience. Most had a doctoral degree in Accounting obtained in a Brazilian institution different from the one to which they were affiliated and less than 15% had attended a sandwich doctoral program or post-doctorate in an international institution, suggesting poor experience with international activities, which may reflect a low level of insertion in international research networks. Only 9.76% had obtained their doctoral degree in the same graduate program to which they were affiliated, which is evidence of decreased endogeneity, an element CAPES consider to indicate improved quality. Additionally, the graduate programs with diversified educational experiences tend to have researchers with very different skills and worldviews, which may contribute to more comprehensive and innovative studies.

Regarding the respondents' experience with research, an important factor to characterize their participation as a panel expert, 34.14% reported up to two years of experience in a graduate program while the remaining 26.83% had between 5 and 7 years of experience. Additionally, 56.10% had received/receive research financial support, approximately ¼ reported a research productivity scholarship and only 2.44% did not have papers published in periodicals classified A1, A2 or B1 in CAPES' last assessment. Most are affiliated to Master and Doctoral programs rated with a grade 4. These data reveal the group presents research experience that qualifies it for such judgment.

4.1 Implementation of the Modified Delphi Technique

The final result of the modified Delphi is detailed in Appendix A. The number of valid answers varied due to the number of negative answers related to the use of research involving human subjects. Additionally, only five propositions did not reach consensus, not justifying the effort needed to perform another round, given the low variability in standard deviation between the two rounds.

The findings reveal that 73.58% (39) of the 53 items obtained a strong level of agreement; 11.32% (6) obtained a moderate level of agreement; 5.66% (3) a low level; and 9.43% (5) did not reach a consensus. The attribute "subjects' voluntary participation", which is linked to integrity, obtained the highest relative score and 10 items obtained a percentage above 90%. The item that obtained the highest absolute score was "objective/problem was presented with accuracy", which is related to adequacy criterion and obtained 395 points out of the 410 possible.



Translation:
Criteria and levels of agreement
Contribution - External validity - Feasibility - Impact - Integrity - Internal validity - Relevance - Reliability - Rigor - Adequacy
Low - Moderate - Strong - Consensus was not obtained

Figure 3. Criteria and levels of agreement

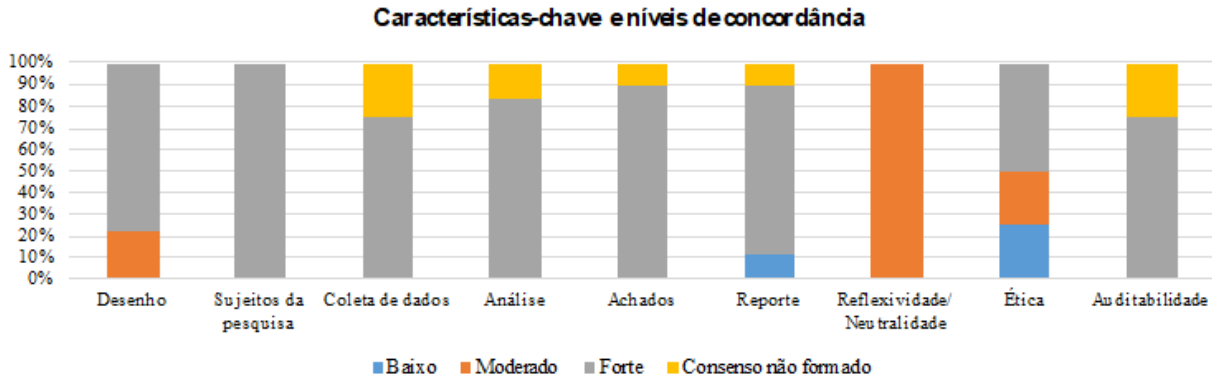
Figure 3 shows that the items concerning contribution/quality of theoretical perspective, external validity or generalization/opportunity, impact, relevance, and adequacy show strong agreement in the respondents' opinion. Approximately 65% of the items related to the feasibility criterion obtained moderate agreement. Even though these do not necessarily indicate severe failures, these should be considered when planning studies, as they are associated with elements such as the size of the study project in terms of time restriction and target population that influence the selection of research strategy. According to VNS (Brinberg & McGrath, 1985), the premise of planning is validity as value, which is called the preparatory stage and should precede the subsequent stages. As a consequence, failure in this stage may compromise a study's feasibility (Rossi *et al*., 2004; Weiss, 1997; Yarbrough *et al*., 2011) and lead to inconsistent data and findings, not meeting internal and external validity as well as a study's impact.

A total of 35% of the integrity items obtained moderate agreement. Integrity requires attention to issues on how researchers deal with errors and biases and how the research team impacts the results. Data integrity is a central element in terms of good research practices and violation of such an element may compromise the scientific credibility of findings (Clarkson, 2012; European Science Foundation, 2011; FAPESP, 2012; OADS, 2012; Shamoo & Resnik, 2003).

Items with a low level of agreement were associated with internal validity/credibility or defensibility and rigor. Approximately ¼ of these obtained a low level of agreement and should be considered by researchers because failure in meeting these criteria may suggest weaknesses during the conduction of studies in the field. Poor agreement in these items may indicate loss of research quality because it may compromise the credibility of findings, as well as reveal aspects that could be improved in the training of researchers in the field, given the importance given to these criteria in the literature (Evans *et al*., 2015; Mays & Pope, 2006; OADS, 2012; Spencer *et al*., 2003; Williams, 2014).

The items that did not obtain consensus are associated with integrity, internal validity/credibility or defensibility, and reliability/auditability. The scores obtained in these items reveal discrepancies or lower average acceptance among respondents in addition to greater dispersion that impeded a consensus to be reached, even when relatively high scores were obtained. These criteria are important for the quality of studies (LeCompte & Goetz, 1982; Libby *et al*., 2002; Mays & Pope, 2006; Spencer *et al*., 2003; Yarbrough *et al*., 2011) and dissonant assessments of these may reveal a lack of clarity regarding the extension of findings in regard to reliability, such as a need to record the reasons changes were implemented in the planning of studies, as well as the reasons for study limitations. Consequently, it may reveal that some research practice requires changes in order to ensure these criteria are met.

Regarding the relationship between key-characteristics and levels of agreement, Figure 4 shows that "research subjects" was the only characteristic with a strong level of agreement. On the other hand, reflectivity/neutrality obtained a moderate level of agreement. Additionally, "ethics" and "report" concentrated the low agreement items. Approximately 25% of the items related to ethics presented a low level of acceptance, suggesting researchers need to be concerned with ethical aspects, possibly indicating weaknesses in terms of methodological rigor and that some essential aspects of research integrity were not met (European Science Foundation, 2011; Mays & Pope, 2006; Spencer *et al*., 2003). Lower levels of agreement suggest potential misalignment between methodological, substantive and theoretical domains proposed by VNS (Brinberg & McGrath, 1985).



Translation:
Key-characteristics and levels of agreement - Design - Study subjects - Data collection - Analysis - Findings - Report - Reflectivity/neutrality - Ethics - Auditability
Low Moderate Strong Consensus was not obtained

Figure 4. Key characteristics and levels of agreement

Among the items with a low level of acceptance, there is the one concerning peer-review, associated with "reporting" and two other items that concern the ethical aspects of formal and mandatory submission of research projects to an institutional review board. Draft review plays an important role in the identification of potential problems and may suggest points that need improvement, also contributing to a manuscript's consistency and logic coherence (Mays & Pope, 2006; Spencer *et al.*, 2003). As a result, peer-review may improve the chances of a manuscript to be approved in periodicals submissions as it decreases potential failures that could lead a manuscript to be rejected (Carmona, 2006; Martins & Lucena, 2014; Shamoo & Resnik, 2003; Valentine, 2009; Webster & Watson, 2002).

Ethical issues are the only characteristics with items that obtained three levels of agreement. Similar to what happens internationally, Brazilian law requires that studies involving human subjects be submitted to the previous assessment and approval of an institutional review board in order to ensure ethical principles, such as respect to participants' dignity and autonomy, are complied with. Informed consent forms should describe the purpose, goals, and procedures of studies, providing clarification regarding potential discomfort and risks, and also specify how participants are monitored and assisted, a condition all participants are entitled to, even after a study is concluded (Brasil, 2012, 2013). Brazilian law also provides that ethical infractions or complaints entailing risks to the participants should be examined and may involve sanctions, including investigation on the part of the Public Prosecution Service. These practices are consistent with international principles such as respect for people, beneficence, and justice (National Institutes of Health, 1979), accountability, respect and integrity (European Science Foundation, 2011; OADS, 2012).

Observance to ethical standards in research involving human subjects has at least three practical effects: (a) ensures that research meets international standards of integrity; (b) mitigates risks to researchers and affiliated institutions concerning potential legal actions on the part of individuals who may feel harmed by their participation in studies; and (c) improves external validity and relevance of a study, which may result in a greater chance of obtaining funding and having greater acceptance on the part of the scientific community.

4.2 Approach to evaluate the quality of the research process in Accounting

Considering the importance of the connection between research quality criteria and the respondents' perceptions regarding their own adherence to the items when doing research, this study suggests a set of elements to assess research processes in Accounting based on the research process stages (key-characteristics), attributes/relationship, general research quality criteria, and research domains.

The Logic model's structure (Kellog Foundation, 2014; Ladd & Jernigan, 2006) was used to outline an approach to assess the quality of the research process in Accounting. This model is widely used to assess programs and projects, both processes and outcomes (or variation). A logic model used to assess processes is structured in inputs, processes/activities, and process indicators/outputs. This approach, based on the Logic model, considers inputs to be stages of the research process (Figure 5). Initially, this approach considered the three stages proposed by Brinberg and McGrath (1985), which in this study are called first (planning), second (implementation), and third (monitoring results or only results). Additionally, the nine key-characteristics (Mays & Pope, 2006; Spencer *et al.*, 2003) were grouped into three stages as the following: (1) design and research subjects; (2) data collection, analysis, results and report; and (3) reflectivity/neutrality, ethics and auditability.

The attributes and relationships adopted in the instrument in the Delphi's data collection were based on the literature and used to characterize the processes and activities proposed in the logic model. In this study, these attributes and relationships properly represent the processes considering that the development of the instrument itself was based on the respondents' research practices, that is, on the activities performed at the different stages of the participants' studies. The items were grouped according to the key-characteristics and quality criteria to which they are associated. In regard to the indicators of processes/outputs, we allocated criteria that were used to group the items, that is, they are associated with attributes/relationships that are defined as processes/activities. Finally, the VNS domains were used to add some potential implications when these criteria are violated or ignored.

Finally, note that this approach (Figure 5) is intended to contribute to the discussion of quality criteria used to develop and conduct research (process), thus is less focused on final results (product's assessment). This approach is only an attempt to help researchers in the accounting field to self-assess the quality of their studies and is not intended to replace other frameworks designed to evaluate research.

Inputs		Processes/Activities	Process Indicators/Outputs	
Stage		Processes/Activities	Criteria	Some potential implications
1st. Planning	Design	• Strategy that is useful to the purpose • Clear overview • Overview x strategy • Establishment of objective/problem • Overview x theory • Reasons for choosing technique	Appropriateness	**Substantive Domain** • Gap not very evidence to justify the research • Poor contribution to advance of knowledge • Inappropriate research strategy • Waste of resources due to inappropriate use of time • Fail to report the study's impact
		• Access to data • Time restriction • Strategy x target audience	Feasibility	
	Research subjects	• Criteria for design/selection of subjects • Representativeness of subjects	Internal validity	
2nd. Implementation	Data Collection	• Record each research stage • Record divergent events	Rigor	
		• Subjects' voluntary participation • Subjects' formal consent	Integrity	
	Analysis	• Describe the nature and shape of data	Integrity	**Conceptual Domain** • Lack of clarity of the main concepts related to the study • Lack of a theory to support data analysis • Difficulty to justify hypotheses based on the literature
		• Describe tools and procedures • Implicit/explicit link – findings and objectives	Rigor	
		• Context x impact for data analysis • Other views to know the context	Reliability	
		• Significance of data to achieve objectives	Relevance	
	Findings	• Path to reach conclusions • Link findings x evidence • Check links findings x purpose • Impact of the nature of divergences	Internal validity	
		• Compare results with those reported by other studies	External validity	
		• New fields based on findings • Insights for the field of knowledge	Contribution	
		• Previous findings x hypotheses	Relevance	
		• Context to enable replications	Reliability	
	Report	• Discussion of the impact for knowledge	Impact	**Methodological Domain** • Failure in ethical procedures • Difficulties that impede replication • Limitations to make comparisons • Study context poorly described • Insufficient record of the research steps • Conclusions are not supported by findings • Risk of impeding the publications of results
		• Report limitations • Study borders	Contribution	
		• Reasons for limitations	Integrity	
		• Literature review x main concepts • Theory underlying propositions • Conclusions x objective • Draft peer-review	Internal validity	
		• Explicit possibility of generalization	External validity	
3rd. Results	Reflectivity/ Neutrality	How to deal with errors and biases Influence of the research team	Integrity	
	Ethics	Institutional Review Board (IRB) to register studies Rules to register studies to IRB Formal submission to an IRB Mandatory submission to an IRB	Rigor	
		Adopt an ethics code Formal respect to human subjects Strategies to minimize harm Confidentiality of participants' data	Integrity	
	Auditability	Record changes in design Record reasons for changes Keep database for checks Keep documents to decrease risks	Reliability	

Figure 5. Logic model for research process evaluation

5. Conclusions

One of this study's conclusions is that certain practices currently used in the development of scientific studies in Accounting in Brazil do not properly meet the attributes of good quality research that are described in the literature. This is evident in the large amplitude of answers obtained in the various attributes, as well as low acceptance of elements concerning ethics, for which Brazilian law does not allow for flexibility. Another conclusion is that ethical criteria lack clarity and need to be complied with in accounting research. We also concluded that the Brazilian institutional environment contributes to a decreased quality of scientific studies in the Accounting field due to institutional failures that may impact the integrity of research, supporting low levels of acceptance of elements concerning rigor and ethics, moderate levels of acceptance for items related to integrity or feasibility in addition to failure in complying with the need to submit projects that involve human subjects to an institutional review board.

One of the conclusions of this study is that certain practices currently used in the construction process of scientific production in accounting in Brazil do not adequately meet the quality attributes of good research described in the literature. This is evident in the wide ranges of answers obtained with regard to the various attributes, as well as in the low acceptance of elements related to ethics, for which Brazilian law does not permit flexibility. Another conclusion is that there is a lack of clarity on ethical criteria and the need to meet them in accounting research. Furthermore, it was concluded that the Brazilian institutional environment contributes to a lower quality of scientific production in Accounting as a result of institutional failures that may impact research integrity, supported by low levels of acceptance of elements related to rigor and ethics, medium levels of acceptance of integrity and feasibility issues, and failure to comply with the obligation to submit research projects involving human beings to an Institutional Review Board.

This study presents limitations, either due to the Delphi technique chosen, which does not allow experts to interact, and potential biases due to the number of panel experts. Another limitation refers to the choice of general criteria, as there may be discrepancies concerning stakeholders' assessment.

This study's contributions include the fact that scientific research assessments, both within the field and among graduate programs, have focused on the results, that is, on the final product. Such an analysis, however, has limitations and fails to reveal to graduate programs and researchers potential failures in the research process. This study is expected to entail implications for the field as it supports researchers to self-assess their studies, identifying deficiencies and limitations, leading to greater acceptance of papers and shortened submission to periodicals. The Logic model is expected to assess the research process and contribute to encourage the target audience to reconsider research strategies and reorient practices that are not consistent with good quality research. Programs can also use the results to improve the process of training new researchers in order to mitigate potential failures in the future.

Given the evidence presented and the model proposed, we suggest that graduate programs intensify training that concerns ethical issues to decrease failures in the conduction of studies. Failure in complying with these requirements may compromise the integrity and quality of studies, possibly impeding the publication of studies. Another suggestion is that programs impose the condition that theses and dissertations' projects involving human subjects be submitted to an institutional review board in order to be accepted for qualification and defense. Future studies are suggested to address the perceptions of graduate students regarding the attributes/relationships addressed here, in order to identify potential errors and failures in the training of new researchers in the field. Additionally, another relevant investigation would be discussing the role played by editors and referees as those responsible for validating the quality of research in terms of their perception of what configures good quality research.

References

Antunes, M. T. P., Mendonça, O. R., Neto, Oyadomari, J. C. T., & Okimura, R. T. (2011). Conduta ética dos pesquisadores em contabilidade: diferenças entre a crença e a práxis. *Revista Contabilidade & Finanças*, *22*(57), pp. 319–337. DOI: https://doi.org/10.1590/S1519-70772011000300006

Aragão, I. R. B. N., Oliveira, J. R. S., & Lima, G. A. S. F. de. (2014). Resonance of articles and impact factor of Brazilian Accounting Journals. *Revista de Educação E Pesquisa Em Contabilidade (Journal of Education and Research in Accounting)*, *8*(1), pp. 5–21. DOI: https://doi.org/10.17524/repec.v8i1.1011

Ballas, A., & Theoharakis, V. (2003). Exploring diversity in accounting through Faculty journal perceptions. *Contemporary Accounting Research*, *20*(4), pp. 619–644. DOI: https://doi.org/10.1506/MLWH-KBTM-ET47-LYKH

Beck, C. T. (1993). Qualitative Research: The Evaluation of Its Credibility, Fittingness, and Auditability. *Western Journal of Nursing Research*, *15*(2), pp. 263–266. DOI: https://doi.org/10.1177/019394599301500212

Brasil. Resolução 466, de 12 de dezembro de 2012 (2012). Conselho Nacional de Saúde.

Brasil. Norma Operacional CNS No 001/2013 (2013). Conselho Nacional de Saúde.

Brinberg, D., & McGrath, J. E. (1985). *Validity and the research process*. Newbury Park: Sage Publications.

Brinn, T., Jones, M., & Pendlebury, M. (2001). Why do UK Accounting and Finance Academics not Publish in Top US Journals? *The British Acc. Review*, *33*(2), pp.223–232. DOI: https://doi.org/10.1006/bare.2001.0160

Brown, L. D., & Gardner, J. C. (1985). Using Citation Analysis to Assess the Impact of Journals and Articles on Contemporary Accounting Research (CAR). *Journal of Accounting Research*, *23*(1), pp. 84–109. http://doi.org/10.2307/2490908

Brown, L. D., & Huefner, R. J. (1994). The Familiarity with and Perceived Quality of Accounting Journals: Views of Senior Accounting Faculty in Leading U.S. MBA Programs. *Contemporary Accounting Research*, *11*(1), pp. 223–250. DOI: https://doi.org/10.1111/j.1911-3846.1994.tb00442.x

Carmona, S. (2006). Performance Reviews, the Impact of Accounting Research, and the Role of Publication Forms. *Advances in Accounting*, *22*(6), pp. 241–267. https://www.sciencedirect.com/science/article/abs/pii/S088261100622011X?via%3Dihub

Chan, K. C., Chang, C. H., Tong, J. Y., & Zhang, F. (2012). An analysis of the accounting and finance research productivity in Australia and New Zealand in 1991-2010. *Accounting and Finance*, *52*(1), pp. 249–265. DOI: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-629X.2011.00440.x

Clarkson, P. M. (2012). Publishing: art or science? Reflections from an editorial perspective. *Accounting & Finance*, *52*(2), pp. 359–376. DOI: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-629X.2012.00472.x

Davidson, E. J. (2005). *Evaluation methodology basics: the nuts and bolts of sound evaluation*. (Sage Publications, Ed.) (1st ed.). Thousand Oaks: Sage Publications.

Denscombe, M. (2010). *The Good Research Guide*. Open University Press.

Doyle, J., & Arthurs, A. (1995). Judging the quality of research in business schools: the UK as a case study. *Omega*, *23*(3), pp. 257–270. DOI: https://doi.org/10.1016/0305-0483(95)00013-E

Dunbar, A. E., & Weber, D. P. (2014). What influences accounting research? A citations-based analysis. *Issues in Accounting Education*, *29*(1), pp. 1–60. DOI: https://doi.org/10.2308/iace-50603

Elwyn, G., O'Connor, A., Stacey, D., Volk, R., Edwards, A., Coulter, A., … Whelan, T. (2006). Developing a quality criteria framework for patient decision aids: online international Delphi consensus process. *BMJ (Clinical Research Ed.)*, *333*(7565), pp. 417. DOI: https://doi.org/10.1136/bmj.38926.629329.AE

European Science Foundation. (2011). *The European Code of Conduct for Research Integrity*. Strasbourg.

Evans, J. H., Feng, M., Hoffman, V. B., Moser, D. V., & Van der Stede, W. A. (2015). Points to Consider When Self-Assessing Your Empirical Accounting Research. *Contemporary Accounting Research*, *32*(3), pp. 1162–1192. http://doi.org/10.1111/1911-3846.12133

Fundação de Amparo à Pesquisa do Estado de São Paulo. (2012). *Code of good scientific practice*. Recuperado 2012 de http://www.fapesp.br/boaspraticas/FAPESP-Codigo_de_Boas_Praticas_Cientificas_2014.pdf

Jones, M. J., Brinn, T., & Pendlebury, M. (1996). Judging the quality of research in business schools: A comment from accounting. *Omega*, *24*(5), pp. 597–602. DOI: https://doi.org/10.1016/0305-0483(96)00036-9

Kelbaugh, B. M. (2003). *Exploration of teamwork in extension: identifying indicators of success using a modified delphi study (Ph.D. dissertation)*. The Ohio State University. https://etd.ohiolink.edu/!etd.send_file?accession=osu1054604099&disposition=inline

Kellog Foundation. (2014). W.K. Kellogg Foundation Logic Model Development Guide. Retrieved September 20, 2014, from https://www.wkkf.org/resource-directory/resource/2006/02/wk-kellogg-foundation-logic-model-development-guide.

Ladd, S., & Jernigan, J. (2006). *Evaluation guide: developing and using a logic model*. Retrieved from https://www.cdc.gov/dhdsp/docs/logic_model.pdf.

LeCompte, M. D., & Goetz, J. P. (1982). Problems of Reliability and Validity in Ethnographic Research. *Review of Educational Research*, *52*(1), pp. 31–60. DOI: https://doi.org/10.3102/00346543052001031

Libby, R., Bloomfield, R., & Nelson, M. W. (2002). Experimental research in financial accounting. *Accounting, Organizations and Society*, *27*(8), pp. 775–810. DOI: https://doi.org/10.1016/S0361-3682(01)00011-3

Lowe, A., & Locke, J. (2005). Perceptions of journal quality and research paradigm: results of a web-based survey of British accounting academics. *Accounting, Organizations and Society*, *30*(1), pp. 81–98. http://doi.org/10.1016/j.aos.2004.05.002

Lowensohn, S., & Samelson, D. P. (2006). An Examination of Faculty Perceptions of Academic Journal Quality within Five Specialized Areas of Accounting Research. *Issues in Accounting Education*, *21*(3), pp. 219–239. DOI: https://doi.org/10.2308/iace.2006.21.3.219

Martins, O. S., & Lucena, W. G. L. (2014). Produtivismo acadêmico: as práticas dos docentes dos programas de pós-graduação em contabilidade. *Advances in Scientific and Applied Accounting*, *7*(1), pp. 66–96. DOI: http://doi.org/10.14392/asaa.2014070103

Mays, N., & Pope, C. (2006). Quality in qualitative health research. In *Qualitative research in health care* (3rd ed., pp. 82–101). London: Blackwell Pub./BMJ Books.

Mendonça, O. R., de Neto., Riccio, E. L., & Sakata, M. C. G. (2009). Dez anos de pesquisa contábil no Brasil: análise dos trabalhos apresentados nos Enanpads de 1996 a 2005. *Revista de Administração de Empresas*, *49*(1), pp. 62–73. DOI: http://dx.doi.org/10.1590/S0034-75902009000100008

Meyrick, J. (2006). What is good qualitative research? A first step towards a comprehensive approach to judging rigour/quality. *Journal of Health Psychology*, *11*(5), pp. 799–808. DOI: 10.1177/1359105306066643

Michaelis. (2012). *Michaelis Moderno Dicionário da Língua Portuguesa (versão online)*. Editora Melhoramentos.

Miranda, G. J., Azevedo, R. F. L., & Martins, G. de A. (2011). Teses das teses em contabilidade na USP. *REPEC Revista de Educação e Pesquisa em Contabilidade*, *5*(2, mai/ago), pp. 21–42. DOI: https://doi.org/10.17524/repec.v5i2.342

National Institutes of Health. (1979). The Belmont Report. *The Belmont Report Ethical Principles and Guidelines for the Protection of Human Subjects of Research*, pp. 4–6.

Northcott, D., & Linacre, S. (2010). Producing spaces for academic discourse: The impact of research assessment exercises and journal quality rankings. *Australian Accounting Review*, *20*(1), pp. 38–54. DOI: http://doi.org/10.1111/j.1835-2561.2010.00079.x

Office of the Associate Director for Science (OADS). (2012). *CDC Guidance on Scientific Integrity*. Retrieved from http://www.cdc.gov/od/science/docs/CDCSIGuide_021612.pdf

Reiter, S. A., & Williams, P. F. (2002). The structure and progressivity of accounting research: the crisis in the academy revisited. *Accounting, Organizations and Society*, *27*(6), pp. 575–607. DOI: https://doi.org/10.1016/S0361-3682(01)00050-2

Rossi, P. H., Lipsey, M. W., & Freeman, H. E. (2004). *Evaluation: a systematic approach* (7th ed.). Thousand Oaks: Sage Publications.

Schwartzman, S. (1988). Qualidade e relevância da pesquisa universitária: 10 teses para discussão. Retrieved from http://www.schwartzman.org.br/simon/10teses.htm

Shamoo, A., & Resnik, D. (2003). *Responsible conduct of research*. New York: Oxford University Press.

Spencer, L., Ritchie, J., Lewis, J., & Dillon, L. (2003). Quality in Qualitative Evaluation: A framework for assessing research evidence A Quality Framework. *Social Research*, *108*(1), p.170. DOI: 10.1093/eurpub/ckv219.

Taylor, J. (2011). The assessment of research quality in UK universities: Peer review or metrics? *British Journal of Management*, *22*(2), pp. 202–217. Available at SSRN: https://ssrn.com/abstract=1841224 or http://dx.doi.org/10.1111/j.1467-8551.2010.00722.x

Valentine, J. C. (2009). Judging the quality of primary research. In H. Cooper, L. V Hedges, & J. C. Valentine (Eds.), *The Handbook of Research Synthesis and Meta-Analysis* (pp. 129–146). Russell Sage Foundation. https://psycnet.apa.org/record/2009-05060-007

Van der Stede, W. a., Young, S. M., & Chen, C. X. (2005). Assessing the quality of evidence in empirical management accounting research: The case of survey studies. *Accounting, Organizations and Society*, *30*(7–8), pp. 655–684. http://doi.org/10.1016/j.aos.2005.01.003

Webster, J., & Watson, R. T. (2002). Analyzing the Past to Prepare for the Future: Writing a Literature Review. *MIS Quarterly*, *26*(2), pp. xiii–xxiii. http://doi.org/10.1.1.104.6570

Weiss, C. H. (1997). *Evaluation* (2nd ed.). New Jersey: Prentice Hall.

Whittemore, R., Chase, S. K., & Mandle, C. L. (2001). Validity in qualitative research. *Qualitative Health Research*, *11*(4), pp. 522–537. DOI: 10.1177/104973201129119299

Williams, P. F. (2014). The Myth of Rigorous Accounting Research. *Accounting Horizons*, *28*(4), pp. 869–887. http://doi.org/10.2308/acch-50880

Yarbrough, D. B., Shulha, L. M., Hopson, R. K., & Caruthers, F. A. (2011). *The program evaluation standards: a guide for evaluators and evaluation users* (3rd ed.). Thousand Oaks: Sage Publications.

Attachment A

Final classification of the Modified Delphi Technique according to key-characteristics (research process stages), attributes/relationships, size, number of valid answers, round in which consensus was obtained, score, score relative percentage, ranking, levels of agreement and associated criteria

Characteristic	Attribute/relationship	N	R	Score	%	Rank	Level	Criterion
Data Collection	Subjects' voluntary participation	23	2	222	96.52	1º	Strong	Integrity
Design	Objective/problem explicitly reported	41	1	395	96.34	2º	Strong	Adequacy
Ethics	Confidentiality of participants' data	26	2	248	95.38	3º	Strong	Integrity
Report	Conclusions x objective	41	1	377	91.95	4º	Strong	Internal validity
Findings	Compare results with other studies	41	1	373	90.98	5º	Strong	Internal validity
Findings	Check links Findings x purpose	41	1	373	90.98	6º	Strong	Internal validity
Report	Literature review x main concepts	41	1	373	90.98	7º	Strong	Internal validity
Report	Theory to support propositions	41	1	371	90.49	8º	Strong	Internal validity
Design	Strategy is useful to the purpose	41	1	369	90.00	9º	Strong	Adequacy
Findings	Previous findings x hypotheses	41	1	369	90.00	10	Strong	Relevance
Auditability	Safeguard database for checks	41	1	367	89.51	11	Strong	Reliability
Design	Reasons to the choice of technique	41	1	366	89.27	12	Strong	Adequacy
Findings	Link Findings x evidence	41	1	363	88.54	13	Strong	Internal validity
Data Collection	Records of each research step	41	1	361	88.05	14	Strong	Rigor
Findings	News fields based on findings	41	1	361	88.05	15	Strong	Contribution
Findings	Path to reach conclusions	41	1	359	87.56	16	Strong	Internal validity
Design	Clear overview guiding the study	41	1	358	87.32	17	Strong	Adequacy
Analysis	Description of tools and procedures	41	1	358	87.32	18	Strong	Rigor
Subjects	Criteria for design/selection of subjects	41	1	352	85.85	19	Strong	Internal validity
Design	Access to data	41	1	351	85.61	20	Strong	Feasibility
Analysis	Description of the nature & shape of data	41	1	351	85.61	21	Strong	Integrity
Analysis	Significance of data to achieve objectives	41	1	351	85.61	22	Strong	Relevance
Findings	Context to allow replication	41	1	351	85.61	23	Strong	Reliability
Design	Overview x strategy	37	2	316	85.41	24	Strong	Adequacy
Report	Report limitations	37	2	315	85.14	25	Strong	Contribution
Design	Overview x theory	41	1	349	85.12	26	Strong	Adequacy
Analysis	Implicit/explicit links-findings x objective	41	1	349	85.12	27	Strong	Rigor
Data Collection	Records regarding diverge events	28	2	237	84.64	28	Strong	Rigor
Report	Explicit possibility of generalization	41	1	345	84.15	29	Strong	External validity
Auditability	Records of changes in design	29	2	242	83.45	30	Strong	Reliability
Analysis	Context x Impact for data analysis	41	1	341	83.17	31	Strong	Reliability
Findings	Insights for the field of knowledge	37	2	302	81.62	32	Strong	Contribution
Ethics	Strategy to minimize potential harm	26	2	212	81.54	33	Strong	Integrity
Report	Discussion of the impact for knowledge	37	2	301	81.35	34	Strong	Impact
Ethics	IRB to register studies	26	2	207	79.62	35	Strong	Rigor
Report	Study boundaries	37	2	294	79.46	36	Strong	Contribution
Ethics	Formal respect human subjects	25	2	195	78.00	37	Strong	Integrity
Auditability	Keeping documents to decrease risks	25	2	194	77.60	38	Strong	Reliability

Characteristic	Attribute/relationship	N	R	Score	%	Rank	Level	Criterion
Subjects	Subjects representativeness	37	2	287	77.57	39	Strong	Internal validity
Ethics	Adopting a code of ethics	26	2	187	71.92	1st	Moderate	Integrity
Refl./Neutr.	How to deal with errors and biases	37	2	266	71.89	2nd	Moderate	Integrity
Design	Research strategy x target population	37	2	264	71.35	3rd	Moderate	Feasibility
Design	Time restriction	37	2	261	70.54	4th	Moderate	Feasibility
Ethics	Rules to register studies (IRB)	25	2	176	70.40	5th	Moderate	Rigor
Refl./Neutr.	Influence of the research team	37	2	249	67.30	6th	Moderate	Integrity
Report	Draft peer-review	37	2	229	61.89	1st	Low	Internal validity
Ethics	Formal submission to IRB	25	2	144	57.60	2nd	Low	Rigor
Ethics	Mandatory submission to IRB	25	2	142	56.80	3rd	Low	Rigor
Auditability	Record reasons for changes	29	2	237	81.72	1st	Not obtained	Reliability
Data Collection	Subjects' formal consent	26	2	212	81.54	2nd	Not obtained	Integrity
Report	Reasons for limitations	37	2	293	79.19	3rd	Not obtained	Integrity
Findings	Impact da the nature of divergences	37	2	289	78.11	4th	Not obtained	Internal validity
Analysis	Other perspective to know the context	37	2	278	75.14	5º	Not obtained	Reliability