



## Replication in experimental management accounting research - an editorial

Alexander Brügger<sup>1</sup> · Victor S. Maas<sup>2</sup> · Alexandra Van den Abbeele<sup>3</sup>

Accepted: 4 June 2021 / Published online: 1 July 2021

© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2021

Over the past two decades, experimental management accounting research has increased in volume and prominence, especially in Europe where it remained something of a niche area well into the early 2000s. As experimental management accounting research became part of the mainstream literature, experimental researchers also faced challenges. Small samples and flexibility in data collection methods (Simmons et al., 2011) casted doubt on the conclusions that could be drawn from early experiments, large collaborative initiatives failed to replicate core findings in fields such as social psychology and behavioral economics (Camerer et al., 2016; Open Science Collaboration, 2012), and experimental studies of prominent researchers such as Hunton (Stone, 2015) and Stapel (Enserink, 2012) turned out to be fraudulent.

In light of these developments, calls for replication studies arose and it is not surprising that replication is often seen as a method to ‘clean up’ the literature by detecting and correcting false conclusions from earlier studies. However, we believe that this is a rather narrow and unfortunate perspective. While we agree that the susceptibility of experimental research to researcher bias and questionable research practices (John et al., 2012) is a serious cause for concern, and that more can and needs to be done by management accounting researchers to address this, we do not think that the main objective of replication studies is unmasking overconfident, opportunistic, or outright fraudulent researchers. Instead, replication serves several different purposes and is a core element of how a research field progresses. Specifically, following Fabrigar et al. (2020), we suggest that replications should be

---

✉ Victor S. Maas  
v.s.maas@uva.nl

Alexander Brügger  
a.brugger@maastrichtuniversity.nl

Alexandra Van den Abbeele  
alexandra.vandenabeele@kuleuven.be

<sup>1</sup> Maastricht University, Maastricht, The Netherlands

<sup>2</sup> University of Amsterdam, Amsterdam, The Netherlands

<sup>3</sup> KU Leuven, Leuven, Belgium

considered within the classic framework of four forms of validity: statistical conclusion validity, internal validity, construct validity, and external validity (Cook & Campbell, 1979; Shadish et al., 2002).<sup>1</sup>

Statistical conclusion validity refers to the validity of conclusions about the association between two constructs in the population. Researchers can make two types of errors. They can incorrectly conclude that an association exists in the population because it exists in their data (Type I error), and they can incorrectly conclude that an association does not exist in the population because it does not exist in their data (Type II error). The narrow view of replication is that its main purpose is identifying Type I errors by replicating existing studies as closely as possible but using larger samples (i.e., with higher statistical power). A failure to replicate is then often seen as an indication that the original finding was a ‘false positive’ (Fabrigar et al., 2020).

Psychologists disagree about the merits of such ‘direct’ replications (e.g., Simons, 2014; Stroebe & Strack, 2014; Zwaan et al., 2018) and it is not our intention to summarize the various viewpoints here.<sup>2</sup> Instead, we would like to highlight that replication is not only important from a statistical conclusion validity perspective, but also for reasons of internal validity, construct validity and external validity. Such non-direct replications are typically labelled ‘conceptual replications’ and introduce changes to the original procedures that might affect the experiment’s outcomes (e.g., Zwaan et al., 2018).

Internal validity refers to the validity that an observed association is due to a specific mechanism, typically a causal mechanism through which variation in an independent variable causes variation in a dependent variable. Compared to other research methods such as archival data analysis and survey research, experiments have relatively high internal validity. This comparative advantage provides important research opportunities. Specifically, we would welcome more experimental research that tests hypotheses that have previously been tested with cross-sectional data. Of course, this form of replication is not new, the benefits of triangulation have been voiced frequently and convincingly (e.g., Hageman, 2008; Modell, 2009). Yet, we feel that there is much untapped potential for experimental researchers here. In addition, as more and more experiments are performed online, often using participants recruited on platforms such as Amazon Mechanical Turk and Prolific, attrition rates go up—posing a potential threat to internal validity (Fabrigar et al., 2020). Thus, we believe that there is a need for experiments that replicate online studies in controlled laboratory settings.

Construct validity refers to the extent to which the operationalization of a variable in a study (e.g., the manipulation of an independent variable or the measure of a dependent variable) corresponds to the conceptual construct that it is supposed to capture. Since no operationalization is perfect, replication studies are needed to

<sup>1</sup> While this framework originates in the work of Cook and Campbell (1979), experimental accounting researchers might be more familiar with the specific graphical representation of Libby (1981): the so-called “Libby-boxes”.

<sup>2</sup> We kindly refer interested readers to the paper of Zwaan et al. (2018), including the comments on this paper in the same issue of the journal and the response of the authors to these comments.

ensure that conclusions about cause-effect relationships between constructs are not conditional upon the operationalization of these constructs. Thus, we encourage experimentalists to test hypotheses that have been tested in previous experiments using alternative manipulations and different measures. Notably, if such a replication leads to a different result (no effect whereas in the original study there was an effect, or vice versa) this could point towards a potential moderating effect and thus facilitate the development of more comprehensive and accurate theoretical models.

Finally, the external validity of a study refers to the extent to which its conclusions can be generalized to other settings and populations. The external validity of findings from controlled lab experiments is a concern in fundamental disciplines such as economics and psychology, but even more so in applied fields such as accounting. The reason is that many of the constructs that accounting researchers are interested in are derived from practice and difficult to isolate from their specific context. Clearly, this is another area where replications can play an important role. We encourage researchers to test hypotheses that have previously been tested in other times, places, cultures, or professional environments, for example. This would allow us to update our beliefs about the existence of cause-effect relationships, and the specific conditions that determine this existence. In addition, replication studies can also add new variables that have not been studied before. As such, research can examine the robustness of prior results and at the same time extend our knowledge.

In summary, replication is of crucial importance to move the literature forward and improve our theoretical models. Failure to replicate findings of an earlier study is not necessarily evidence of questionable research practices, but more likely just a signal that under certain circumstances results can be different. For example, relationships between certain variables of interest are more complex than initially thought or relationships have changed over time. In other words, ‘unsuccessful’ replication should not be considered a threat to existing studies, but an important extension of our knowledge.

This special issue contains five experimental papers that can all be classified as replication studies. Their diversity underscores our main point that replication is much more than checking whether previous studies might have produced false positives.

First, the study by *Lindquist and Rausch* is an example of a direct replication. These authors exactly replicate an experiment reported in Lindquist (1995). Previous studies have also replicated (parts of) that experiment and found different results from the original study. However, these studies sometimes used different experimental tasks and different operationalizations. The result of Lindquist and Rausch’s exact replication are largely consistent with those of the original study. This paper raises important questions about construct validity and external validity.

Next, *Hermans, Cools, and Van den Abbeele* is a conceptual replication of Bol et al. (2016). At a high level of abstraction, Hermans et al. are interested in the same causal relationships as Bol et al. (2016), however they use different operationalizations and a slightly different setting to examine these relationships. The results are similar to those of the original study. Importantly, Hermans et al. also use their experiment to dig deeper into the underlying process, using data collected in a

post-experimental questionnaire. This analysis not only helps us to better understand their own results, but also sheds new light on the original study.

*Knox* partly replicates an experiment of Cheng and Humphreys (2012) with the intention of examining a potential moderating effect. The original study suggests that managers' judgments about their firm's strategy are affected by whether or not this strategy is presented in a format with clearly specified links between leading and lagging performance measures. Knox hypothesizes that this effect is conditional upon participants perceiving these links as being 'causally related.' For his replication he uses a new instrument that 'removes some richness from the context developed by Cheng and Humphreys (2012)' but is otherwise explicitly designed to be comparable to the instrument used in the original study. As the author describes, this study is an explicit attempt to investigate the generalizability of one of the conclusions of Cheng and Humphreys (2012). Since he finds support for his moderation hypothesis, the results of this replication point towards a possible boundary condition for this generalizability.

*Nikias, Schwartz, and Young* also replicate a prior experiment and look at a potential moderator. These authors start out from the same experimental design as Hannan et al. (2010) but then add a condition. Specifically, in the original experiment two subordinates are not aware of each other's' project cost when they make a budget request to a superior. Nikias et al. replicate this 'low transparency' condition but also add a 'high transparency' condition in which the subordinates do know their colleague's project cost. Consistent with their expectations, they find that transparency matters: subordinates who observe their colleague's costs request higher budgets. Thus, like Knox, Nikias et al. suggests that the generalizability of a finding from prior research may be conditional upon a specific characteristic of the setting.

Finally, while the paper of *van Veen-Dirks, Leliveld, and Kaufmann* does not replicate a specific existing study, we nevertheless included in the special issue because it illustrates two important points. First, van Veen Dirks et al. show how experiments can be used to reexamine relationships that have been studied using other research methods. The authors designed an experiment to investigate the effects of the use of enabling performance measurement systems. While the consequences of using performance measurement systems in an enabling (as opposed to coercive) way have been studied extensively in the management accounting literature, most of this research has used qualitative case studies or questionnaire surveys (e.g., Ahrens & Chapman, 2004; Hartmann & Maas, 2011; Wouters & Wilderom, 2008). The paper of van Veen-Dirks et al. contributes to the literature by complementing these richer field studies with an experiment that primarily stands out for its internal validity. Second, van Veen-Dirks run their experiment with two groups of participants: students in a research lab at a Netherlands-based university and MTurk participants located in the United States. It can thus be said that these authors replicated themselves, increasing both the statistical conclusion validity and the external validity of their study.

We hope that the readers of this special issue will appreciate the different ways in which replications contribute to our knowledge and will agree with us and others that it is time to make replication 'mainstream' (Zwaan et al., 2018). A shift away from a focus on developing exciting new ideas towards validation and refinement of

existing ideas does not signal stagnation, it signals that we are taking ourselves seriously and become mature as an academic discipline.

We would like to thank the authors for submitting their papers for consideration for publication in this special issue, and the anonymous reviewers for their effort and constructive feedback to the authors. We would also like to thank the Editors of the Journal of Management Control for initiating the special issue and for inviting us as Guest Editors. Special thanks to the Managing Editor Thomas Günther for his guidance, patience, and for taking on most of the actual work in preparing this special issue.

## References

- Ahrens, T., & Chapman, C. S. (2004). Accounting for flexibility and efficiency: A field study of management control systems in a restaurant chain. *Contemporary Accounting Research*, 21(2), 271–301.
- Bol, J. C., Kramer, S., & Maas, V. S. (2016). How control system design affects performance evaluation compression: The role of information accuracy and outcome transparency. *Accounting, Organizations and Society*, 51, 64–73.
- Camerer, C. F., Dreber, A., Forsell, E., Ho, T. H., Huber, J., Johannesson, M., Kirchler, M., Almenberg, J., Altmeld, A., Chan, T., & Heikensten, E. (2016). Evaluating replicability of laboratory experiments in economics. *Science*, 351(6280), 1433–1436.
- Cheng, M. M., & Humphreys, K. A. (2012). The differential improvement effects of the strategy map and scorecard perspectives on managers' strategic judgments. *The Accounting Review*, 87(3), 899–924.
- Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design and Analysis Issues for Field Settings*. Houghton Mifflin.
- Enserink, M. (2012). Final report on Stapel also blames field as a whole. *Science*, 338, 1270–1271.
- Fabrigar, L. R., Wegener, D. T., & Petty, R. E. (2020). A validity-based framework for understanding replication in psychology. *Personality and Social Psychology Review*, 24(4), 316–344.
- Hageman, A. M. (2008). A review of the strengths and weaknesses of archival, behavioral, and qualitative research methods: Recognizing the potential benefits of triangulation. *Advances in Accounting Behavioral Research*, 11, 1–30.
- Hannan, R. L., Rankin, F. W., & Towry, K. L. (2010). Flattening the organization: The effect of organizational reporting structure on budgeting effectiveness. *Review of Accounting Studies*, 15(3), 503–536.
- John, L. K., Loewenstein, G., & Prelec, D. (2012). Measuring the prevalence of questionable research practices with incentives for truth telling. *Psychological Science*, 23(5), 524–532.
- Libby, R. (1981). *Accounting and human information processing: Theory and applications*. Prentice Hall.
- Lindquist, T. M. (1995). Fairness as an antecedent to participative budgeting: Examining the effects of distributive justice, procedural justice and referent cognitions on satisfaction and performance. *Journal of Management Accounting Research*, 7, 122–147.
- Hartmann, F. G., & Maas, V. S. (2011). The effects of uncertainty on the roles of controllers and budgets: An exploratory study. *Accounting and Business Research*, 41(5), 439–458.
- Modell, S. (2009). In defence of triangulation: A critical realist approach to mixed methods research in management accounting. *Management Accounting Research*, 20(3), 208–221.
- Open Science Collaboration. (2012). An open, large-scale, collaborative effort to estimate the reproducibility of psychological science. *Perspectives on Psychological Science*, 7(6), 657–660.
- Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Houghton Mifflin.
- Simons, D. J. (2014). The value of direct replication. *Perspectives on Psychological Science*, 9(1), 76–80.
- Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, 22(11), 1359–1366.

- Stone, D. N. (2015). Post-Hunton: Reclaiming our integrity and literature. *Journal of Information Systems*, 29(2), 211–227.
- Stroebe, W., & Strack, F. (2014). The alleged crisis and the illusion of exact replication. *Perspectives on Psychological Science*, 9(1), 59–71.
- 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.
- Zwaan, R. A., Etz, A., Lucas, R. E., & Donnellan, M. B. (2018). Making replication mainstream. *Behavioral and Brain Sciences*. <https://doi.org/10.1017/S0140525X17001972>

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.