Quantitative Research in Management Accounting: A Proposed Typology and Implications for Research Design

Andson Braga de Aguiar

Universidade de São Paulo

Daniel Magalhães Mucci

Universidade de São Paulo

Myrna Modolon Lima

Universidade de São Paulo

Abstract

This paper aims to propose a typology of quantitative research in management accounting based on three key design features: the presence of a control group, the availability of pre-structured data and the ability to extrapolate results. The presence of a control group in a study allows the creation of a counterfactual that increases the claims in favour of causal relationship and, consequently, contributes to enhance internal validity. The use of pre-structured data facilitates data collection, but at the same time, creates challenges in the operational definition of the key variables and, thus, in construct validity. The ability to extrapolate results has implications for the generalizability to other persons or settings, affecting the external validity. Based on these three design features, we propose eight quantitative research method alternatives that management accounting researchers can use: (1) laboratory experiments, (2) crowdsourcing experiments, (3) field experiments, (4) natural Experiment, (5) pre-structured archival study, (6) proprietary archival study, (7) large scale survey and (8) single entity survey. We discuss the main implications of the proposed typology for research design—emphasizing validity concerns—and provide examples of studies that fit to each research method of this typology. The contribution of this study is twofold. First, this paper helps expand the toolkit available to management accounting scholars, specially, junior scholars, in the design of their studies. Second, this study provides insights on the validity concerns that management accounting researchers face when selecting a particular research design, in particular, research designs that are still incipient in the field, especially in Brazil.

Keywords: Quantitative research methods; Management Accounting; Experiment; Archival Study; Survey.

1 Introduction

A challenge management accounting researchers typically face is the choice of the research method to be used in a particular study, being it a master thesis, a dissertation, or a research article. The general guideline is to select a research method that is better aligned with the research question to be investigated in the study (Kinney, 2019). While helpful, this guideline does not inform the researcher of the different research methods that could be used to

22º USP International Conference in Accounting
"Accounting in favor of sustainable development"
22º
USP International
Conference in Accounting
São Paulo 27 a 29 de julho 2022.

address a particular research question, the advantages and disadvantages of each available method to address that research question, or the most appropriate method among the available ones. In other words, management accounting researchers may not be sufficiently well-informed when it comes to the selection of the research method to be used in a particular research project. The lack of knowledge about the available research methods has fundamental consequences for the advancement of a research field; particularly, the lack of familiarity with the diverse options of available methods to address the research question can actually narrow down the scope of research questions that management accounting researchers investigate as well as overlook the complementarities of using different research methods to address the same research question (Bloomfield, Nelson, & Soltes, 2016).

The main purpose of this study is to propose a typology of quantitative research in management accounting. We focus on quantitative research because the use of quantitative research methods is prevalent in the accounting research, either in international (Hesford, Lee, Van der Stede, & Young, 2006) as well as in local journals (Aguiar, 2018; Nascimento, Junqueira, & Martins, 2010). In addition, the attributes and validity criteria for quantitative studies differ in type and importance from qualitative research methods. Likewise, we focus on management accounting because a recent typology has been proposed to help accounting researchers to select appropriate methods based on the data gathering tasks involved (Bloomfield et al., 2016). While useful in general, this typology does not consider all the singularities that management accounting researchers face in the process of selecting the particular research method to be used for the research project. For instance, even when selecting a survey method, management accounting researchers have still to decide the level of analysis and representativeness of the sample (Van der Stede, Young, & Chen, 2006).

The proposed typology in this study is based on three key design features. First, we separate experimental from non-experimental studies if the research design includes a control group. Second, similar to Bloomfield et al. (2016), we separate studies with from those without pre-structured data based on the availability of dataset prior to the research. Finally, we separate research projects whose results can or cannot be extrapolated from the observed sample to other samples, located in other places and at other times (Shadish, Cook, & Campbell, 2002; Trochim, Donnelly, & Arora, 2016). Building on these three criteria, we propose eight different quantitative research methods that management accounting researchers can use when addressing their research questions.

Secondarily, we also discuss the main implications of the proposed typology for research design in terms of the validity framework. Precisely, the presence of a control group in a study allows the creation of a counterfactual that increases the claims in favour of causal relationship and, consequently, contributes to enhance internal validity (Shadish et al., 2002; Trochim et al., 2016). The use of pre-structured data creates challenges in the operational definition of the key variables and, thus, in construct validity, since pre-structured data are typically not collected for academic purposes (Das, Jain, & Mishra, 2016; Moers, 2006). Finally, the ability to extrapolate results from the observed sample to other persons or settings has implications for the generalizability and, thus, affects the external validity of the study (Shadish et al., 2002; Trochim et al., 2016).

The contribution of this study is twofold. First, we propose a typology of quantitative research in management accounting based on three key design features that can help expand

the toolkit available to scholars, specially, junior scholars, in the design of their studies. For instance, we highlight that the choice of a survey research design requires from management accounting researchers to make a subsequent choice related to the unit of observation between a large-scale survey or a single-entity survey. The research methods included in the proposed typology can help management accounting researchers increase the familiarity with the available alternatives, some of which that are still incipient in Brazil, such as is the case for experimental designs (Aguiar, 2017; Nascimento et al., 2010). Second, the discussion about the implications of the proposed typology provides insights on the trade-offs among the validity types management accounting researchers have to pay attention when selecting a particular research design. For instance, the choice of a research method including a control group may favour internal validity while, at the same time, pose challenges in terms of external validity if results obtained from the observed sample are not generalizable. In other words, the choice of a particular research method to address a research question implies a simultaneous choice of the validity types that will be favoured in the study and the validity types that will create challenges for the researcher.

In the next section we broadly discuss philosophical perspectives of research and underlying consequences for a methodological design, positioning that quantitative management accounting research usually fits to positivism. We then present the three design features that are used to develop the proposed typology. Following, we introduce the proposed typology including eight different research methods and provide examples of studies that fit to each research method of this typology. Finally, we conclude by discussing the implications of the proposed typology for research design.

2 Research Onion

Saunders, Lewis, and Thornhill (2019) introduced a general scheme, called the research onion, in which they deployed several aspects related to methodological decisions. The scheme starts from the understanding of research philosophies and approaches to theory to the methodological choices and strategies per se. For the purpose of this study, we actually take into account a part of the whole research onion due to the nature of quantitative research as well as the particular research methodology discussed afterwards. Our scope in this paper addresses Positivism as the main research philosophy, which means that objectivism is the prevailing/dominant view in both the assumptions about the nature of reality named ontology (i.e., real and universal) and the assumptions about knowledge named as epistemology (i.e., observable, factual and generalizable) (Burrell & Morgan, 1979; Crotty, 1998; Saunders et al., 2019). Regarding the approach to theory, methodologists debate three different approaches being the deductive, inductive and abductive. Due to the positioning of positivist researchers and underlying nature of their studies, quantitative researchers usually employ a deductive approach, which means that research is anchored in theories and is subject to tests of falsification or verification through hypotheses (Saunders et al., 2019).

After these two first philosophical understandings of research and the role of theory, Saunders et al. (2019) present what they call methodological choices and strategies, what in fact is the focus of our paper. Hence, we debate different quantitative methodologies that are broadly characterized by these authors as experiments, archival studies and surveys. These quantitative research methods also can be developed considering different time-horizons being a cross-section (common in surveys) and longitudinal (common in archival studies), which are

22º
USP International
Conference in Accounting

São Paulo 27 a 29 de julho 2022.

22º *USP International Conference in Accounting*
"*Accounting in favor of sustainable development*"

also one of the layers from the research onion. Although it is not our focus, another layer of the research onion is the use of a mono or multi-method research design.

In sum, our paper discusses three design features of quantitative research in management accounting. We highlight these explanations in advance to point out the existence of other "*research layers*" that are not the object of this paper, but that also deserve attention as they forego the discussions about the features of research methods/strategies. Hence, our debate leaves from this point being that quantitative research in management accounting thrives from positivist paradigm (objectivism in terms of ontology and epistemology) as well as from a deductive approach from theory.

3 Design Features

The main purpose of this study is to propose a typology of quantitative research in management accounting. The proposed typology is based on three key design features: control group, pre-structured dataset, and extrapolable results. In this section, we briefly describe each of the three design features and highlight how the choices associated with each of them lead to different research designs. We end this section introducing our proposed typology.

3.1 Control Group

The first design feature is associated with the existence of a **control group** as part of the research design. The presence of a control group is the hallmark of experimental studies (Shadish et al., 2002; Trochim et al., 2016), so that this design choice allows us to separate experimental (and quasi-experimental) from non-experimental studies[1]. A control group[2] represents a group of respondents or participants who are comparable to the non-control group or, more precisely, to the treatment group, in every way possible, with the main difference being that the control group is not exposed to the treatment or program (Trochim et al., 2016).

For instance, in a study examining the effect of providing performance-based compensation on employee motivation, the treatment group would include participants who receive a performance-based compensation, while the control group would include participants who do not receive a performance-based compensation. For such research design, the researcher would compare employee motivation between the treatment group and the control group and examine whether or not the fact that a performance-based compensation is provided would alter (increase or decrease) employee motivation.

The main benefit of using control group is that researchers can mitigate the likelihood of alternative explanations for a causal relationship and, thus, increase the internal validity of the study (Shadish et al., 2002; Trochim et al., 2016). In other words, the presence of a control group creates a useful counterfactual inference, essential for research studies interested in establishing causal relationships (Lonati, Quiroga, Zehnder, & Antonakis, 2018). For that, it is essential that participants in the control group are as similar as possible to participants in the treatment group (Trochim et al., 2016).

---

[1] Both experimental and non-experimental designs include a control group. The main difference between experimental and quasi-experimental designs is the random allocation of participants to the experimental conditions in the former design, while the latter design uses non-equivalent groups (Trochim et al., 2016).
[2] A control group can also be denominated as the comparison group or baseline condition.

4

3.2 Pre-structured Dataset

Our second design feature is similar to what Bloomfield et al. (2016) label as **pre-structured data**. The presence of preestructured data is the key characteristic of archival studies (Moers, 2006). This design choice allows us to separate observational from non-observational studies (Oehlert, 2003). Pre-structured data typically are data gathered and stored prior to the beginning of the research, intended for other purposes than for academic purposes (Das et al., 2016; Moers, 2006).

For instance, in a study examining the association between the weight placed on different performance measures and investment decisions, the researcher could obtain public available data using database platforms such as Compustat and Formulario de Referência. Alternatively, the research could get access to confidential data granted by an organization or an owner of the data, such as third-party surveys. In either case, researchers conducting studies with pre-structured data would have to identity appropriate proxies from the available data to measure the relevant variables of interest to test for associations between them.

The main benefit of using pre-structured data is the relatively easiness for researchers obtain a large dataset, at a relatively low cost, to examine a wide variety of relevant research questions (Das et al., 2016; Moers, 2006). In other words, the researchers using pre-structured data can more easily collect relevant data from several respondents, be organizations, organizational units, or individual respondents.

3.3 Extrapolable Results

Our third and final design feature is associated with the extent to which the study has **extrapolable results**. Extrapolable results refer to the ability of a study to make inferences from the observed sample to other (non-observed) samples, located in other places and at other times (Shadish et al., 2002; Trochim et al., 2016). The higher the ability of a study to make inferences to other samples, the more generalizable are the results from this study.

For instance, a study examining the effect of a communicated value statement on middle level employees' behavior can use a sample of undergraduate students as the unit of observation. In such a study, the ability to extrapolate the results from the observed sample (i.e, undergraduate students) to other samples (i.e., middle level employees) is challenging. Similarly, in a study examining the relationship between the use of management controls, whether diagnostic or interactive, and managers' creative task performance in a single company, it is challenging to extrapolate results from the sample of managers in this company to other managers in other companies.

The main benefit of extrapolating results by using more generalizable samples is to increase external validity (Shadish et al., 2002). This allows the researcher more confidently makes inferences from the study's results to other individual units of observation (e.g., individual, groups, organizational units, or organizations) in other places at other time than just the units represented by the study`s sample (Trochim et al., 2016).

3.4 Proposed Typology

Building on the three aforementioned criteria, we propose eight quantitative research method alternatives that management accounting researchers can use when addressing their research questions (Figure 1). The proposed typology resembles a decision tree that guides

5

researchers to the alternative research methods based on the choices associated with each of the three criteria.
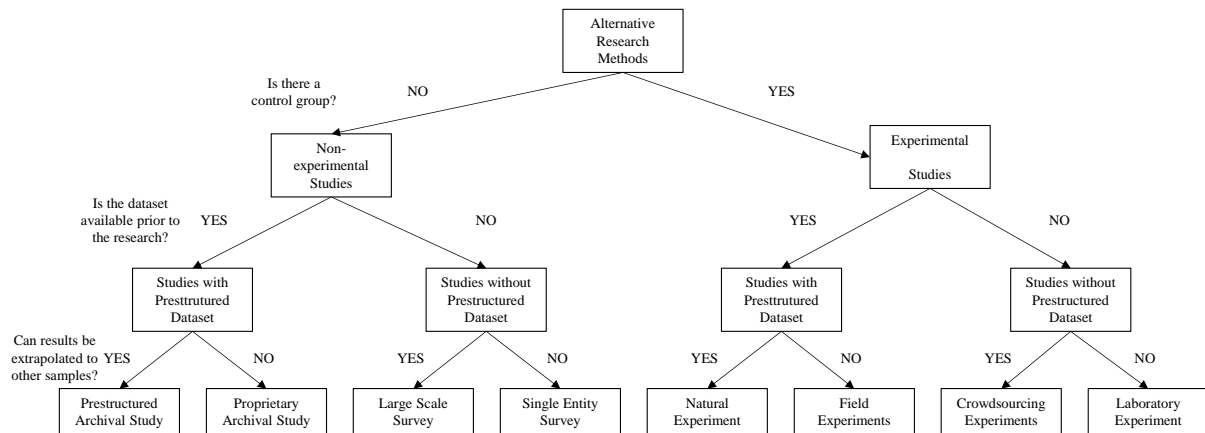

Figure 1. Proposed typology of quantitative research in management accounting

In the **first layer**, researchers have to decide whether or not their research design includes a control group. If a control group is included (right side of Figure 1), the researcher chooses to conduct an experimental study; on the contrary, a non-experimental study is chosen (left side of Figure 1). In the **second layer**, researchers have to decide whether or not they will use an available pre-structured dataset. For both experimental and non-experimental studies, the researcher can select a quantitative research method that either has an available pre-structured dataset or does not have an available pre-structured dataset. Finally, in the **third layer**, the decision is about the extent to which researchers will choose a unit of observation from whom results can be extrapolated. Whether experimental or non-experimental, with or without pre-structured dataset, researchers can select a sample that either does or does not allow for extrapolation of the results from the unit observed in the study to other units of observation.

4 Alternative Quantitative Research Methods

4.1 Laboratory experiments (Lab experiments)

Laboratory experiment is a common research method that employs a control group, given that they are more restrictive in scope and design compared to other research methods (Bloomfield et al., 2016; Swieringa & Weick, 1982). Mainly, lab experiments allow the research to establish a comparison between a control and a treatment, being ideal for theory testing (Oehlert, 2003). A lab experiment can be conducted in a lab or, for social scientists, in a room where the experimenter can ensure physical control over participants, such as a room with individual spaces that prevent participants from looking to other participants or to use any technological gadgets during the experiment. Overall, lab experiments are great at ensuring theory testing given that researchers can control any other influences that might harm the process.

Lab experiment displays the highest internal validity compared to alternative research methods because of the aforementioned control group, and given the amount of perceived control that researchers have during the experiment (Trochim et al., 2016). Lab experiment requires that participants are randomized between conditions, avoiding creating patterns among the conditions (e.g., gender patterns, cultural patterns) (Shadish et al., 2002; Trochim et al.,

6

22º
USP International
Conference in Accounting

São Paulo 27 a 29 de julho 2022.

22º *USP International Conference in Accounting*
"Accounting in favor of sustainable development"

2016). This means that the ideal participant for a lab experiment is a *tabula rasa*, so the only influence that the participant experiences is the one(s) being manipulated. However, it is not possible for researchers to ensure such a thing, so they have to rely on the best alternative, that is, to guarantee that participants have similar characteristics: same undergraduate course, approximately the same age, and similar backgrounds (Shadish et al., 2002; Sprinkle & Williamson, 2006; Trochim et al., 2016). In particular, students are well suited for experiments that require some specific knowledge (e.g., analysing financial statements or a Balanced Scorecard) but do not require specific experiences (Houghton & Hronsky, 1993; Liyanarachchi, 2007; Mortensen, Fisher, & Wines, 2012; Trottier & Gordon, 2018).

In a lab experiment, at least one independent variable is manipulated to emulate operationally what the researcher is willing to conceptually capture, so that this research design must be established ensuring construct validity (Asay, Guggenmos, Kadous, Koonce, & Libby, 2021). Comparatively, lab experiments make it easier for the researcher to think about ways to emulate the independent variable; however, some procedures to enhance construct validity must be observed. For example, the use of manipulation checks that are required to ensure different experimental conditions is a great way to check for construct validity (Shadish et al., 2002). Mainly, "manipulation checks support the existence of a relationship between causal and effect constructs by providing evidence of construct validity" (Rose, 2018, p. 127). In practice, the manipulation checks ensure that the reason for the study's outcome is the experimental treatment (Rose, 2018; Shadish et al., 2002). It is worth mentioning that one important aspect of the manipulation checks is that they are singular to each study; therefore, they have to be designed to fit the study's variables (Rose, 2018), which is consistent with studies that do not have a pre-structured dataset.

Of particular relevance for management accounting studies, lab experiments are suitable for analysing relationships at the employee-level or managerial level (Trottier & Gordon, 2018). Conversely, lab experiments are less suitable for examining higher hierarchical levels, such as Chief Executive Officers (CEOs), since their decision making process is less proximal to the student's reality compared to managers' or employees' decision making process. Overall, researchers conducting lab experiments should be aware that they will have difficulty in extrapolating their results to a larger sample since the unit of observation tends to be not representative. In other words, lab experiments display the lowest external validity among alternative quantitative research methods, since this research method concentrates more on establishing and ensuring internal validity with similar participants and diminishing the possibility of confounding variables than on extrapolating results to other units of observation (Asay et al., 2021).

An example of a lab experiment that focuses on managers' decision making process is Haesebrouck's (2021) study on the effects of effort to acquire information and the induced sense of psychological ownership from this action over managers' reporting. As part of their job, managers might have to acquire, synthesize and analyse data from several sources–exerting great effort–or they might easily acquire this information if the company has a good information sharing system (i.e., a software) in place. To test her predictions, the author used a 2x2 between-participants experimental design, where she manipulated acquisition of information (endowed vs. earned) and saliency of honesty in the reporting context (less vs. more salient honesty

22º USP International Conference in Accounting
"Accounting in favor of sustainable development"

22º
USP International
Conference in Accounting

São Paulo 27 a 29 de julho 2022.

context)[3]. The paper ensures the study's internal validity by following all the experimental procedures (i.e., randomization, participants' characteristics, direct observation of participants). The study also ensure that the effects on the dependent variable—managers' slack creation—are solely caused by the manipulated variables, ensuring construct validity, by asking participants several post-experimental questions to understand their decision making process. Finally, the author states that the experiment focuses on theory testing and results might not be extrapolated or generalized to other settings.

4.2 Crowdsourcing experiments (or online experiments)

A crowdsourcing experiment is a type of experiment that is relatively similar to the lab experiment. Crowdsourcing experiment can be seen just as a type of lab experiment since participants are aware that they are participating in an experiment (Bloomfield et al., 2016). Also, the two research designs manipulate at least one independent variable and tend to follow similar procedures to deal with construct validity threats. Crowdsourcing experiments and lab experiments, however, have two main differences: location and participants. In order to understand their unique benefits, it is then important to explore the differences between the two research method alternatives.

The first main difference regards the crowdsourcing experiment's location. Given that crowdsourcing platforms used in experimental studies are online (e.g., MTurk, Prolific, and CrowdFlower), there is a significant decrease on the researchers' control over the experiment. On the one side, both lab and crowdsourcing experiments include a control group. However, on the other side, crowdsourcing experiments involves a reduced control over other aspects of the research design. In particular, the researcher has less control over noises and bias such as lack of attention or effort, or even unqualified participants (Bentley, 2021). Researcher can also face fraudulent behaviour from respondents, what affects data integrity and reliability (Aguinis & Ramani, 2021; Dennis, Goodson, & Pearson, 2020). Mainly, researchers give up part of the laboratory experiment's ability to monitor participants in-person, which renders the crowdsourcing experiment lower internal validity compared to the lab experiment.

The second main difference regards the crowdsourcing experiment's participants which are a direct response to the lab experiment's lack of extrapolable results (i.e., external validity). Mainly, crowdsourcing platforms allow researchers to recruit human participants for research (Peer, Brandimarte, Samat, & Acquisti, 2017), which extends the participant pool beyond the undergraduate/graduate students from a certain geographical location. The unit of observation then tend to be more representative in crowdsourcing experiments. The platforms used in crowdsourcing experiments also allow researchers to filter some characteristics that can be helpful to narrow down participants, such as "years of work experience" or "management experience" (Palan & Schitter, 2018; Peer et al., 2017). This feature, of course, does not guarantee that the participants will be 100% accurate given that they can always be untruthful about their personal characteristics (Aguinis & Ramani, 2021; Bentley, 2021). However, the

---

[3] In the acquisition of information manipulation, managers might have to acquire the information by determining the cost of a project's implementation in the earned condition or they might read the implementation cost in the computer screen in the endowed condition. In the saliency of honesty in the reporting context manipulation, the report is framed either as an allocation of profits (less salient) or as revelation of costs (more salient).

22º USP International Conference in Accounting
"Accounting in favor of sustainable development"

22º
USP International
Conference in Accounting

São Paulo 27 a 29 de julho 2022.

platforms provide a more diverse participant pool, allowing researchers to investigate the theory among participants with different ages, cultures, and backgrounds.

An example of a crowdsourcing experiment is Murphy, Wynes, Hahn, and Devine 's (2019) study on the internal and external motivations to honest reporting. The authors use MTurk participants in an experiment where they have both opportunity and incentive to misreport in order to test the different motivations behind honesty. Since the authors' goal is to test the motivation behind participants' decisions, they have three manipulations: baseline (control group) vs. reward vs. punishment. The authors also explain that MTurk participants are suitable for the experiments given the nature of the tasks (i.e., not specialized). Consistent with crowdsourcing experiments, this study combines internal and external validity. The authors ensure the internal validity of the experimental method by applying the different manipulations, consistent with theory testing, and ensure the increase in external validity by using participants with different characteristics, extrapolating the results to other settings.

4.3 Field Experiments

A field experiment is a field study that uses the experimental method in which either interventions or effects are observed for longer periods of time (Lourenço, 2019; Bloomfield et al., 2016). In other words, it applies the experimental method–with control and treatment groups–in the field using longitudinal data. Field experiments are similar to other types of experiments in their development, requiring a counterfactual to establish whether the interventions are the sole responsible for the effects on employees, managers, or the company (Bloomfield et al., 2016; Lourenço, 2019). The biggest advantage of the field experiment in management accounting is the possibility to establish a causal relationship between management accounting practices and controls and the company's outcomes (Lourenço, 2019). In a nutshell, field experiments are conducted in the field (i.e., inside the companies) and have the same requirements as the prior experiments, such as randomized participants and the manipulation of independent variables.

There is an increasing interest in field experiments in management accounting due to its potent combination: the benefits of internal validity, consistent with the experimental method, and real world data that comes from the field (Bloomfield et al., 2016; Lourenço, 2019). This combination provides a great mix between "control and realism usually not achieved in the laboratory or with uncontrolled data" (Floyd & List, 2016, p. 438). Although this feature might increase the external validity of the field experiment, due its real applicability, it is still bounded to a less representative unit of observation (e.g., a business unit, a single company).

One important aspect of the field experiment is its difference in terms of construct validity. As mentioned in the prior section, field experiments are different than crowdsourced experiments as to the awareness of participants[4]. In the field experiment, the researchers will coordinate their interventions without the participants' knowledge in order to avoid any bias or noise to the results (e.g., selection bias) (Floyd & List, 2016). This design characteristic, however, harms more direct approaches to ensure construct validity, such as asking manipulation checks, since it requires more subtle ways to ask participants about the

---

[4] This is not a fundamental rule, but is generalized across literature, as seen on Bloomfield et al., (2016), Floyd and List (2016), and Lourenço (2019).

22º USP International Conference in Accounting
"Accounting in favor of sustainable development"

22º
USP International
Conference in Accounting

São Paulo 27 a 29 de julho 2022.

manipulations (Lourenço, 2019). Finally, field experiments also allow researchers to use pre-structured data to complement their design through measured variables (Asay et al., 2021). Researchers can assess different proprietary information to address a moderating variable, or to complement their study with the underlying explanations (i.e., mediating variable) of the causal effects on the investigated variable, or even for the dependent variable (e.g., company's financial results).

An example of a field experiment is Cronin, Erkens, Schloetzer, & Tinsley's (2021) study on the effects of controlling failure perceptions over performance. The authors manipulate the video-based message that sales workers see during their weekly meeting in one of the 20 Brazilian distributorships of a multinational direct sales organization. In the treatment condition, the sales workers see a video message from the regional head encouraging workers to look at failure as a "natural part of history". While, in the control condition, the sales workers see the same regional head summarizing the organization's history. The authors conducted the experiment during a four-week period, with a control and a treatment group, and with pre-structured data (e.g., weekly sales commission pre and post-treatment). Additionally, despite the fact that field experiments have more extrapolable results compared to lab and crowdsourcing experiments, the authors still disclosed both the limitations of the experimental method and the confounding factors that come from field studies.

4.4 Natural Experiment

A natural experiment is a naturally-occurring event that is not manipulated but can establish a "contrast between a treatment and a comparison condition" (Shadish et al., 2002, p. 17). Accounting researchers consider that an event that does not happen through nature's intervention (e.g., flood, hurricane) can also be treated as a natural experiment (e.g., adoption of a law or standard) (Mcvay, 2011). Therefore, natural experiments happen when an event occurs–either because of human intervention or natural intervention–and researchers are able to compare ex-ante to ex-post information. These events are also known as exogenous shocks (Bloomfield et al., 2016). One point of constant discussion regarding natural experiments is whether they are truly experiments since the "cause cannot usually be manipulated" (Shadish et al., 2002, p. 12). That is why natural experiments can also be classified as quasi-experiments (Aguinis & Bradley, 2014). There is substantial accounting research with natural experiments with a concurrent increase of sophisticated statistical tools to analyse the events (e.g., Diff-in-Diff, RDD, synthetic controls) (Lonati et al., 2018).

While natural experiments include a control group as a counterfactual, similar to other experimental designs, this research method tends to face more challenges with internal validity. The main reason is that the control and treatment groups are not designed by the researchers, but determined exogenously. This means that participants are not randomly assigned to experimental conditions and, as consequence, researcher cannot assume that the experimental groups are equivalent (Trochim et al., 2016). As such, researchers that conduct natural experiments have to find alternative methods to minimize threats to internal validity, in particular, by developing sophisticated statistical analysis (Lourenço, 2019).

Natural experiments can benefit from large available pre-structured data or proprietary archival data (to be discussed in detail below). It might even happen that this data is not yet structured to investigate the exogenous shock and researchers will have to structure it. In any case, however, the researcher has less flexibility in how to operationalize the dependent

10

variables relative to the other experimental designs and will have to rely on standard statistical procedures to enhance construct validity of the available proxies, such as examining their convergent validity. Also, since the very characteristic of natural experiments is the natural occurrence of the event, researchers cannot manipulate the independent variables and include appropriate procedures (e.g., manipulation checks) to deal with construct validity threats.

The extent to which results from natural experiments can be extrapolated depends on the unit of observation of a particular study. When the unit of observation is at the organizational level, that is, when the exogenous event has affected a large sample of organizations, the ability to extrapolate and generalize results is higher. However, when the unit of observation is at the subunit organizational or individual level, the challenge with external validity is higher, since researchers will have a hard time in convincing that the organization examined to collect the data is somehow similar to other organizations, located in other places.

An example of a natural experiment in management accounting is Flammer and Kacperczyk's (2016) study about the effect of stakeholder orientation on innovation on a company's business decisions. The authors explore the enactment of state-level constituency statutes, particularly the statutory change of the company's responsibility from shareholders to stakeholders. From 1980 to 2006, 34 US states have adopted constituency statutes that focus on stakeholder value creation. The authors use information about patent creation from the National Bureau of Economic Research (NBER) Patent Data Project database from 1976 to 2006 to assess their dependent variable, innovative productivity, measured as the number of patents and citations divided by the number of company's employees. In order to deal with internal validity issues, the authors collected a number of control variables that could act as confounding factors. This study is consistent with both characteristics of the natural experiment, in which the authors use pre-structured data—NBER database—that can capture exogenous shock and are consistent with the investigated variables, and the higher external validity that the large dataset provides.

4.5 Pre-structured archival study

Pre-structured archival studies represent a quantitative research method that uses archival data, that is, data recorded and structured by third-parties and whose primary purpose is not academic research (Bloomfield et al., 2016; Moers, 2006). The dataset used in pre-structured archival studies typically include observations for several respondents, mostly at the organizational level of analysis. Pre-structured archival studies are the primary research method used in the overall accounting literature (Bloomfield et al., 2016). For management accounting studies, the lack of available public data makes more challenging the use of pre-structured archival studies relative to alternative research methods, such as surveys, experiments, and case studies (Hesford et al., 2006; Moers, 2006; Aguiar, 2018).

Pre-structured archival studies are classified as a non-experimental study as they do not include a control group[5]. Due to the lack of a control group, this research method has difficulty in ruling out alternative explanations as an observed association between the variables of interest can be attributed to reverse causality, omitted correlated variables, or a miss-specified functional form (Gassen, 2014; Lourenço, 2019). Pre-structured archival studies face several

---

[5] Archival studies that include a control group are typically classified as natural experiments.

internal validity problems associated with selection bias and endogeneity that researchers try to solve by using econometric solutions (e.g., the use of instrumental variables) (Lourenço, 2019).

Pre-structured archival studies can take advantage of the richness of the available data that are suited for the investigation of several research questions of interest to the accounting community. However, the fact that pre-structured archival studies use dataset that are available prior to the beginning of the research creates contruct validity challenges as well. In pre-structured archival studies, the researcher has to define the operationalization of the variables of interest within the pre-existing dataset (Bloomfield et al., 2016). As such, the available pre-structured dataset may not always include data that is ammenable to the operationalization of theoretical constructs the research is willing to investigate. In other words, the researcher may not find suitable proxies for the variables of interest within the available pre-structured dataset.

Finally, given that pre-structured archival studies involve large samples (Das et al., 2016; Moers, 2006), their ability to make inferences from the observed sample to samples located in other places and at other time is relatively high. The unit of observation in pre-structured archival studies tends to be highly representative and then results from these studies may be extrapolable and generalizable to the population of interest. However, selection bias may pose challenges for the representativeness of the samples used in pre-structured archival studies, in particular, self-selection bias since the disclosure of (management) accounting information is not random (Moers, 2006; Lourenço, 2019).

Laviers, Sandvik, and Xu, (2021) is an example of a preestructured archival study as it does not include a control group and uses several large available dataset. In their study, Laviers et al. (2021) examine investor reactions to CEO pay ratio voluntary disclosures. They collect proxy statements from firms listed in the Standard & Poors 1500 index with mandated CEO pay ratio disclosure and classify firms as having low, middle, or high CEO pay ratio. The authors combine the CEO pay ratio information with information collected from several large database platforms, such as stock returns from CRSP, financial information from Compustat, and executive compensation data from Execucomp. Because they do not have a control group, Laviers et al. (2021) conduct several additional analysis and robustness tests using different empirical specifications to increase internal validity. As to the generalizability of results, the authors deal with self-selection by using an estimation procedure by which they include the inverse Mills ratio (Heckman, 1979).

4.6 Proprietary archival study

Similar to pre-structured archival studies, Proprietary archival studies is a quantitative research method that uses archival data. The key difference is that the dataset used in this case is proprietary, meaning that the data is confidential rather than public and can be accessed only if the proprietor or owner of the data grants the access (Moers, 2006). Examples of proprietary data includes third-party surveys (e.g., consulting firms) and firm internal data (Das et al., 2016; Moers, 2006). The dataset used in proprietary archival studies tends also to include observations for several respondents, but this time, not necessarily at the organizational level of analysis. In particular, firm internal data can include observations at the subunit organizational level as well as at the individual level of analysis. Because of the possibility of getting access to data at a level of analysis lower than the organizational level, proprietary archival studies are more suitable to address research questions that are of interest to management accounting researchers.

Proprietary archival studies are also classified as non-experimental as a control group is lacking. This research method faces the same internal validity challenges already mentioned due to the lack of a control group as they cannot easily rule out alternative explanations to the results. Proprietary archival studies then follow similar procedures as pre-structured archival studies in the use of econometric solutions to solve internal validity problems

Different from pre-structured archival studies, proprietary archival studies cannot only take advantage of the available data, but also gather additional unstructured data and structure it to create measures that are suitable for addressing the relevant research questions of interest to a particular study, similar to what Bloomfield et al. (2016) labeled as a hand-collected archival study. As such, the possibility of structure additional data not yet available in the current dataset gives more flexibility to the researcher to find suitable proxies for the variables of interest, reducing the challenges with construct validity. In addition, proprietary data can be combined with field interviews to help identify suitable proxies for the relevant variables. The main challenge, though, is to get access to proprietary data in the first place (Moers, 2006).

Finally, proprietary archival studies also involve large and comprehensive samples (Moers, 2006). Yet, their ability to make inferences from the observed sample to samples located in other places and at other time may be more challenging than for pre-structured archival studies, particularly in relation to firm internal data. Whether the proprietary data comes from third-party surveys or from firm internal data, proprietary archival studies face the same challenge associated with selection bias and, particularly, self-selection bias. When using firm internal data, the additional challenge the representativeness of the sample and, then, the ability to extrapolate and generalize results from the observed sample to other samples. The reason is that observations at the subunit or the individual (e.g., employee) level of analysis may be unique to the sampled firm characteristics, making difficulty to extrapolate results to other firms with different characteristics. Because of that, similar to laboratory experiments, proprietary archival studies can better argue in favour of a generalization to the theory be tested.

Ikäheimo, Kallunki, University, and Schiehll (2018) is an example of proprietary archival study without a control group and using firm internal data. Ikäheimo et al. (2018) examine the relationship between performance-based incentives for white-collar employees and firm future profitability and if this relationship depends on task complexity. They use a large proprietary compensation panel data set from a survey questionnaire administrated by the Confederation of the Finnish Industries. The data set includes over 564,000 individual employee-year and 7,820 firm-year observations over the years 2002–2011. The authors conduct robustness checks to deal with endogeneity issues and this way increase internal validity. They use different proxies to capture the same construct to deal with construct validity threats. Finally, given the comprehensive dataset Ikäheimo et al. (2018) use in their study, the ability to extrapolate results is high and can be generalized to other organizations that use incentive schemes and plan to change them, or do not currently have incentive schemes but plan to adopt them.

4.7 Large scale survey

Large scale survey consists in a primary data collection method that is operationalized through a questionnaire composed by research instruments which are submitted to a broad set of potential respondents (e.g., an e-mail database or social media). Large scale surveys are one of the main research methods used in management accounting research (e.g., Van der Stede et

13

22º USP International Conference in Accounting
"Accounting in favor of sustainable development"

22º
USP International
Conference in Accounting

São Paulo 27 a 29 de julho 2022.

al., 2006) which are usually implemented when data is not available in a pre-structured manner or when, although data may be available, the available data is not adequate for addressing a particular research problem. Then, researchers use large scale surveys to translate complex events or phenomena mostly into quantitative data through the adoption of research instruments (Spekle & Widener, 2018a). These data can express facts, opinions, or perceptions considering different levels of analysis such as individual, team and organizational levels. An illustration of the importance of this method in the management accounting field is a Special Forum that was published recently in the Journal of Management Accounting Research (Spekle & Widener, 2018b) and that treated common issues and future directions related to the implementation of surveys in the area.

Large scale survey studies are not designed with a control group as well as there is not a manipulation of independent variables. Both dependent and independent variables are elicited in the research instrument (Bloomfield et al., 2016). Large scale surveys are usually designed in only one period of time (i.e., cross-section) what limits its potential for causal claims (i.e., absence of a counterfactual or the absence of time difference between the cause and the effect). Even when they involve different periods (i.e., longitudinal design), large scale surveys suffer from a lack of information since respondents might not be available to participate in a second or third survey waves. Due to threats to internal validity, researchers conducting large scale survey adopt several methodological strategies, such as anchoring their studies in clear arguments (coherent arguments) derived from theories Van der Stede (2014), defining the theoretical population, target population and target respondents, and considering the different types of responses (facts, opinions, etc.). Hence, although some methodological strategies can be applied in large scale surveys, these remedies cannot completely overcome the threats to internal validity.

Large scale surveys have to look at on how well the instruments/questions in the study operationalize the definition and frontiers of a construct (i.e., in the theoretical level) (Bedford & Spekle, 2018a). As a consequence, researchers conducting large scale surveys are faced with several design decisions aiming at increasing construct validity: the choice of the instruments used to capture data, the validation of the selected instruments, and respondents' knowledge to what information is being asked. In particular, this research method demands careful attention to construct dimensionality, measurement types for the constructs (formative and reflexive), and the use of single and multiple items to measure a construct (Bedford & Spekle, 2018a). Overall, construct validity is a critical issue in large scale surveys since it involves a broad set of respondents and henceforth researchers are not capable of understanding and treating biases that influence individual's responses (e.g., hallo effect, social desirability, lack of knowledge). An additional problem is that, by not contemplating contextual information (i.e., organizational particularities), large scale surveys face problems related to definition of the most appropriate respondent and the suitability of the research instruments.

Finally, large scale surveys benefit from the use of large samples and are thus able to provide evidence that can be extrapolated and generalized to a population of firms although the most reliable arguments around surveys consider the generalization for theory. In large scale surveys, the researcher is usually aware that the potential of extrapolation of the results depends on respondents and nonresponse bias as well as the response rates obtained (Hiebl & Richter, 2018; Spekle & Widener, 2018a). In addition, the ability of large scale surveys to extrapolate

14

results depends on whether probability or nonprobability sampling is used, with the latter being the most common strategy.

Bedford, Spekle, and Widener (2022) is an example of a large scale survey with Business Unit (BU) managers from the Netherlands, obtaining a final sample of 83 respondents. The authors developed a cross-sectional survey design using an online questionnaire, addressed to a population of 172 BU managers. Their selection criteria involve that target respondents are BU managers in for-profit organizations, have either profit or investment centre accountability and report to a higher hierarchical level and that BUs employ at least 15 full time employees and serve external customers. Although the sample size is considered small, they obtain a high response rate of 48.3%. The authors follow several procedures to mitigate internal, construct, and external validity threats. They address nonresponse bias and common method bias, in addition to use validated instruments and control variables in the research model. Bedford, Speklé & Widener (2022) study how do firms change budget tightness in response to a global crisis and the implications of budget tightness for employee stress and emotional exhaustion, also considering an enabling budget design as a moderation for this relationship.

4.8 Single entity survey

Single entity survey[6] consists in a primary data collection method that is operationalized through a questionnaire composed by research instruments which are submitted to respondents from one organizational setting. Because of that, studies that employ this research method usually focus on a research problem at an individual or team level of analyses and also on more subjective constructs (i.e., perception, feelings and opinion). Although context is important for surveys in general, for single entity surveys the organizational context is a pillar for the definition of both the theoretical model and methodological strategies. In other words, qualitative information of the organizational context has a central role in shaping and "calibrating" the research problem and the research instrument, and this affect the validity issues faced by single entity surveys. This research method requires researchers to map and diagnose phenomena and relationships within a specific context, which is usually carried out through interviews, prior to the administration of the survey. Single entity survey terminology is not commonly presented in methodological books (i.e., (Saunders et al., 2019; Smith, 2019) and has been applied in the management accounting research mostly as an alternative and complementary research method in multi(mixed)-method research (e.g., Demeré, Krishnan, Sedatole, & Woods, 2016; Wouters & Wilderom, 2008), notwithstanding in Brazil it has been used as a mono-method strategy (e.g., Mucci, Frezatti, & Bido, 2021; Souza & Beuren, 2018).

Similar to large scale surveys, single entity surveys are not designed with a control group and independent variables are measured rather than manipulated. Single entity surveys can be administered using cross-sectional or longitudinal data. Since researchers have access to the field, they might not suffer from the lack of information when adopting a longitudinal design. This research method faces the same internal validity threats as large scale surveys and thus researchers can use mostly similar procedures to mitigate these threats. However, single entity

---

[6] Single entity surveys cannot be confused with field studies, in which field information obtained through interviews, observations and questionnaires integrate the findings (through the triangulation procedure). In particular, in single entity surveys, despite the importance of qualitative information, the information obtained in the questionnaire is predominant and is used to test a quantitative model.

surveys might suffer less from internal validity as researchers can strictly define the theoretical and empirical model in light of the context being investigated (Luft & Shields, 2002) as well as control for confounding effects that emerge from the context. Also, single entity surveys can better limit the target respondents and even use random samples as well as design a longitudinal study, which might benefit the internal validity (i.e., causal claims) of a study.

In single entity surveys researchers are able to calibrate the research instruments to the organizational setting, which involves for instance, the adaptation of construct definition and research instruments, as well as a more adequate choice of which employee-title is a candidate for an ideal survey respondent. Hence, by understanding the context before implementing the survey, researchers are able to obtain more qualified and less biased responses by identifying the knowledgeable potential respondents and by engaging them to appropriately respond to the survey. In order to mitigate bias in single entity surveys, researchers should be sensible about the confidentiality terms, the invitation letter and the survey administration tools used. All these procedures contribute to mitigate construct validity concerns.

Single entity surveys face similar challenges as proprietary archival studies in terms of making inferences to other organizations since the research model is context-dependent. This does not mean that the findings are not extrapolable to other organizational settings, but they are only empirically applicable to organizations sharing similar characteristics and phenomena. Hence, when researchers discuss about external validity of single entity surveys, they are usually referring to the extrapolation of the population of individuals within that organizational setting in a certain period of time, for instance current middle-managers or assembly workers. Single entity surveys benefit from the close contact with the organizations to increase external validity. Researchers usually obtain larger response rates than in large scale surveys (Hiebl & Richter, 2018) and are more capable of addressing the issues related to nonresponse bias.

Mucci, Frezatti and Bido (2021) is an example of a single entity survey developed with a sample of 75 middle managers from different areas in an organization that operates in the electric utilities industry. These authors develop a cross-sectional survey design using an online questionnaire which is operationalized with the support of the budgeting manager of the firm. The authors obtained a high response rate of 42%. The authors follow several procedures to mitigate construct (i.e., pre-survey interviews, adaptation of the instruments, pilot test with the target respondents of the organization) and internal validity threats (i.e., strict theoretical model, use control variables, test for common-method bias).

5 Conclusion

This study proposes a typology of quantitative research in management accounting based on three key design features: The presence of a control group, the use of pre-structured data, and the ability to extrapolate results. Building on these three criteria, we propose eight different quantitative research methods that management accounting researchers can use when addressing their research questions, including experimental (lab experiments, crowdsourcing experiments, field experiments, and natural experiments) and non-experimental alternatives (pre-structured archival studies, proprietary archival studies, large sample surveys, and single entity surveys). Secondarily, we discuss the main implications of the proposed typology for research design in terms of the validity framework, focusing on internal validity, construct validity, and external validity.

The proposed typology and its implications for internal validity can help management accounting researchers broaden the scope of research questions to be investigated. For instance, if the researcher is interested to examine the role of a value statement on employees' behaviour, the researcher can use alternative research methods such as designing a lab experimental study and manipulate the value statement to capture employees' responses; accessing an organization to use available data on employees' understanding of the organization's value statement and on their performance relative to different metrics; or collect survey responses from employees located in different organizations about their perception of the organization's value statement and attitudinal and/or behavioural responses, such as organizational citizenship behaviour. The key point is that researchers will be better off by choosing a quantitative research method that is viable to be conducted given the existing constraints (e.g., data availability, access to the organization, time and money).

The proposed typology has the potential to help management accounting researchers identify the alternative quantitative research methods available to address a particular research question. In particular, a group of researchers interested in a similar research question can examine it sequentially using different quantitative research methods, with the evident benefit of replicating, extrapolating, and generalizing results. For instance, the behavioural effects of using tight budgetary controls identified in a large scale survey can be further examined in a clever experimental design in order to provide stronger evidence of causal association. Also, the proposed typology can help management accounting researchers examine a given research question that are not suitable to be addressed through conventional research methods (e.g., large sample surveys) by using one of the alternative and available quantitative research methods. For instance, if the relevant question is to examine COVID-19 effects on employees' use of accounting information for decision making, the researcher could have access to an organization and collect proprietary data about the frequency of use of accounting information prior and post the beginning of the pandemics in a natural experimental design and then examine whether this use has increased or not.

In any case, researchers conducting quantitative management accounting studies have to pay attention to the design choices in a specific study. First of all, the choice of the research method is dependent on the research question of interest, not the other way around. While the exploration of a different research method can be valuable in terms of acquiring new research skills, the research question should come in the first place to guide this decision. The benefits in terms of increased chances of publication are higher when the researcher obtains deepen knowledge on how to use a particular research method to address the relevant research questions of interest. The reason is that each research method involves a different set of procedures or protocol to be followed in order to deal with the validity threats. Then, regardless of the quantitative research method selected by the management accounting researcher to address the relevant research question, the researcher is expected to apply the associated protocol, according to the best practices stablished in the area, when conducting the study.

Second, as we try to emphasize in this study, the choice of a particular quantitative research method involves an associated choice of the validity threats to which the researcher will be more exposed to. For instance, the choice of a lab experimental design increases the internal validity, while creates challenges for the external validity of the study. Conversely, the choice of a pre-structured archival study increases the external validity, but poses increased internal validity threats. Again, following the research protocol for a given quantitative research

17

method, accepted in the research community of interest, increases the chances of the researcher being able to deal with the different validity threats. The use of the predictive validity framework, also known as Libby Boxes, can be a helpful tool for the researcher better visualize the research design for her study as well as to identify potential validity threats.

Finally, management accounting researchers are becoming increasingly creative in how to collect and use data to address a particular research question. This creativity can be evident in mono-method research designs as well as in the use of multi-method research designs. For mono-method research designs, management accounting researchers are increasingly taking advantage of the internal validity associated with experimental studies and the advantages of the external validity associated with observational studies by conducting quasi-experimental studies, using proprietary archival data and design choices typical of experimental studies such as pre- and post-measures for the relevant outcomes (e.g., Brüggen, Grabner, and Sedatole, 2021; Forker, Grabner, and Sedatole, 2020). In terms of multi-method research designs, it is not new that management accounting researchers are combining the use of different data collection procedures in the same study to address the relevant research question (e.g., Bol, Braga de Aguiar, and Lill, 2020; Wouters and Wilderom, 2008). The main benefit of combining different research methods is that the researcher can increase the sources of relevant data and can provide stronger results by using alternative research methods that complement each other, such as the use of proprietary archival data on employees performance combined with perceptual measures on employees' motivation captured through a single entity survey.

References

Aguiar, A. B. (2017). Pesquisa Experimental Em Contabilidade: Propósito, Desenho E Execução. *Advances in Scientific and Applied Accounting*, *10*(2), 224–244. https://doi.org/10.14392/asaa.2017100206

Aguiar, A. B. de. (2018). O pequeno mundo da pesquisa em contabilidade gerencial no Brasil: discussão sobre desenhos alternativos de pesquisa. *Revista de Contabilidade e Organizações*, *12*, e151933. https://doi.org/10.11606/issn.1982-6486.rco.2018.151933

Aguinis, H., & Bradley, K. J. (2014). Best Practice Recommendations for Designing and Implementing Experimental Vignette Methodology Studies. *Organizational Research Methods*, *17*(4), 351–371. https://doi.org/10.1177/1094428114547952

Aguinis, H., & Ramani, R. S. (2021). MTurk Research : Review and Recommendations. *Journal of Management*, *47*(4), 823–837. https://doi.org/10.1177/0149206320969787

Asay, H. S., Guggenmos, R. D., Kadous, K., Koonce, L., & Libby, R. (2021). Theory Testing and Process Evidence in Accounting Experiments. *The Accounting Review*.

Bedford, D. S., & Spekle, R. F. (2018). Construct Validity in Survey-Based Management Accounting and Control Research. *Journal of Management Accounting Research*, *30*(2), 23–58. https://doi.org/10.2308/jmar-51995

Bedford, D. S., Spekle, R. F., & Widener, S. K. (2022). Accounting , Organizations and Society Budgeting and employee stress in times of crisis : Evidence from the Covid-19 pandemic. *Accounting, Organizations and Society*, (xxxx). https://doi.org/10.1016/j.aos.2022.101346

Bentley, J. W. (2021). Improving the Statistical Power and Reliability of Research Using Amazon Mechanical Turk. *Accounting Horizons*, *35*(4), 45–62.

https://doi.org/10.2308/HORIZONS-18-052

Bloomfield, R., Nelson, M. W., & Soltes, E. (2016). Gathering Data for Archival, Field, Survey, and Experimental Accounting Research. *Journal of Accountig Research*, *54*(2), 341–395. https://doi.org/10.1111/1475-679X.12104

Bol, J. C., Braga de Aguiar, A., & Lill, J. B. (2020). *Peer-Level Calibration of Performance Evaluation Ratings : Are There Winners or Losers ?*

Brüggen, A., Grabner, I., & Sedatole, K. L. (2021). The Folly of Forecasting: The Effects of a Disaggregated Demand Forecasting System on Forecast Error, Forecast Positive Bias, and Inventory Levels. *The Accounting Review*, *96*(2), 127–152. https://doi.org/10.2308/tar-2018-0559

Burrell, G., & Morgan, G. (1979). Sociological Paradigms and Organisational Analysis. *Sociological Paradigms and Organisational Analysis*, *41*(1), 448. https://doi.org/10.1177/003803858001400219

Cronin, M., Erkens, D. H., Schloetzer, J. D., & Tinsley, C. H. (2021). How controlling failure perceptions affects performance: Evidence from a field experiment. *Accounting Review*, *96*(2), 205–230. https://doi.org/10.2308/TAR-2018-0146

Crotty, M. (1998). *The Foundations of Social Research: Meaning and Perspective in the Research Process*. SAGE.

Das, R., Jain, K. K., & Mishra, S. K. (2016). Archival Research: A Neglected Method in Organization Studies. *Benchmarking: An International Journal*. https://doi.org/https://doi.org/10.1108/BIJ-08-2016-0123

Demeré, B. W., Krishnan, R., Sedatole, K. L., & Woods, A. (2016). Do the incentive effects of relative performance measurement vary with the ex ante probability of promotion? *Management Accounting Research*, *30*, 18–31. https://doi.org/10.1016/j.mar.2015.09.002

Dennis, S. A., Goodson, B. M., & Pearson, C. A. (2020). Online Worker Fraud and Evolving Threats to the Integrity of MTurk Data: A Discussion of Virtual Private Servers and the Limitations of IP-Based Screening Procedures. *Behavioral Research in Accounting*, *32*(1), 119–134. https://doi.org/10.2308/bria-18-044

Flammer, C., & Kacperczyk, A. (2016). The impact of stakeholder orientation on innovation: Evidence from a natural experiment. *Management Science*, *62*(7), 1982–2001. https://doi.org/10.1287/mnsc.2015.2229

Floyd, E., & List, J. A. (2016). Using Field Experiments in Accounting and Finance. *Journal of Accounting Research*, *54*(2), 437–475. https://doi.org/10.1111/1475-679X.12113

Forker, E., Grabner, I., & Sedatole, K. (2020). *Does learning by disaggregating accelerate learning by doing? The effect of forecast disaggregation on the rate of improvement in demand forecast accuracy*.

Gassen, J. (2014). Causal inference in empirical archival financial accounting research. *Accounting, Organizations and Society*, *39*(7), 535–544. https://doi.org/10.1016/j.aos.2013.10.004

Haesebrouck, K. (2021). The Effects of Information Acquisition Effort, Psychological Ownership, and Reporting Context on Opportunistic Managerial Reporting*. *Contemporary Accounting Research*, *38*(4), 3085–3112. https://doi.org/10.1111/1911-3846.12712

Heckman, J. J. (1979). Sample Selection Bias as a Specification Error. *Econometrica*, *47*(1), 153–161.

Hesford, J. W., Lee, S. H. (Sam), Van der Stede, W. A., & Young, S. M. (2006). Management

Accounting: A Bibliographic Study. *Handbooks of Management Accounting Research*, *1*, 3–26. https://doi.org/10.1016/S1751-3243(06)01001-7

Hiebl, M. R. W., & Richter, J. F. (2018). Response Rates in Management Accounting Survey Research. *Journal of Management Accounting Research*, *30*(2), 59–79. https://doi.org/10.2308/jmar-52073

Houghton, K. A., & Hronsky, J. J. F. (1993). The Sharing of Meaning Between Accounting Students and Members of the Accounting Profession. *Accounting & Finance*, *33*(2), 131–147. https://doi.org/10.1111/j.1467-629X.1993.tb00324.x

Ikäheimo, S., Kallunki, J.-P., University, S. M., & Schiehll, E. (2018). Do White-Collar Employee Incentives Improve Firm. *Journal of Management Accounting Research*, *30*(3), 95–115. https://doi.org/10.2308/jmar-51902

Kinney, W. R. (2019). The Kinney Three Paragraphs (and More) for Accounting Ph.D. Students. *Accounting Horizons*, *33*(4), 1–14. https://doi.org/10.2308/acch-52451

Laviers, L., Sandvik, J., & Xu, D. (2021). *CEO Pay Ratio Voluntary Disclosures and Investor Reactions*.

Liyanarachchi, G. A. (2007). Feasibility of using student subjects in accounting experiments: a review. *Pacific Accounting Review*, *19*(1), 47–67. https://doi.org/10.1108/01140580710754647

Lonati, S., Quiroga, B. F., Zehnder, C., & Antonakis, J. (2018). On doing relevant and rigorous experiments: Review and recommendations. *Journal of Operations Management*, *64*(April), 19–40. https://doi.org/10.1016/j.jom.2018.10.003

Lourenço, S. M. (2019). Field Experiments in Managerial Accounting Research. *Foundations and Trends® in Accounting*, *14*(1), 1–72. https://doi.org/10.1561/1400000059

Luft, J., & Shields, M. (2002). Zimmerman's Contentious Conjectures: Describing the Present and Prescribing the Future of Empirical Management Accounting Research. *European Accounting Review*, *11*(4), 795–803. https://doi.org/10.1080/0963818022000047091

Mcvay, S. E. (2011). Discussion of Do Control Effectiveness Disclosures Require SOX 404(b) Internal Control Audits? A Natural Experiment with Small U.S. Public Companies. *Journal of Accounting Research*, *49*(2), 449–456. https://doi.org/10.1111/j.1475-679X.2011.00403.x

Moers, F. (2006). Doing Archival Research in Management Accounting. *Handbooks of Management Accounting Research*, *1*, 399–413. https://doi.org/10.1016/S1751-3243(06)01016-9

Mortensen, T., Fisher, R., & Wines, G. (2012). Students as surrogates for practicing accountants: Further evidence. *Accounting Forum*, *36*(4), 251–265. https://doi.org/10.1016/j.accfor.2012.06.003

Mucci, D. M., Frezatti, F., & Bido, D. de S. (2021). Enabling design characteristics and budget usefulness. *RAUSP Management Journal*, *56*, 38–54. https://doi.org/10.1108/RAUSP-04-2019-0058

Murphy, P. R., Wynes, M., Hahn, T.-A., & Devine, P. G. (2019). Why are People Honest? Internal and External Motivations to Report Honestly. *Contemporary Accounting Research*, *53*(9), 1689–1699. https://doi.org/10.1017/CBO9781107415324.004

Nascimento, A. R. do, Junqueira, E., & Martins, G. de A. (2010). Pesquisa Acadêmica em Contabilidade Gerencial no Brasil: Análise e Reflexões sobre Teorias, Metododologias e Paradigmas. *Revista de Administração Contemporânea*, *14*(6), 1113–1133.

Oehlert, G. W. (2003). A First Course in Design and Analysis of Experiments. In *The American*

Statistician (Vol. 57). https://doi.org/10.1198/tas.2003.s210

Palan, S., & Schitter, C. (2018). Prolific.ac—A subject pool for online experiments. *Journal of Behavioral and Experimental Finance*, *17*, 22–27. https://doi.org/10.1016/j.jbef.2017.12.004

Peer, E., Brandimarte, L., Samat, S., & Acquisti, A. (2017). Beyond the Turk : Alternative platforms for crowdsourcing behavioral research. *Journal of Experimental Social Psychology*, *70*, 153–163. https://doi.org/10.1016/j.jesp.2017.01.006

Rose, J. M. (2018). Manipulation and attention checks. In T. Libby & L. Thorne (Eds.), *The Routledge Companion to Behavioural Accounting Research* (1st ed., pp. 1–541). https://doi.org/10.4324/9781315710129

Saunders, M. N. K., Lewis, P., & Thornhill, A. (2019). Research Methods for Business Students. In *Pearson* (8th ed., Vol. 3). https://doi.org/10.1108/qmr.2000.3.4.215.2

Shadish, W. R. ., Cook, T., & Campbell, D. (2002). Experimental and quasi-experimental designs for generalized causal inference. In *Houghton Mifflin Company*. https://doi.org/10.1016/j.evalprogplan.2004.01.006

Smith, M. (2019). *Research Methods in Accounting* (5th ed.). SAGE.

Souza, G. E. De, & Beuren, I. M. (2018). Reflexos do sistema de mensuração de desempenho habilitante na performance de tarefas e satisfação no trabalho. *Revista de Contabilidade e Finanças*, *29*(77), 194–212. https://doi.org/10.1590/1808-057x201805850

Spekle, R. F., & Widener, S. K. (2018a). Challenging Issues in Survey Research: Discussion and Suggestions. *Journal of Management Accounting Research*, *30*(2), 3–21. https://doi.org/10.2308/jmar-51860

Spekle, R. F., & Widener, S. K. (2018b). Special Forum on Survey Research. *Journal of Management Accounting Research*, *30*(2), 1–2. https://doi.org/10.2308/jmar-10584

Swieringa, R. J., & Weick, K. E. (1982). An Assessment of Laboratory Experiments in Accounting. *Journal of Accounting Research*, *20*, 56–101.

Trochim, W. M., Donnelly, J. P., & Arora, K. (2016). *Research Methods - The essential knowledge base* (2nd ed.). Cengage Learning.

Trottier, K., & Gordon, I. M. (2018). Students as surrogates for managers: Evidence from a replicated experiment. *Canadian Journal of Administrative Sciences*, *35*(1), 146–161. https://doi.org/10.1002/cjas.1377

Van der Stede, W. A., Young, S. M., & Chen, C. X. (2006). Doing Management Accounting Survey Research. *Handbooks of Management Accounting Research*, *1*, 445–478. https://doi.org/10.1016/S1751-3243(06)01018-2

Wouters, M., & Wilderom, C. (2008). Developing performance-measurement systems as enabling formalization: A longitudinal field study of a logistics department. *Accounting, Organizations and Society*, *33*(4–5), 488–516. https://doi.org/10.1016/j.aos.2007.05.002