Taking Stock of Research on the Levers of Control with Meta-Analytic Methods

Bellora-Bienengräber, Lucia; Derfuss, Klaus; Endrikat, Jan

Published in:
Accounting, Organizations and Society

DOI:
10.1016/j.aos.2022.101414

Publication date:
2023

Accounting, Organizations and Society 106 (2023) 101414

Contents lists available at ScienceDirect

Accounting, Organizations and Society

journal homepage: www.elsevier.com/locate/aos

Taking stock of research on the levers of control with meta-analytic methods: Stylized facts and boundary conditions

Lucia Bellora-Bienengräber [a,*], Klaus Derfuss [b], Jan Endrikat [c]

[a] University of Groningen, Faculty of Economics and Business, Department of Accounting, Groningen, the Netherlands
[b] University of Hagen, Faculty of Business Administration and Economics, Hagen, Germany
[c] University of Applied Sciences (FHD), Faculty of Economics, Dresden, Germany

A R T I C L E I N F O

Article history:
Received 29 October 2019
Received in revised form
20 October 2022
Accepted 1 November 2022
Available online 22 November 2022

Keywords:
Management control systems
Levers of control
Organizational capabilities
Organizational performance
Meta-analysis

A B S T R A C T

In levers of control (LoC) research, empirical and conceptual ambiguities have hampered the establishment of a coherent body of knowledge. Mixed findings, variability in the approaches to account for the levers' combined use, and variability in conceptual choices (e.g., the conceptualization of interactive and diagnostic control) have caused this unsatisfactory state. In response, we validate and extend theory on the LoC framework by meta-analytically synthesizing quantitative evidence from 58 independent samples and 10,374 observations. We develop two models of the combined use of the levers, which portray their simultaneous use and mutual relationships, and relate them to capabilities and performance. For theory validation, we uncover stylized facts that demonstrate that organizations use the four levers in combination, not in isolation. Moreover, following the logic of the resource-based view, the levers are related to performance via capabilities. These relationships are robust to moderating influences of the dimensions and conceptualization of interactive control and managers' hierarchical level. For theory extension, we systematically uncover the need to complement the resource-based view with other theories and offer related suggestions. Our moderator analyses identify boundary conditions that limit the generalizability of the LoC framework. For example, surprisingly, the conceptualization of diagnostic control emerges as a boundary condition. On a general level, our findings might serve as an inspiration for better appreciating future survey-based knowledge creation in management control research and also provide researchers from other disciplines with a more comprehensive understanding of the enhancement of capabilities and performance.

1. Introduction

Within the past 25 years, Simons' (1995) levers of control (LoC) framework has inspired a considerable amount of research in management control systems (MCS). Simons maintains that four levers—beliefs control, boundary control, and diagnostic and interactive use of management control practices—facilitate the emergence and implementation of business strategy and the management of organizational tensions to achieve desired outcomes. A central theme in empirical LoC research is whether and how the levers help organizations to gain superior performance. LoC studies, which often explicitly or implicitly rely on the resource-based view (RBV) of the firm (Barney, 1991), suggest that the emphasis on the levers entails an organizational context expected to foster performance by enhancing organizational capabilities, such as innovativeness and organizational learning (Bedford, Malmi, & Sandelin, 2016; Henri, 2006; Widener, 2007).[1]

LoC research has attracted both enthusiasm and criticism. A considerable number of survey studies have investigated the relationships between the levers and capabilities as well as performance. Nevertheless, LoC research is criticized for its conceptual and empirical ambiguities, which may hamper progress of

* Corresponding author.
E-mail addresses: l.bellora-bienengraeber@rug.nl (L. Bellora-Bienengräber), klaus.derfuss@fernuni-hagen.de (K. Derfuss), j.endrikat@fh-dresden.eu (J. Endrikat).

[1] For concision, we label 'diagnostic use of management control practices' and 'interactive use of management control practices' as 'diagnostic control' and 'interactive control'. We also simplify the 'emphasis on the levers' use' and similar terms (e.g., emphasis on diagnostic control) by using 'the levers' use' and similar terms (e.g., diagnostic control). Finally, we use the terms 'capabilities' for 'organizational capabilities' and 'performance' for 'organizational performance.'

https://doi.org/10.1016/j.aos.2022.101414

knowledge and practical applicability (Curtis, Lillis, & Sweeney, 2017; Martyn, Sweeney, & Curtis, 2016; Tessier & Otley, 2012).

Although we concur with these criticisms, we take a more positive angle. We believe that LoC research is ripe for making sense of these ambiguities. Using meta-analysis, we validate and extend theory about the relationship among the levers, capabilities, and performance.

Specifically, we aim to achieve two goals. First, we capitalize on the existing, seemingly heterogeneous, studies to derive stylized facts (Kaldor, 1961), or "stable patterns that emerge from many different sources of empirical data" (Heine, Meyer, & Strangfeld, 2005, p. 3).[2] We aggregate results from 58 independent samples from quantitative studies (see Web Appendixes 1 and 2), which total 10,374 observations. On this basis, we propose and test two meta-analytic path models to estimate how the levers are related to each other and with capabilities and performance. We identify the *what*, *how*, and *why* of the nomological network of LoC[3] (Whetten, 1989). Second, we explore the effects of the conceptual choices (e.g., different conceptualizations of the levers) of prior studies on the relationships among the levers and with capabilities and performance via meta-analytic moderator analyses. Thereby, we assess whether these relationships are robust to potential moderators and identify their boundary conditions.[4] We thus highlight limits to these relationships' generalizability, or the *who*, *where*, and *when* of the nomological network of the levers (Whetten, 1989).

Overall, this endeavor is timely and non-trivial due to the sizable ambiguities inherent in the LoC literature. We address three areas of ambiguity: mixed findings, variability in the approaches to account for the levers' combined use, and variability in conceptual choices. We summarize these ambiguities, related examples, and the conclusions from our meta-analysis in Table 1.

The first area of ambiguity arises from mixed findings in the literature in regard to specific relationships and their mechanisms (e.g., Curtis et al., 2017). Not only do some findings differ in the strength, significance, or direction of the relationships among levers, capabilities, and performance, but the studies also differ in whether they disentangle the levers' direct performance effect from their indirect performance effects through capabilities. Therefore, ambiguity regarding the relationships and the specific type of mediation that is at play, i.e., indirect-only, complementary, or competitive mediation,[5] diminishes the opportunity for researchers to identify potentially omitted theoretical links and related mediators (Zhao, Lynch, & Chen, 2010).

The second area of ambiguity concerns the disparate approaches used to account for the levers' combined use. Despite Simons' (2000) claim that the levers' power lies in their combined use, many studies investigate only one or two in isolation (Martyn et al., 2016). This raises the question of whether the respective findings hold when all four levers are considered in combination (Bedford et al., 2016). Moreover, studies that focus on all four levers model them in conceptually different ways by either treating them as independent (Janka & Guenther, 2018) or acknowledging their combined use via second-order constructs (Speklé, van Elten, & Widener, 2017) or mutual paths among the levers (Heinicke, Guenther, & Widener, 2016). Whereas second-order constructs allow testing *broad theories* about the overall LoC environment at a high level of abstraction, mutual paths models allow testing *specific theories* of the individual levers and their antecedents and/or outcomes (Edwards, 2001). Although both models are appropriate in their respective contexts, a conceptual and empirical comparison of these models is missing.

The third and final area of ambiguity concerns the variability in studies' conceptual choices, such as those regarding the conceptualization of the levers. This variability may cause divergent findings for the same relationship (Bisbe, Batista-Foguet, & Chenhall, 2007; Tessier & Otley, 2012). These divergent choices, if not explicitly conceptually and empirically acknowledged, hamper the establishment of generalizable knowledge in the field.

By addressing the three areas of ambiguities outlined above, our study makes important contributions to the MCS literature. First, we validate theory on the nomological network of the levers. We show that, for enhancing performance, organizations use the four levers in combination and not in an isolated manner. Moreover, we find positive total levers—performance effects. From a broad theory perspective, these positive effects are partially mediated by capabilities. From a specific theory perspective, with mutual paths that link the levers, the indirect effects partially or completely explain the levers—performance relationships. Additionally, and in contrast with prior conceptual work, our moderator analyses show that the relationships are robust against differences in conceptual choices related to the operationalization of interactive control and the hierarchical levels of primary studies' respondents. Because we aggregate a considerable amount of data and account for (1) measurement and sampling error (i.e., between-study fluctuations in reliability and sample size, respectively) and common method bias that may have caused some of the previous mixed findings, (2) the context variables of environmental uncertainty and size, and (3) conceptual choices, we provide a more robust and fine-grained account of the nomological network of the levers than has prior research.

Second, we extend theory on the nomological network of the levers. Most studies explicitly or implicitly draw on the RBV, whose logic explains parts of these levers—performance relationships. We show, however, that the direct performance effects likely mask mediators rooted in other theories. We provide examples of psychological or sociological theories that suggest such mediators. Moreover, we extend knowledge on the nomological network of the levers by unveiling boundary conditions rooted in the conceptualizations of the levers, capabilities, and performance as well as in industry differences. These differences partially explain prior studies' heterogeneous findings.

Third, on a more general level, we contribute to a shift in the perception of survey-based MCS studies—from regarding them as producing idiosyncratic findings that result from construct reliability and validity issues or small samples (Chenhall, 2003; Speklé & Widener, 2020) to seeing them as an indispensable part of the "evolution of science" (Combs, Ketchen, David, Crook, & Roth, 2011,

[2] 'Stylized facts' explain the essential characteristics of a phenomenon in an adequate, but parsimonious, way. Deriving stylized facts requires a high level of structure and intersubjective reproducibility in the process (Heine et al., 2005). Methods such as meta-analyses and systematic literature reviews fulfil these requirements. Deriving stylized facts from LoC research means, for example, understanding how the levers relate to performance based on prior empirical evidence.

[3] A nomological network is a "web of relationships" that specifies "how the construct of interest relates to other constructs" (Schwab, 2005, p. 26). We use the term 'nomological network of the levers' for the relationships among the levers, capabilities, and performance and 'LoC framework' for the levers and their combined use.

[4] Boundary conditions are variables that "place limitations on the propositions generated from a theoretical model" and "constitute the range of the theory" (Whetten, 1989, p. 492). They are empirically captured by moderator variables (Schwab, 2005). A likely boundary condition for the nomological network of the levers is whether interactive and diagnostic controls refer to the use of budgeting or to the use of performance measurement.

[5] An indirect-only mediation implies that the mediator fully accounts for the relationship, and no substantial direct effect remains after accounting for the mediator (Zhao et al., 2010). Both complementary and competitive mediation are indicators of an "incomplete theoretical framework" (Zhao et al., 2010, p. 201). The mediator is consistent with the theory, but the remaining significant direct effect points to the existence of another, omitted mediator. If the direct and the indirect effect have the same sign, then the omitted mediation effect is a complementary one, whereas different signs indicate competitive mediation (Zhao et al., 2010).

Table 1
Main areas of empirical and conceptual ambiguities in the nomological network of the levers and conclusions from this meta-analysis.

Panel A: Mixed findings regarding specific relationships and their mechanisms	
Examples for ambiguities	**Related conclusions from this meta-analysis**
Relationships among the levers	
- *Diagnostic with interactive control* → e.g., Bedford (2015): positive; Acquaah (2013): negative	All relationships among the levers are positive.
-*Beliefs with boundary control* → e.g., Bedford (2015): positive; Naranjo-Gil (2016): insignificant	
-*Boundary with diagnostic control* → e.g., Heinicke et al. (2016): positive; Schaefer and Guenther (2016): insignificant	
Relationships between the levers and capabilities	
-*Diagnostic control with capabilities* → e.g., Widener (2007): positive; Bisbe and Malagueño (2015): insignificant	All relationships between the levers and capabilities are positive.
Relationships between the levers and performance	
-*Interactive control with performance* → e.g., Bedford (2015): positive; Garcia Osma, Gomez-Conde, and de las Heras (2018): insignificant	All relationships between the levers and performance are positive.
Mechanisms linking the levers with performance	
-*Direct performance effects only* → e.g., Guenther and Heinicke (2019) or Bedford (2015) -*Indirect effects* via *capabilities* → e.g., Henri (2006) or Widener (2007)	The relationship between interactive control and performance is fully mediated by capabilities. The relationships between the other levers and performance are partially mediated by capabilities.
Panel B: Variability in the approaches to account for the levers' combined use	
Examples for ambiguities	**Related conclusions from this meta-analysis**
Studies investigating all four levers together rely on conceptually different types of models, among others: -*Independent levers* → e.g., Janka and Guenther (2018) or Naranjo-Gil (2016) -*Second-order construct* → e.g., Speklé et al. (2017) or Bellora-Bienengräber, Radtke, and Widener (2022) -*Mutual paths* → e.g., Heinicke et al. (2016) or Journeault, De Rongé, and Henri (2016)	Models that acknowledge the levers' combined use (i.e., by using a second-order construct or mutual paths among the levers) outperform a model that implies independence of the levers.
Panel C: Variability in conceptual choices	
Examples for ambiguities	**Related conclusions from this meta-analysis**
A. Conceptualization of the levers	
Variability in the conceptualization of diagnostic control	
Studies use conceptually different items and follow (1) Henri (2006), (2) Widener (2007), or (3) other approaches (e.g., Bedford, 2015).	Boundary condition
Variability in the dimensions of interactive control	
Studies focus on different dimensions and include (1) only the intensity of use dimension (Bedford, Bisbe, & Sweeney, 2019), (2) add the enabling role (Heinicke et al., 2016), and (3) additionally add the strategic validity dimension (Bedford, 2015).	No boundary condition
Variability in the conceptualization of interactive control	
Studies use conceptually different items and follow (1) Henri (2006), (2) Widener (2007), or (3) other approaches (e.g., Bedford, 2015).	No boundary condition
Management control practice addressed in the conceptualizations of interactive and diagnostic control	
Studies refer to management control practices, such as (1) performance measurement (Henri, 2006), (2) budgeting (Hofmann, Wald, & Gleich, 2012), or (3) other practices (e.g., Bisbe & Malagueño, 2015).	Boundary condition
General vs. specific control problem addressed in the conceptualizations of interactive and diagnostic control	
Studies refer to the (1) general control of the organization (Bedford et al., 2019) or to (2) specific control problems, like implementing an environmental strategy (e.g., Journeault et al., 2016).	Boundary condition
Intention for vs. perception of control	
Studies capture (1) superiors' intentions for (Bedford, 2015) or (2) subordinates' perceptions of control (Kruis, Speklé, & Widener, 2016).	Boundary condition
Hierarchical level at which the data is gathered	
Studies sample (1) senior (Bisbe & Otley, 2004) or (2) lower-level managers (Kruis et al., 2016).	No boundary condition
B. Conceptualization of capabilities and performance	
Variability in the conceptualization of capabilities	
Studies capture (1) innovativeness (Bedford, 2015), (2) organizational learning (Widener, 2007), or (3) other capabilities (e.g., Heggen, 2019).	Boundary condition
Variability in the conceptualization of performance	
Studies use (1) financial (Bisbe & Malagueño, 2015), (2) non-financial (Kruis et al., 2016), or (3) general indicators (including both financial and non-financial ones; e.g., Bedford and Malmi (2015)).	Boundary condition
C. Industry	
Manufacturing vs. service industry samples	
Studies use (1) manufacturing (Lopez-Valeiras, Gonzalez-Sanchez, & Gomez-Conde, 2016), (2) service (Budianto & Yuliansyha, 2014), or (3) mixed samples (Bedford, 2015).	Boundary condition

p. 178). For LoC research, we dismantle some of the concerns and show that survey studies contribute effectively to our understanding of MCS relationships. Indeed, after accounting for measurement and sampling error and for conceptual choices as sources of variation, we show that survey studies, even with a limited sample size, provide a solid and coherent picture of the functioning of MCS and, thus, are essential for generating knowledge. Moreover, the diversity of survey studies, for example, in regard to industries or respondents, has often been blamed as a source of conflicting findings. Although single-survey studies certainly cannot capitalize on this diversity, meta-analyses such as ours can leverage it for testing theories' robustness and identifying boundary conditions. Thus, survey studies collectively contribute to the evolution of science. Thus, overall, we hope that this exemplary case of aggregating survey studies helps shed a more positive light on future survey-based MCS research.

In line with evidence-based management[6] (Rousseau, 2006), our study also has implications for practice and teaching. Prior MCS studies, even those strongly anchored in practice, seemingly fail to yield convergent findings and, thus, could mislead practice (Merchant & Otley, 2020). In contrast, our study uncovers robust positive levers—performance relationships via capabilities and, thus, can encourage organizations to increase their emphases on the levers and their combined use. At the same time, we show that caution is required because some boundary conditions (e.g., industry differences) must be considered to contextualize the use of the levers.

2. Theoretical background and hypothesis development

This study aims to validate and extend the theory about the relationships among the levers, capabilities, and performance. In this section, we build on the RBV to provide a rationale for these relationships. Then, we theoretically develop two alternative models and related hypotheses to explain these relationships. Finally, we outline conceptual choices that may constitute boundary conditions for the nomological network of the levers. We define the focal constructs in Table 2.

2.1. Levers of control, capabilities, and performance

According to the RBV, capabilities such as innovativeness, organizational learning, entrepreneurship, or market orientation (Henri, 2006) allow organizations to gain sustainable competitive advantage and, thus, improve performance. Capabilities are valuable, rare, imperfectly imitable, and non-substitutable resources (Barney, 1991, 1995) that have widely documented positive performance effects (Karna, Richter, & Riesenkampff, 2016; Schweiger, Stettler, Baldauf, & Zamudio, 2019).[7]

The levers relate to capabilities for two reasons. First, to gain sustained competitive advantage, an organization must be organized and controlled in a way that facilitates leveraging its capabilities. In this context, Barney (1995) explicitly emphasizes the role of formal MCS. Second, to overcome the danger that capabilities become dysfunctional routines, continuous monitoring of capabilities—as a double-loop learning process that can be facilitated by the levers' use—is necessary (Schreyögg & Kliesch-Eberl, 2007).[8] Simons (1995) argues that managers use the levers for successful strategy implementation, organizational learning, and effective resource allocation. Specifically, they rely on beliefs, boundary, diagnostic, and interactive control to monitor the four key strategic variables of core values, risks to be avoided, critical performance variables, and major strategic uncertainties, respectively. The combined use of the levers can generate a fruitful tension between "opportunistic innovation and predictable goal achievement" (Simons, 1995, p. 153) that contributes to the development of capabilities and, ultimately, performance (Henri, 2006; Mundy, 2010; Widener, 2007).

The four levers relate to performance via capabilities for distinct, levers-specific reasons. First, compared to those of other organizations, employees of organizations that clearly communicate their mission and core values via beliefs control are more empowered and encouraged to develop new ideas and initiatives (Blount & Leinwand, 2019; Speklé et al., 2017), thereby enhancing existing and new capabilities. Beliefs control fosters stability and continuity but also helps to enable organizational change (Simons & Dávila, 2021), thus reducing the possibility that routines become dysfunctional.

Second, boundary control assures that employees' opportunity seeking is restricted to areas that are defined as acceptable. On the one hand, boundary control establishes business conduct guidelines and clearly defined limits for precluding dysfunctional behaviors (Simons, 1995). For example, it helps the organization to protect its capabilities from becoming imitable by precluding competitors from hiring executives and, thus, acquiring capability-related information. On the other hand, boundary control establishes strategic boundaries that specify those opportunities in which managers "do not want the organization to expend resources" (Simons, 1995, p. 48). Thus, boundary control helps to focus subordinates' efforts and discourages them from seeking adjustments beyond optimal and timely solutions (Bedford, 2015; Mundy, 2010).

Third, diagnostic control enables single-loop learning. Although the development of capabilities generally requires double-loop learning, diagnostic control also contributes to enhanced capabilities because it provides essential feedback on how well existing capabilities are applied. It thus increases organizations' understanding of these capabilities and their application and helps to avoid wasting resources by establishing clear targets (Simons, 1995). Further, the related management-by-exception approach frees up managers' time to engage in the development of new capabilities (Chenhall & Moers, 2015). Finally, by giving orientation and structure to employees, diagnostic control facilitates the coordination and embedment of new routines in the organization's practices (Mundy, 2010; Simons & Dávila, 2021; Speklé et al., 2017).

Fourth, interactive control is intended to inspire double-loop learning processes and induce organizational creativity and innovativeness (Simons, 1995). The continual challenge to and debate about set assumptions allows the defining of new and innovative routines, processes, or product ideas (Mundy, 2010) and, thus, fosters the development and exploitation of capabilities.

These RBV-based arguments focus on the individual levers.

[6] Evidence-based management "derives principles from research evidence and translates them into practices that solve organizational problems" (Rousseau, 2006, p. 256). Its usefulness depends on the clarity of evidence and on its suitability for the specific practical setting (Rousseau, 2006). This study generates evidence that can be translated into practice.

[7] During our data collection process, we reviewed all of the variables used in the LoC survey studies. In this process, we noticed that the majority include some type of capability in their models. Some of the studies explicitly refer to the RBV, while others include the logic of the RBV without explicitly mentioning it. This strong reliance on the RBV logic guided our choice to use the RBV as a basis for our nomological network because it enables us to validate and extend the extant LoC research.

[8] "The single loop learning keeps a process within desired bounds; double loop learning leads to question about the very basis upon which strategies have been constructed" (Simons, 1995, p. 106; also see Argyris (1976)).

Table 2
Definitions of the constructs included in this meta-analysis.

Constructs	Definition in this meta-analysis
Constructs included in the examined relationships	
Beliefs control	"… the explicit set of organizational definitions that senior managers communicate formally and reinforce systematically to provide basic values, purpose, and direction for the organization" (Simons, 1995, p. 34).
Boundary control	The formal system that "… delineates the acceptable domain of strategic activity for organizational participants" (Simons, 1995, p. 39).
Diagnostic control	The formal system that aligns employees' behavior with organizational goals based on critical success factors and management by exceptions (Simons, 1995).
Interactive control	The formal system that stimulates "… search and learning, allowing new strategies to emerge as participants throughout the organization respond to perceived opportunities and threats" (Simons, 1995, p. 91).
Simultaneous use of the levers	This construct captures a combined use that is coherent in that it is simultaneously reflected in the emphasis on the four levers; it is concerted because if the emphasis on the simultaneous use of the levers increases (decreases), the emphasis on each lever correspondingly also increases (decreases); and it is purposefully directed towards the solving of control problems. Empirically, this combined use of the levers is captured as their common variance (Edwards, 2001; Law & Wong, 1999; Speklé et al., 2017).
Capabilities	"… specific processes, activities, or competences that enable the organization to perform and gain competitive advantages" (Franco-Santos, Lucianetti, & Bourne, 2012, p. 80). Among others and in accordance with Henri (2006), we include innovativeness, organizational learning, entrepreneurship, and market orientation capabilities in our notion of capabilities.
Performance	"… the economic outcomes resulting from the interplay among an organization's attributes, actions, and environment" (Combs, Crook, & Shook, 2005, p. 261). Under this notion, we include financial, non-financial, and general (i.e., including both) conceptualizations.
Environmental uncertainty	The gap between the information required and the information available to make a decision (Galbraith, 1973), i.e., the predictability and stability of the environment (Gordon & Narayanan, 1984).
Size	The size of the focal entity, i.e., an organization or unit, measured as the absolute value or logarithm of number of employees or sales. We use "size" in the text as an umbrella term for both organizations and units.
Factors potentially moderating the examined relationships	
Conceptualization of diagnostic control	We distinguish between the conceptualizations suggested by Henri (2006), by Widener (2007), and others. Other conceptualizations include studies that use a mix of both of these conceptualizations and studies using a different and/or a broader conceptualization. We exclude studies that do not fully disclose their conceptualization from the moderator analysis.
Dimensions of interactive control	We distinguish, in line with the definitions by Tessier and Otley (2012), between conceptualizations (1) including only the intensity of use dimension, (2) combining the intensity of use and the enabling role dimensions, and (3) combining the intensity of use, enabling role, and the strategic validity dimensions. We exclude other combinations from the moderator analysis.
Conceptualization of interactive control	We distinguish between the conceptualizations suggested by Henri (2006), by Widener (2007), and others. Other conceptualizations include studies that use a mix of both of these conceptualizations and studies using a different and/or broader conceptualization. We exclude studies that do not fully disclose their conceptualization from the moderator analysis.
Management control practice	We distinguish between performance measurement and budgeting. We exclude studies including other or more general management control practices from the moderator analysis.
Control problem	We distinguish between general control problems and specific control problems (e.g., new product development).
Intention for vs. perception of control	We distinguish between studies addressing superiors' intention for control and subordinates' perception of control. We exclude studies that remain unclear about the concrete questions asked from the moderator analysis.
Conceptualization of capabilities	We distinguish between innovativeness, organizational learning, and other capabilities. We exclude studies that include several capabilities in one aggregated construct from the moderator analysis.
Conceptualization of performance	We distinguish between financial, non-financial, and general (i.e., including both financial and non-financial) conceptualizations of performance.
Hierarchical level	Senior managers encompass all C-suite managers (e.g., CEO, CFO); all other respondents with supervisory duties belong to the lower-level managers' category. We exclude studies mixing the level of respondents or not including managers from the moderator analysis.
Industry	Categorization of samples based on the sectors of operations of the organizations investigated in the primary studies. We distinguish between manufacturing vs. services industries. We exclude studies of mixed industries or those that do not report the industries included from the moderator analysis.

Simons (1995), however, notes that "the four control levers are nested—they work simultaneously but for different purposes" (p. 5), such that their power "does not lie in how each is used alone" (p. 153). Therefore, combining the levers' uses is essential because it provides an organizational context that simultaneously allows for developing, exploiting, and monitoring capabilities.

Despite these theoretical arguments, conceptual and empirical ambiguities remain. First, research findings in regard to the relationships among the levers, capabilities, and performance are mixed (Table 1, Panel A). Second, some prior studies focus only on selected levers and/or ignore their combined use (Table 1, Panel B). Seeking to address this tension, we conceptually synthesize these heterogeneous approaches into two theory-based models, each of which captures the combined use of the levers and the respective hypotheses.

2.2. Broad and specific theories on the relationships of the levers of control with capabilities and performance

Simons (1995, 2000) leaves room for interpretation of the form of the combined use of levers, labelling the levers as "nested" (Simons, 1995, p. 5) or "not used alone" (Simons, 2000, p. 301). Thus, for modeling the levers' combined use, existing studies use various approaches (Table 1, Panel B). We structure these approaches and related theorizing around the distinction between the perspectives of broad and specific theories (Edwards, 2001).

To capture the multidimensional nature of the LoC framework, scholars draw on a broad (e.g., Bellora-Bienengräber et al., 2022; Speklé et al., 2017) or a specific theory (e.g., Heinicke et al., 2016), and these theories differ in their levels of abstraction. A broad theory focuses on relationships between general constructs (e.g., the levers' combined use) that are reflected in specific dimensions (e.g., the levers' combined use as reflected in the four levers) and generates a higher-order understanding of the relationships at hand. A specific theory focuses on the relationships among the specific dimensions of the focal constructs (e.g., the four levers) and teases out idiosyncrasies in these relationships that would be invisible (and deemed irrelevant) in broad theories (Edwards, 2001; Law & Wong, 1999).[9]

[9] We concur with Judge and Kammeyer-Mueller (2012) who stress that artificially dichotomizing broad and specific theories may hamper research progress. Researchers should be aware of the respective differences and shape their research projects accordingly. These perspectives may complement each other, as theories might explain how a multidimensional construct and its dimensions relate to each other and to other variables (Edwards, 2001).

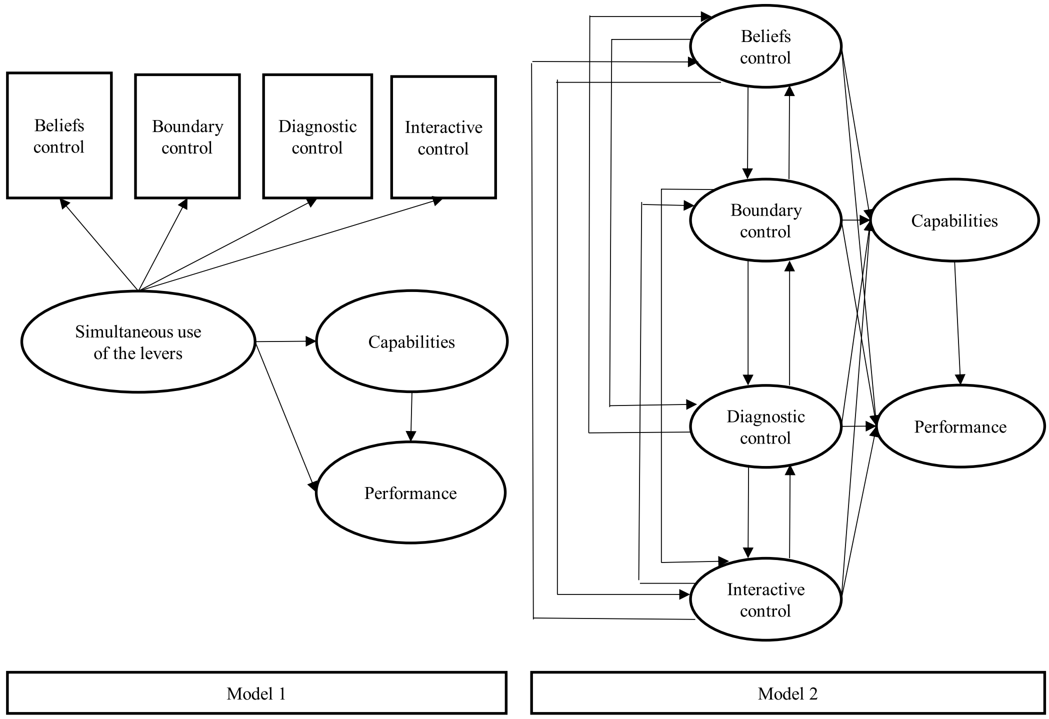

Fig. 1. Theoretical models: simultaneous use of the levers (Model 1) and mutually related levers (Model 2)
Note: Both models include environmental uncertainty and size as control variables (modelled as antecedents to all constructs).

For the LoC framework, a broad theory investigates whether the combined use of the levers—in our broad theory model, labelled 'simultaneous use of the levers'—enhances capabilities and, subsequently, performance, thus explaining these relationships on a higher level of abstraction than that for the individual levers (Tanriverdi & Venkatraman, 2005). The simultaneous use of the levers refers to their common core (George, 2011; Speklé et al., 2017) and concerns how the four levers are used in combination. This combined use is coherent in that it is simultaneously reflected in the emphasis on the four levers; it is concerted because, if the emphasis on the simultaneous use of the levers increases (decreases), the emphasis on each lever correspondingly increases (decreases); and it is purposefully directed toward the solving of control problems. This is in line with Simons' (1995) reasoning that "control of business strategy is achieved by integrating the forces of beliefs control, boundary control, diagnostic control, and interactive control" and that "they complement each other when used together" (p. 153). Building on our RBV-based arguments in Section 2.1, the simultaneous use of the levers provides the organizational context necessary for developing, exploiting, and monitoring capabilities and, ultimately, fosters performance (Fig. 1, Model 1). Stated formally:

H1. The simultaneous use of the levers relates positively to performance through enhanced capabilities.

A specific theory facilitates testing whether the relationships between the individual levers and their relationships with capabilities and performance have the same direction and/or strength.[10] Advocates of specific theories favor these theories' clarity and unambiguous character, rooted in their low level of abstraction (Edwards, 2001). By relying on the arguments in Section 2.1, our specific theory model—labelled 'mutually related levers'—captures the levers' distinct relationships with performance through capabilities. Specifically, it clarifies that the emphasis on one of the four levers (e.g., beliefs control) is more or less strongly coupled with the emphases on the other levers (e.g., boundary, diagnostic, and interactive control). Thus, it mirrors their interdependencies and acknowledges that the levers are not used in isolation. This combined use is in line with Simons' (1995) reasoning, in which he stresses the levers' "continual interplay" (p. 30) and states that they "are mutually reinforcing" (p. 161). For example, beliefs and boundary control are mutually related because a strengthening of boundaries should go hand in hand with a stronger communication of the core values (and vice versa) to guide employees' opportunity-seeking. Overall, we expect positive indirect relationships of the individual levers with performance through the other levers and capabilities (Fig. 1, Model 2). Stated formally:

H2. (a) Beliefs control, (b) boundary control, (c) diagnostic control, and (d) interactive control relate positively to performance through the enhancement of the respective other levers and capabilities.

2.3. Conceptual choices as potential moderators

Conceptual works (Bisbe et al., 2007; Tessier & Otley, 2012) highlight ambiguities in the LoC literature that translate into different conceptual choices as well as vagueness on the empirical level. If these choices cause variability in primary studies' findings, they are important moderators and, thus, represent boundary conditions for the nomological network of the levers. Otherwise, the relationships would be robust to these conceptual choices. Motivated by prior studies (Bisbe et al., 2007; Tessier & Otley, 2012) and meta-analyses (Derfuss, 2015), we examine at the bivariate

[10] In principle, specific LoC models work best if multidimensional constructs that are used as dependent variables also are studied at the dimension level. Unfortunately, LoC research currently does not allow disentangling meta-analytically the multidimensional constructs of capabilities and performance.

level the influence of the conceptualization of the levers (conceptualization of diagnostic control, dimensions and conceptualization of interactive control, management control practice, control problem, intention for vs. perception of control, and hierarchical level) and conceptualization of capabilities and performance as well as industry. We introduce these choices in Table 1, Panel C, define them in Table 2, and describe them in detail in Web Appendix 3.[11]

3. Method

Meta-analysis is a method for establishing reliable stylized facts, reconciling seemingly inconsistent findings, and identifying theories' boundary conditions (Geyskens, Krishnan, Steenkamp, & Cunha, 2009). After we describe our data collection, we present the meta-analytic procedures and close with ex ante analyses of potential validity threats from common method bias and endogeneity issues.

3.1. Literature search and coding of relevant studies

We adopted a literature search procedure that guarantees a sample of studies that is as complete as possible. As a primary means for retrieving studies eligible for inclusion, we searched the Google Scholar platform for studies that cite Simons' (1995) book. Second, we searched the *Academic Search Complete*, *Business Source Complete*, *EconLIT*, *Elsevier ScienceDirect*, *Emerald|Insight*, *Social Science Research Network*, and *Wiley Online* electronic databases, using the following search string: *AB (control AND ("lever* of control" OR LOC OR "belief* system*" OR "boundar* system*" OR diagnostic OR interactive)) AND TX (simons).*[12] Fourth, we searched the reference sections of relevant review papers (Chenhall, 2003; Langfield-Smith, 2007; Martyn et al., 2016). Finally, after selecting those studies that are relevant for our analyses, we examined their references sections to identify further papers.

We included all studies that focus on at least one of the levers, follow a quantitative data collection approach, and were publicly available by October 31, 2018. After attempting to obtain study effects for those investigations that do not report correlations by contacting the authors, we included all studies that provide correlation coefficients. To guarantee statistical independence of the samples, we used Wood's (2008) procedure for detecting multiple uses of samples.[13] Each of the three authors coded the studies separately. We then compared our codes, discussed any differences, and came to an agreement by referring to the respective studies. We categorized the variables based on our definitions (Table 2) and a thorough examination of the construct operationalizations provided in the studies, which means that we sometimes deviated from the construct labels used in the primary studies.[14] This led to 69 independent samples, each investigating parts of our 28 bivariate relationships. Due to potential common method bias issues, we further reduced this dataset to 58 independent samples for our analyses, as detailed in section 3.3. We list all independent samples and related papers in Web Appendix 1.

3.2. Meta-analytic procedures and data preparation

To analyze our data, we use meta-analytic procedures with corrections for sampling and measurement errors, as developed by Schmidt and Hunter (2015) and in line with prior MCS meta-analyses (Derfuss, 2015; Endrikat, Guenther, & Titus, 2020). For each distribution of correlations, we correct each observed correlation for measurement error, using the respective reliability coefficients. To do so, we calculate an attenuation factor, which is the product of the square roots of the respective reliability coefficients. When the reliability coefficient is not available, we use the means of all reliability coefficients for the respective variable. To estimate the weighted mean true-score correlation (ρ) and the related standard deviation, we aggregate the correlations by calculating their weighted average, using the product of the square of the attenuation factor and the respective sample size as weight. Finally, we use 95% confidence intervals to determine the significance of ρ.

To test our hypotheses, we use meta-analytic path modeling with IBM SPSS Amos 28. Because the meta-analytic correlations are already corrected for measurement and sampling error, we model all variables as observed (Hogreve, Iseke, Derfuss, & Eller, 2017). The basis for our analyses is a meta-analytic correlation matrix. The sample size for model estimation is the harmonic mean of the correlations' total sample sizes (Schmidt & Hunter, 2015).

For the broad theory perspective, we model the simultaneous use of the levers (Model 1) as a reflective second-order construct (Speklé et al., 2017). A reflective second-order construct manifests itself via its dimensions such that changes in the second-order construct affect the first-order dimensions (i.e., the individual levers). Only the variance common to all four levers, and, thus, the levers' common core, pertains to the second-order construct, while group, specific, and random variances are attributed to the dimensions' error terms (George, 2011; Law & Wong, 1999; Petter, Straub, & Rai, 2007).[15] For the levers, this means that, for example, if the emphasis on the simultaneous use of the levers increases, the emphases on the beliefs, boundary, diagnostic, and interactive control increase correspondingly. For the specific theory perspective, we empirically account for the interdependencies among the levers with mutual paths (Model 2), in line with

[11] Although these conceptual and methodological choices may be perceived as mere empirical operationalization issues, we emphasize that they may constitute boundary conditions for the nomological network of the levers (Libby, Bloomfield, & Nelson, 2002). Different measurement approaches of constructs may tap distinct construct domains and, thus, result in different theoretical relationships. Therefore, as theoretical differences drive empirical ones (Libby et al., 2002), we treat these choices as conceptual moderators. For example, one can view the dimensions of the interactive control construct as merely an empirical matter, as reflected in different (groups of) items used in a survey. The difference in (groups of) items, however, reflects an underlying difference in the conceptual domain of the construct, as described by Bisbe et al. (2007) and Tessier and Otley (2012). Thus, empirical differences translate into conceptual differences. Therefore, we refrain from artificially distinguishing methodological from conceptual moderators.

[12] AB means abstract and TX, text.

[13] If multiple studies build on one sample, we included the correlation only once (e.g., Bisbe & Malagueño, 2009; Bisbe & Otley, 2004). If a study contains conceptual replications, such as two subscales for a variable (e.g., Bedford, 2015, for the combination of environmental hostility and dynamism as subscales of environmental uncertainty), we computed composite correlations to adjust for interdependence. Using the Mosier formula, we computed the respective reliability coefficients (Schmidt & Hunter, 2015).

[14] For the decision about how to categorize and include a variable in our meta-analysis, we relied on the conceptual definition provided by the respective primary study. Sometimes, however, we observed differences in the empirical measurement of constructs meant to be conceptually equivalent. We used these differences, whenever possible, as the basis for our moderator analyses (see Section 2.3 and Web Appendix 3). We excluded variables from the aggregation in our meta-analysis if they are conceptually too distinct from our conceptual definition

[15] An alternative specification of the broad theory perspective is a formative one, which assumes that the four levers 'cause' (i.e., 'form') the higher-order construct (Bellora-Bienengräber et al., 2022) and which could be labelled as 'additive use of the levers'. The distinction between reflective and formative second-order constructs is conceptually meaningful (Speklé et al., 2017). Compared to our reflective model specification, however, we find similar results for the relationships among levers, capabilities, and performance such that the implications remain qualitatively unchanged. We tabulate and briefly discuss the formative model specification in Web Appendix 4. For simplification, we use the term 'second-order' to refer to our reflective model specification.

Heinicke et al. (2016).[16] For the levers, this means that, for example, increasing the emphasis on beliefs control might be associated with an increased emphasis on boundary control.

Similar to Tanriverdi and Venkatraman (2005), we contrast Models 1 and 2 with a baseline model—labelled 'isolated use of the levers'—that assumes that the levers are not associated with each other via mutual paths (i.e., their relationships are constrained to zero). Specifically, we compare the models' relative fit to the data and parsimony using their AIC and BIC coefficients. Moreover, we compare the respective R^2 of performance.

For our moderator analyses, we focus on the meta-analytic correlations. We first assess the homogeneity of each ρ across samples via the 75% rule and 95% credibility intervals (Geyskens et al., 2009). According to the 75% rule (Schmidt & Hunter, 2015), moderating variables need to be analyzed only if less than 75% of the variance is attributable to sampling and measurement error and, thus, in turn, more than 25% of the variance is attributable to moderator variables. By relying on the respective standard deviation, 95% credibility intervals provide an estimate of the variability of a correlation's observed distribution. Moderating variables should be analyzed if the interval is wide and/or includes zero.

Because our moderating variables are categorical, we use subgroup analyses (Geyskens et al., 2009; Schmidt & Hunter, 2015). If the subgroup ρs differ significantly and the average corrected variance is lower across subgroups than in the overall analysis, a moderating influence exists. As a test, we rely on 95% confidence intervals around the difference of the subgroup ρ (Hunter & Schmidt, 2000). We seek to conduct subgroup analyses with regard to all potential moderators, as presented in Section 2.3, for the focal relationships, i.e., those (1) among the levers, (2) between the levers and capabilities as well as performance, and (3) between capabilities and performance. At least three correlations, however, must be available for each subgroup (Dalton, Daily, Certo, & Roengpitya, 2003). When possible, we conduct hierarchical analyses to disentangle subgroups of one moderator variable in one already disentangled subgroup of another moderator variable (Schmidt & Hunter, 2015).

3.3. Ex ante analyses of potential common method bias and endogeneity issues

Because we aggregate survey-based data, we assure that common method bias does not drive our findings. Higher-quality studies employ procedural and/or statistical remedies to limit common method variance (Podsakoff, MacKenzie, Lee, & Podsakoff, 2003). For the initial dataset of 69 independent samples, we thus investigate whether the relationships differ, depending on whether the studies use remedies to common method bias. Our analyses (Table 3) show that whether a study addresses common method bias is a significant moderator in nine out of the 11 focal relationships for which this test is possible. As expected, when common method bias is (not) addressed, the mean corrected correlation is significantly lower (higher). Table 3 also displays the inflation of correlations (i.e., the ratio of the mean corrected correlation without addressing common method bias to the mean corrected correlation addressing this bias). These inflations range from 118.11% to 261.85%. This range is slightly lower than that reported by Podsakoff, MacKenzie, and Podsakoff (2012) for their comparison of effect sizes from single-source vs. multi-source organizational behavior studies. This indicates that common method bias is (at least) not more serious in LoC research than in other social science research. Because we intend to uncover stylized facts for research and practice, we want to avoid a potential type I error and, thus, exclude from our further analyses the 11 samples that do not address common method biases at all.[17] Hence, we base our following analyses on 58 independent samples and a harmonic mean sample size of $n = 2458$.[18]

Moreover, endogeneity might be a concern, which we thus address in three ways. (1) Measurement error is a major source of endogeneity, but, as described above, we correct all correlations for differences in reliability. (2) To limit the threat of omitting correlated variables (Merchant & Otley, 2020; Speklé & Widener, 2020), we account for environmental uncertainty and size as joint determinants of all variables in our models.[19] These context variables are the most important antecedents to the levers, capabilities, and performance (Simons, 1995) and, thus, most frequently investigated in prior research (Chenhall, 2003; Kruis, Speklé, & Widener, 2016). Moreover, our hypotheses and models do not only relate the levers to performance but also include capabilities as a mediating variable. Our models thus account for a nuanced theory-based relationship, which further limits the likelihood of omitting correlated variables (Bergh et al., 2016). (3) Although most primary studies do not include time lags, we partially account for simultaneity as a further source of endogeneity. Following Bergh et al. (2016), for each of our two models, we estimate three competing ones that model the levers as consequences rather than as

[16] For the purpose of identification, we constrain the mutual paths among each pair of the levers to be equal. This model allows for feedback loops (Kline, 2011) and, thus, for indirect effects of the levers through capabilities on performance via the other levers. Given that we include environmental uncertainty and size as crucial antecedents of the levers (Chenhall, 2003; Kruis et al., 2016), we base our theorizing on the likely assumption that the mutual paths capture substantial bivariate relationships rather than omitted joint determinants.

[17] We mark these samples with asterisks in Web Appendix 1. The remaining samples use different remedies for common method bias (see Web Appendix 2). Following the classification by Podsakoff et al. (2003), the majority employs procedural *and* statistical remedies. We test whether the results of studies using only procedural remedies and those employing statistical remedies differ significantly, but we do not find significant results (untabulated). We thus do not exclude further studies from our analyses. The most frequent statistical remedy is Harman's single-factor test. Few studies use advanced statistical remedies, such as marker variables or other partial correlation procedures. This impedes a more rigorous test that compares the effects with and without partialling out surrogates of method variance. Of the 58 samples that address potential common method bias, 35 (60%) use statistical remedies. Of these 35 samples, 34 (97%) apply Harman's single-factor test; four (11%), partial correlation procedures; five (14%), an unmeasured latent methods factor; and one (3%), multiple methods factors (multiple methods per sample are possible).

[18] To explore how addressing common method bias (and, thus, excluding or including the 11 samples that do not address common method biases at all) affects the results of this meta-analysis, we ran both the meta-analytic path models and the meta-analytic moderator analyses with and without those 11 samples (see Tables 4 and 5 as well as Web Appendix 5). Compared with the version excluding the 11 samples, in the version including the 11 samples, as expected, a few findings change. That is, overall, the mean corrected correlations in the bivariate and moderator analyses as well as the path coefficients are (mostly) larger if the 11 samples are included. This reinforces our decision to drop the 11 samples from further analyses.

[19] Instead of using them as control variables, in line with the contingency tradition of MCS research, environmental uncertainty and size could also be used as moderator variables in our path models. Due to data restrictions, however, we cannot use this approach. Statistically, it would require either subgroup analyses or interaction terms. First, for subgroup analyses, we would require at least three independent samples for each single effect in the models per subgroup (e.g., small vs. large organizations). Most studies, however, use mixed samples (e.g., including both small and large organizations) and cover only some of the variables in our models, thus preventing us from using this option. Second, interaction terms can be included in meta-analyses (which are based on primary studies' correlations and not on raw data) only if the primary studies (1) include such interaction terms and (2) report the correlations between them and all other variables. Unfortunately, this is not common in LoC studies. In the same vein, strategy might constitute another potential moderator variable. Unfortunately, studies do not consider the same conceptualization of strategy sufficiently often, thus preventing us from using this option.

Table 3
Analysis of common method bias.

Relationships	*k*	*N*	*r*	SD_r	*ρ*	SD_ρ	95% CI lower	95% CI upper	95% CrI lower	95% CrI upper	% Var. unacc.	95% CIdiff comp.	95% CIdiff lower	95% CIdiff upper	% Inflation
Beliefs control – boundary control	**20**	**4,218**	**0.501**	**0.172**	**0.564**	**0.183**	**0.480**	**0.648**	**0.205**	**0.923**	**90.87**				
CMB not addressed	3	222	0.704	0.065	0.791	0.055	0.694	0.889	0.683	0.899	40.71	(a) - (b)	0.107	0.373	143.56
CMB addressed	17	3,996	0.489	0.169	0.551	0.180	0.462	0.641	0.198	0.904	91.26				
Beliefs control – diagnostic control	**20**	**4,235**	**0.403**	**0.147**	**0.448**	**0.155**	**0.375**	**0.522**	**0.144**	**0.753**	**85.51**				
CMB not addressed	3	222	0.621	0.079	0.686	0.098	0.543	0.828	0.494	0.877	59.97	(a) - (b)	0.088	0.412	157.34
CMB addressed	17	4,013	0.391	0.141	0.436	0.148	0.360	0.512	0.146	0.725	85.34				
Beliefs control – interactive control	**19**	**3,997**	**0.420**	**0.128**	**0.474**	**0.143**	**0.403**	**0.544**	**0.194**	**0.754**	**83.15**				
CMB not addressed	3	222	0.655	0.008	0.714	0.059	0.608	0.821	0.599	0.829	39.05	(a) - (b)	0.128	0.384	155.90
CMB addressed	16	3,775	0.406	0.118	0.458	0.134	0.386	0.530	0.196	0.721	82.58				
Beliefs control – performance	**12**	**2,482**	**0.294**	**0.115**	**0.327**	**0.135**	**0.241**	**0.413**	**0.063**	**0.591**	**78.06**				
CMB not addressed	3	301	0.480	0.213	0.555	0.221	0.287	0.824	0.123	0.988	86.28	(a) - (b)	-0.019	0.535	186.87
CMB addressed	9	2,181	0.269	0.060	0.297	0.081	0.229	0.365	0.139	0.455	59.29				
Boundary control – diagnostic control	**19**	**3,997**	**0.369**	**0.117**	**0.415**	**0.133**	**0.348**	**0.482**	**0.155**	**0.676**	**79.52**				
CMB not addressed	3	222	0.579	0.000	0.655	0.000	0.611	0.700	0.655	0.655	0.00	(a) - (b)	0.172	0.336	163.34
CMB addressed	16	3,775	0.356	0.111	0.401	0.125	0.332	0.470	0.155	0.647	79.24				
Boundary control – interactive control	**19**	**3,997**	**0.351**	**0.193**	**0.400**	**0.225**	**0.294**	**0.506**	**-0.041**	**0.841**	**91.26**				
CMB not addressed	3	222	0.620	0.000	0.693	0.000	0.661	0.724	0.693	0.693	0.00	(a) - (b)	0.194	0.430	181.89
CMB addressed	16	3,775	0.335	0.188	0.381	0.221	0.268	0.494	-0.051	0.814	91.67				
Diagnostic control – interactive control	**50**	**8,836**	**0.561**	**0.190**	**0.646**	**0.209**	**0.584**	**0.704**	**0.234**	**1.054**	**92.60**				
CMB not addressed	8	729	0.624	0.155	0.750	0.139	0.640	0.860	0.478	1.022	76.65	(a) - (b)	-0.013	0.243	118.11
CMB addressed	42	8,107	0.556	0.192	0.635	0.211	0.568	0.701	0.220	1.049	93.26				
Diagnostic control – capabilities	**17**	**3,359**	**0.378**	**0.193**	**0.453**	**0.206**	**0.350**	**0.557**	**0.049**	**0.858**	**89.19**				
CMB not addressed	3	207	0.567	0.050	0.633	0.037	0.521	0.745	0.561	0.706	13.92	(a) - (b)	0.031	0.353	143.54
CMB addressed	14	3,152	0.366	0.193	0.441	0.208	0.327	0.556	0.035	0.848	90.25				
Diagnostic control – performance	**28**	**5,504**	**0.244**	**0.141**	**0.285**	**0.165**	**0.217**	**0.352**	**-0.039**	**0.608**	**81.46**				
CMB not addressed	5	532	0.527	0.124	0.652	0.100	0.536	0.767	0.456	0.847	57.60	(a) - (b)	0.273	0.533	261.85
CMB addressed	23	4,972	0.213	0.105	0.249	0.121	0.190	0.307	0.011	0.487	71.92				
Interactive control – capabilities	**20**	**3,564**	**0.428**	**0.181**	**0.507**	**0.186**	**0.420**	**0.594**	**0.142**	**0.872**	**87.14**				
CMB not addressed	3	207	0.636	0.018	0.696	0.000	0.608	0.784	0.696	0.696	0.00	(a) - (b)	0.074	0.332	140.89
CMB addressed	17	3,357	0.415	0.179	0.494	0.186	0.399	0.588	0.129	0.859	87.92				
Interactive control – performance	**38**	**6,770**	**0.224**	**0.136**	**0.261**	**0.156**	**0.205**	**0.317**	**-0.044**	**0.566**	**77.92**				
CMB not addressed	7	735	0.418	0.162	0.514	0.151	0.381	0.646	0.219	0.809	70.81	(a) - (b)	0.139	0.425	234.70
CMB addressed	31	6,035	0.200	0.112	0.232	0.128	0.179	0.285	-0.019	0.483	72.05				

Note: Lines in bold report meta-analytic results for the respective full samples, whereas lines in normal font report meta-analyses for moderator variable subgroups. Gray shaded lines report the significant moderator effects of addressing common method bias. CMB: common method bias; *k:* number of correlation coefficients per relationship; *N:* total sample size across *k* samples; *r*: weighted mean observed correlation; SD_r: standard deviation of *r*; *ρ*: estimated weighted mean correlation corrected for artefacts; SD_ρ: standard deviation for the estimated *ρ*; 95% CI: lower and upper bounds of the confidence interval for *ρ*; 95% CrI: lower and upper bounds of the credibility interval for each meta-analysis distribution; % Var. unacc.: percentage of unexplained variance in correlations; 95% CIdiff.: lower and upper bounds of the confidence interval of the difference between compared (comp.) subgroup *ρ*; % Inflation: percentual inflation of the estimated weighted mean correlation corrected for artefacts, measured as the ratio between the correlation without addressing common method bias and the correlation addressing common method bias.

antecedents of capabilities and/or performance.[20] None of these (untabulated) models fits the data better and more parsimoniously than do our three models. Altogether, although we obviously cannot fully rule out endogeneity concerns, we provide evidence that endogeneity is not a major issue in this study.

4. Results

4.1. Meta-analytic path model results

The bivariate meta-analytic correlations show significant positive relationships among the four levers and between the levers and capabilities and performance (Web Appendix 6). Environmental uncertainty and size are mostly significantly positively related to the levers. These meta-analytic correlations are the basis for estimating our models. In the following, before discussing Models 1 and 2 in detail, we compare them with a baseline model that assumes an isolated use of the levers to establish whether they are suitable for further investigation.

4.1.1. Isolated vs. combined use of the levers

To establish the credibility of our hypothesized models and, thus, to provide evidence for Simons' (1995) theoretical claim that the levers' power "does not lie in how each is used alone" (p. 153) but in their "continual interplay" (p. 30), we compare the baseline model with Models 1 and 2. The respective AIC and BIC are considerably lower for Models 1 and 2, thus displaying better fit and parsimony such that both models outperform the baseline model in describing the data.[21] Compared with the baseline model, for Models 1 and 2, we also find higher total effects and higher R^2 of performance. This supports Simons' theoretical reasoning. In contrast to the isolated use, the combined use of the levers better describes how organizations in the samples actually use the levers to enhance performance.

4.1.2. Broad theory perspective

We find support for our Model 1 (Table 4, Panel A). Although the fit indexes deviate slightly from the common threshold values (Kline, 2011), our model shows reasonable fit with the data. The simultaneous use of the levers has a significant positive total effect on performance ($\beta = 0.337$, $p < 0.01$). This is partially rooted in a significant positive indirect effect through capabilities ($\beta = 0.109$, $p < 0.01$), thus supporting H1. This indirect effect emanates from significant positive relationships between the simultaneous use of the levers and capabilities ($\beta = 0.627$, $p < 0.01$) and between capabilities and performance ($\beta = 0.174$, $p < 0.01$). We also find a significant positive direct effect of the simultaneous use of the levers on performance ($\beta = 0.227$, $p < 0.01$), which indicates complementary mediation (Zhao et al., 2010).

Therefore, although we find support for our RBV-based argument, complementary explanations for the relationship between the simultaneous use of the levers and performance also are necessary. Indeed, the indirect-to-total effect ratio is 32.34% (Iacobucci, Saldanha, & Deng, 2007), indicating that the indirect effect via capabilities does not explain almost 70% of the total effect.[22] Unfortunately, prior LoC research has not yet intensively explored such potential mechanisms.

4.1.3. Specific theory perspective

We also find support for Model 2 (Table 4, Panel B). The fit indexes display (artifactual) perfect fit to the data (Kline, 2011).[23] The model shows significant positive total effects of the four levers on performance (beliefs: $\beta = 0.292$; boundary: $\beta = 0.067$; diagnostic: $\beta = 0.196$; interactive: $\beta = 0.172$; all $p < 0.01$). We find positive indirect effects via the other levers and capabilities (beliefs: $\beta = 0.102$; boundary: $\beta = 0.123$; diagnostic: $\beta = 0.119$; interactive: $\beta = 0.158$; all $p < 0.01$), which support H2. In addition to the significant positive mutual paths among the levers, these effects are composed of significant positive direct levers–capabilities effects (beliefs: $\beta = 0.203$; boundary: $\beta = 0.093$; diagnostic: $\beta = 0.145$; interactive: $\beta = 0.262$; all $p < 0.01$) and a significant positive capabilities–performance effect ($\beta = 0.211$; $p < 0.01$), resulting in (untabulated) specific indirect levers–capabilities–performance effects (beliefs: $\beta = 0.043$; boundary: $\beta = 0.020$; diagnostic: $\beta = 0.031$; interactive: $\beta = 0.055$).

Interestingly, the direct interactive control–performance effect is insignificant ($\beta = 0.014$, $p > 0.05$; indirect-to-total effect ratio: 91.86%), indicating indirect-only mediation (Zhao et al., 2010). We thus see great theoretical clarity for this relationship because it is fully explained by the RBV-based mediation via capabilities and the mutual relationships among the levers. This implies that either additional theories are not needed to explain this relationship or the related effects mutually offset each other.

The direct effects of beliefs ($\beta = 0.190$; $p < 0.01$) and diagnostic control ($\beta = 0.077$, $p < 0.01$) on performance are positive and significant. For both relationships, this finding implies complementary mediation and that our RBV-based model does not capture at least one additional positive indirect effect. The indirect-to-total effect ratios for beliefs and diagnostic control are 34.93% and 60.71%, respectively. Thus, for beliefs and diagnostic control, a considerable part of the total performance effects is not explained by the indirect effects via capabilities. The finding for diagnostic control is notable because this control typically allows single-loop learning and, thus, might be perceived as less important than the other levers for developing capabilities. The large indirect-to-total effect ratio, however, underscores this lever's vital role for enhancing capabilities.

Finally, we find a significant negative direct boundary control–performance effect ($\beta = -0.055$, $p < 0.05$; indirect-to-total effect ratio: 183.58%). This indicates competitive mediation. Thus, our RBV-based model does not capture at least one additional negative indirect effect.

Model 2 better describes the use of the levers in the organizations in our samples than does Model 1. As previously indicated,

[20] Specifically, we estimate the following competing models: (1) capabilities-levers-performance; (2) performance-capabilities-levers; (3) performance-levers-capabilities. In these models, we include environmental uncertainty and size as antecedents of the levers, capabilities, and performance.

[21] Moreover, Model 2 and the baseline model are nested, allowing for a chi square difference test (Δchi square = 2895.636, $df = 6$, $p < 0.001$). This test indicates that Model 2 fits the data better than does the baseline model. For the details of our baseline model, see Web Appendix 7.

[22] We cannot rule out the possibility that other capabilities than those included in the capabilities construct are important, such that the RBV completely explains (together with the mutual relationships among the levers) the levers-performance relationships. Because the included capabilities (Table 2) "are recognized as primary capabilities to reach competitive advantage" (Henri, 2006, p. 532), we deem it more reasonable that variables from other theories might be at play. Moreover, statistically, the direct levers-performance effects that remain after accounting for the mediating effect of capabilities also may be artefacts of joint determinants of the levers and performance. For example, because the available primary studies do not offer the necessary correlations, we cannot include strategy as an antecedent. The inclusion of environmental uncertainty and size as prominent joint antecedents of the levers, capabilities, and performance, however, considerably limits the portion of the direct effect attributable to other antecedents, also, because environmental uncertainty and size are related with strategy (Chenhall, 2003; Langfield-Smith, 2007). In sum, this underscores the need for complementary or competitive theories.

[23] This finding is common in meta-analytic path models, which generally have a limited number of degrees of freedom (Karna et al., 2016).

Table 4
Meta-analytic path model results.

Panel A: Broad theory model: Simultaneous use of the levers (Model 1)			
Paths	**Estimates**	**95% CI**	
Total effect:			
Simultaneous use of the levers → Performance	0.337**	0.294	0.377
Indirect effect (mediation of capabilities):			
Simultaneous use of the levers → Performance	0.109**	0.076	0.140
Direct effects:			
Simultaneous use of the levers → Capabilities	0.627**	0.596	0.659
Simultaneous use of the levers → Performance	0.227**	0.172	0.283
Capabilities → Performance	0.174**	0.121	0.220
Measurement model:			
Beliefs control ← Simultaneous use of the levers	0.665**	0.634	0.698
Boundary control ← Simultaneous use of the levers	0.588**	0.551	0.623
Diagnostic control ← Simultaneous use of the levers[a]	0.739**	0.710	0.763
Interactive control ← Simultaneous use of the levers	0.764**	0.737	0.789
Context variables' effects (controlled for):			
Environmental uncertainty ↔ Size[b]	0.019	–	–
Environmental uncertainty → Simultaneous use of the levers	0.119**	0.078	0.163
Environmental uncertainty → Capabilities	0.051**	0.018	0.085
Environmental uncertainty → Performance	−0.136**	−0.173	−0.099
Size → Simultaneous use of the levers	0.159**	0.117	0.205
Size → Capabilities	−0.002	−0.038	0.030
Size → Performance	0.101**	0.066	0.140
Model fit:			
Chi square	425.467**		
df	15		
CFI	0.909		
RMSEA	0.106		
SRMR	0.041		
AIC	467.467		
BIC	589.416		
Explained variance in dependent variables—R^2:			
Simultaneous use of the levers	0.040		
Beliefs control	0.443		
Boundary control	0.346		
Diagnostic control	0.546		
Interactive control	0.584		
Capabilities	0.403		
Performance	0.158		

Panel B: Specific theory model: Mutually related levers (Model 2)			
Path	**Estimates**	**95% CI**	
Total effects:			
Beliefs control → Performance	0.292**	0.246	0.337
Boundary control → Performance	0.067**	0.025	0.114
Diagnostic control → Performance	0.196**	0.146	0.242
Interactive control → Performance	0.172**	0.123	0.221
Indirect effects (mediation of the other levers and capabilities):			
Beliefs control → Performance	0.102**	0.084	0.122
Boundary control → Performance	0.123**	0.107	0.141
Diagnostic control → Performance	0.119**	0.098	0.139
Interactive control → Performance	0.158**	0.136	0.181
Direct effects:			
Beliefs control ⇄ Boundary control[c]	0.245**	0.224	0.265
Beliefs control ⇄ Diagnostic control[c]	0.100**	0.081	0.119
Beliefs control ⇄ Interactive control[c]	0.134**	0.115	0.153
Boundary control ⇄ Diagnostic control[c]	0.106**	0.087	0.126
Boundary control ⇄ Interactive control[c]	0.064**	0.044	0.083
Diagnostic control ⇄ Interactive control[c]	0.302**	0.282	0.322
Beliefs control → Capabilities	0.203**	0.158	0.244
Boundary control → Capabilities	0.093**	0.053	0.132
Diagnostic control → Capabilities	0.145**	0.100	0.185
Interactive control → Capabilities	0.262**	0.219	0.304
Beliefs control → Performance	0.190**	0.140	0.237
Boundary control → Performance	−0.055*	−0.098	−0.009
Diagnostic control → Performance	0.077**	0.030	0.124
Interactive control → Performance	0.014	−0.039	0.062
Capabilities → Performance	0.211**	0.168	0.255
Context variables' effects (controlled for):			
Environmental uncertainty ↔ Size[b]	0.019	–	–
Environmental uncertainty → Beliefs control	0.011	−0.020	0.044
Environmental uncertainty → Boundary control	0.099**	0.064	0.135
Environmental uncertainty → Diagnostic control	−0.027	−0.058	0.003
Environmental uncertainty → Interactive control	0.101**	0.069	0.130

(*continued on next page*)

Table 4 (*continued*)

Environmental uncertainty → Capabilities	0.064	0.030	0.099
Environmental uncertainty → Performance	−0.123**	−0.158	−0.084
Size → Beliefs control	0.060**	0.026	0.092
Size → Boundary control	0.019	−0.015	0.049
Size → Diagnostic control	0.063**	0.033	0.094
Size → Interactive control	0.078**	0.048	0.109
Size → Capabilities	0.018	−0.014	0.051
Size → Performance	0.106**	0.070	0.144
Model fit:			
Chi square	0.000		
df	1		
CFI	1.000		
RMSEA	0.000		
SRMR	0.000		
AIC	70.000		
BIC	273.249		
Explained variance in dependent variables—R^2:			
Beliefs control	0.334		
Boundary control	0.293		
Diagnostic control	0.384		
Interactive control	0.398		
Capabilities	0.327		
Performance	0.163		

Notes for both Panels:
*p < 0.05, **p < 0.01 (one-tailed for hypothesized paths, two-tailed for all other paths).
n (harmonic mean) = 2458.
Standardized coefficient estimates reported, based on maximum likelihood estimation using IBM SPSS Amos 28. We compute standard errors and significance levels of the total and indirect effects and all confidence intervals based on 2000 bootstrap samples (parametric Monte Carlo bootstrap).
Thresholds for the approximate fit indexes: CFI > 0.95 (Hu & Bentler, 1999); RMSEA < 0.08 (Browne & Cudeck, 1993); SRMR < 0.08 (Hu & Bentler, 1999).
Predictive fit indexes (i.e., AIC and BIC) are used to compare alternative nonhierarchical models. Lower index values indicate better model fit and higher parsimony (Kline, 2011).
[a] parameter fixed to one for model identification.
[b] parameter fixed to the estimated weighted mean correlation corrected for artefacts.
[c] mutual paths between variables constrained to be equal.

both the broad and specific theory perspectives have their merits. This can be seen empirically by the good fit to the data of both models.[24] It is, however, possible to compare Models 1 and 2 to better understand which of the two types of the combined use of the levers better represents their predominant use in the organizations included in our primary studies. For this comparison, we again use the AIC and BIC indexes and find that Model 2 outperforms Model 1. Compared with the simultaneous use of the levers, a use that favors a more or less strong coupling of the emphases on the respective levers better describes the actual use of the levers in the organizations in our samples. Models 1 and 2, however, yield a similar R^2 for performance (0.158 and 0.163, respectively). Due to the conceptual difference between the broad and the specific theory perspectives (see Section 2.2), we believe that, on theoretical grounds, it is not appropriate to a priori favor one of these models.

4.2. Moderator analysis of conceptual choices

Table 5 presents the *significant* moderator effects uncovered in the non-hierarchical and hierarchical moderator analyses. Web Appendixes 9 and 10 provide the full non-hierarchical and hierarchical analyses, respectively, and include both *significant* and *insignificant* effects. In the following, we consider every significant between-subgroup difference as falsifying the null hypothesis that no moderation effect exists. In contrast, when our data do not provide a single indication for an effect of a potential moderator variable, we stipulate that our findings are robust to this variable's influence.

The focal relationships appear to be quite robust (for related insignificant moderator effects, see Web Appendixes 9 and 10). First, for the different dimensions of interactive control, in none of the five tests do we find a significant difference. This is surprising, as previous research (Bisbe et al., 2007; Tessier & Otley, 2012) provides an extensive discussion of the heterogeneity of the dimensions of interactive control. We distinguish approaches that (1) include only the intensity of use dimension, (2) combine the dimensions of intensity of use and enabling role, and (3) combine the dimensions of intensity of use, enabling role, and strategic validity. As a common core, these approaches share the intensity of use dimension, or the “attention pattern” for the control practices (Tessier & Otley, 2012, p. 177). Our findings show that the surplus meaning covered by the enabling role and strategic validity dimensions does not give rise to systematically different relationships with other variables. Therefore, these three variants appear to tap the same construct domain and likely belong to the same nomological network. For example, a more intensive use of control practices may facilitate frequent exchanges and discussions that likely entail the perception of control practices as enabling and facilitating (Mahama & Cheng, 2013). Similarly, strategic validation requires an intensive use of performance measures or risk management tools (Tuomela, 2005) to continuously monitor the environment for strategic uncertainties, such as threats to the competitive position.

Second, for the conceptualization of interactive control, according to Henri (2006), Widener (2007), or other approaches, in the four possible tests, we do not find any significant difference. In line with our arguments above in regard to the constructs' dimensions, we conclude that the construct domain does not change systematically across different conceptualizations.

Third, in none of nine tests do we find that the distinction

[24] As mentioned above, Model 2 displays an artifactually perfect fit to the data due to the low number of degrees of freedom. To alleviate this concern, we stepwise trim insignificant non-hypothesized paths from our models, thus increasing the degrees of freedom and improving the assessment of model fit. Our inferences about fit and model comparison hold when considering the trimmed models, which we document in Web Appendix 8.

Table 5
Meta-analytic subgroup analyses: Significant moderator effects.

Relationships	k	N	r	SD_r	ρ	SD_ρ	95% CI lower	95% CI upper	95% CrI lower	95% CrI upper	% Var. unacc.	95% CIdiff comp.	95% CIdiff lower	95% CIdiff upper
Conceptualization of diagnostic control														
Beliefs control - diagnostic control	**17**	**4013**	**0.391**	**0.141**	**0.436**	**0.148**	**0.360**	**0.512**	**0.146**	**0.725**	**85.34**			
Conceptualization Henri	8	1970	0.329	0.030	0.374	0.000	0.333	0.416	0.374	0.374	0.00	(a) - (b)	−0.295	−0.145
Conceptualization Widener	3	963	0.556	0.028	0.594	0.036	0.532	0.656	0.523	0.664	43.07	(b) - (c)	0.048	0.452
Conceptualization other	5	924	0.313	0.183	0.344	0.207	0.152	0.537	−0.061	0.750	88.72	(a) - (c)	−0.166	0.226
Diagnostic control − capabilities	**14**	**3152**	**0.366**	**0.193**	**0.441**	**0.208**	**0.327**	**0.556**	**0.035**	**0.848**	**90.25**			
Conceptualization Henri	5	1281	0.236	0.091	0.302	0.094	0.197	0.406	0.118	0.485	61.54	(a) - (b)	−0.409	−0.027
Conceptualization other	8	1749	0.452	0.202	0.520	0.223	0.360	0.680	0.083	0.956	92.88			
Management control practice														
Diagnostic control − performance	**23**	**4972**	**0.213**	**0.105**	**0.249**	**0.121**	**0.190**	**0.307**	**0.011**	**0.487**	**71.92**			
Performance measurement	12	2921	0.204	0.112	0.240	0.125	0.158	0.322	−0.006	0.486	75.35	(a) - (b)	0.040	0.244
Budgeting	3	430	0.074	0.000	0.098	0.000	0.038	0.158	0.098	0.098	0.00			
Interactive control − performance	**31**	**6035**	**0.200**	**0.112**	**0.232**	**0.128**	**0.179**	**0.285**	**−0.019**	**0.483**	**72.05**			
Conceptualization other (performance measurement)	5	1002	0.213	0.055	0.242	0.060	0.157	0.327	0.125	0.359	37.67	(a) - (b)	0.106	0.418
Conceptualization other (budgeting)	3	264	−0.020	0.000	−0.020	0.000	−0.150	0.111	−0.020	−0.020	0.00			
General vs. specific control problem														
Boundary control - interactive control	**16**	**3775**	**0.335**	**0.188**	**0.381**	**0.221**	**0.268**	**0.494**	**−0.051**	**0.814**	**91.67**			
General control problem	10	2486	0.271	0.173	0.302	0.203	0.170	0.435	−0.095	0.700	90.23	(a) - (b)	−0.436	−0.060
Specific control problem	6	1289	0.458	0.152	0.550	0.154	0.417	0.683	0.249	0.851	85.21			
Intention for vs. perception of control														
Diagnostic control - interactive control	**42**	**8107**	**0.556**	**0.192**	**0.635**	**0.211**	**0.568**	**0.701**	**0.220**	**1.049**	**93.26**			
Conceptualization other (interactive) (intention)	20	3371	0.530	0.255	0.605	0.283	0.478	0.733	0.051	1.160	95.25	(a) - (b)	−0.289	−0.017
Conceptualization other (interactive) (perception)	3	535	0.694	0.000	0.758	0.007	0.710	0.807	0.745	0.772	2.38			
Conceptualization of capabilities														
Diagnostic control − capabilities	**14**	**3152**	**0.366**	**0.193**	**0.441**	**0.208**	**0.327**	**0.556**	**0.035**	**0.848**	**90.25**			
Innovativeness	9	2200	0.260	0.126	0.329	0.142	0.224	0.422	0.050	0.608	78.64	(a) - (b)	−0.292	0.130
Organizational learning	4	733	0.339	0.159	0.410	0.169	0.228	0.592	0.079	0.741	82.51	(b) - (c)	−0.431	0.113
Other capabilities	5	1146	0.495	0.210	0.569	0.224	0.366	0.772	0.130	1.008	93.92	(a) - (c)	−0.468	−0.012
Conceptualization of performance														
Diagnostic control − performance	**23**	**4972**	**0.213**	**0.105**	**0.249**	**0.121**	**0.190**	**0.307**	**0.011**	**0.487**	**71.92**			
Financial performance	10	2619	0.165	0.063	0.193	0.073	0.130	0.256	0.049	0.336	52.04	(a) - (b)	−0.227	0.111
Non-financial performance	5	946	0.198	0.136	0.251	0.156	0.094	0.407	−0.055	0.557	76.65	(b) - (c)	−0.211	0.129
General performance	14	2766	0.249	0.089	0.292	0.101	0.225	0.359	0.093	0.490	62.77	(a) - (c)	−0.191	−0.007
Interactive control − performance	**31**	**6035**	**0.200**	**0.112**	**0.232**	**0.128**	**0.179**	**0.285**	**−0.019**	**0.483**	**72.05**			
Financial performance	12	2811	0.141	0.092	0.158	0.104	0.086	0.229	−0.045	0.361	67.35	(a) - (b)	−0.278	−0.042
Non-financial performance	6	1078	0.258	0.084	0.318	0.081	0.224	0.411	0.159	0.476	47.84	(b) - (c)	−0.083	0.165
General performance	20	3502	0.236	0.087	0.277	0.098	0.197	0.357	0.084	0.470	58.07	(a) - (c)	−0.226	−0.012
Capabilities − performance	**14**	**3171**	**0.264**	**0.089**	**0.311**	**0.099**	**0.247**	**0.376**	**0.116**	**0.505**	**65.22**			
Financial performance	6	1645	0.164	0.000	0.193	0.000	0.138	0.247	0.193	0.193	0.00	(a) - (b)	−0.361	−0.143
General performance	8	1737	0.382	0.105	0.445	0.119	0.350	0.539	0.212	0.678	75.56			
Manufacturing vs. service industry samples														
Diagnostic control − performance	**23**	**4972**	**0.213**	**0.105**	**0.249**	**0.121**	**0.190**	**0.307**	**0.011**	**0.487**	**71.92**			
Manufacturing	8	1658	0.225	0.109	0.266	0.125	0.164	0.368	0.021	0.511	72.07	(a) - (b)	0.016	0.238
Service	3	459	0.126	0.000	0.139	0.000	0.096	0.182	0.139	0.139	0.00			
Interactive control − performance	**31**	**6035**	**0.200**	**0.112**	**0.232**	**0.128**	**0.179**	**0.285**	**−0.019**	**0.483**	**72.05**			
Conceptualization intensity (manufacturing)	3	241	0.025	0.000	0.027	0.000	−0.094	0.148	0.027	0.027	0.00	(a) - (b)	−0.547	−0.021
Conceptualization intensity (service)	3	308	0.248	0.176	0.311	0.174	0.079	0.544	−0.030	0.653	71.82			

Note: Lines in bold report meta-analytic results for the respective full samples, whereas lines in normal font report meta-analyses for moderator variable subgroups. *k*: number of correlation coefficients per relationship; *N*: total sample size across *k* samples; *r*: weighted mean observed correlation; *SDr*: standard deviation of *r*; *ρ*: estimated weighted mean correlation corrected for artefacts; *SDρ*: standard deviation for the estimated *ρ*; 95% CI: lower and upper bounds of the confidence interval for *ρ*; 95% CrI: lower and upper bounds of the credibility interval for each meta-analysis distribution; % Var. unacc.: percentage of unexplained variance in correlations; 95% CIdiff: lower and upper bounds of the confidence interval of the difference between compared (comp.) subgroup *ρ*.

between managers' hierarchical levels exerts any influence. Therefore, by extending Simons' (1995) original reasoning, it is possible for researchers to apply the LoC framework in the same manner not only to senior management but also to managers at other hierarchical levels in the organization.

Although these three conceptual choices do not affect the generalizability of the focal relationships, other choices warrant attention in future research because they significantly moderate some of the focal relationships. The conceptualization of diagnostic control moderates the relationships in two out of five tests. First, the beliefs control—diagnostic control relationship is significantly larger for Widener's (2007) conceptualization than for Henri's (2006) and the other approaches. Second, the diagnostic control—capabilities relationship is significantly larger for other conceptualizations than for Henri's approach. These findings suggest that Widener's conceptualization likely covers a broader conceptual domain than does Henri's and that of other approaches, also implying differences in the respective nomological networks. Compared with Henri's focus on results monitoring, Widener's conceptualization also includes items in regard to the challenge to and debate about the underlying assumptions of control practices and the development of a common language in the organization—akin to the core values of the organization (and, herewith, its beliefs control). This may drive the larger beliefs control—diagnostic control relationship and ultimately make diagnostic control more powerful in its relationship with capabilities.

The distinction between performance measurement and budgeting as management control practices used to conceptualize diagnostic and interactive control is significant in two out of nine tests. First, the diagnostic control–performance relationship is larger for the diagnostic use of performance measurement than for budgeting. Second, when interactive control is conceptualized with approaches other than Henri's (2006) or Widener's (2007), we find a significant positive interactive control–performance relationship for the use of performance measurement but an insignificant one for budgeting. Overall, compared with budgeting, performance measurement systems also capture leading indicators, such as customer satisfaction, and, thus, allow organizations to better cope with strategic uncertainties. This, in turn, enables them to generate meaningful competitive advantage, thus having a larger impact on performance than the use of budgets, which merely focus on lagging indicators.

Whether the levers address a general or a specific control problem is a significant moderator in one out of ten tests. Compared with its application to general control problems, the boundary control–interactive control relationship is larger when a specific control problem is addressed. For specific control problems, such as safeguarding sustainable product development (Bellora-Bienengräber, 2019), changes in one of these two levers more likely translate into changes in the other levers, likely a manifestation of a focused effort to address a specific, pressing issue by concomitantly emphasizing the two levers.

The distinction between the conceptualization of the levers as the intention for versus the perception of control is a significant moderator in one out of eight tests. If conceptualizations other than Henri's (2006) or Widener's (2007) capture interactive control, the diagnostic control–interactive control relationship is larger for subordinates' perception of control than for superiors' intentions. Subordinates appear to perceive their superiors' emphases on interactive and diagnostic control as more strongly related than it is intended by superiors, thus potentially increasing subordinates' perceived exposure to control practices and likely entailing increased stress and decreased job satisfaction.

For the conceptualization of capabilities as innovativeness, organizational learning, or other capabilities, one of three tests is significant. The diagnostic control–capabilities relationship is smaller for innovativeness than for other capabilities. Thus, other capabilities, such as market orientation, seem to profit more from the recurring and structured results monitoring inherent in diagnostic control, while interactive control, for which the relationship with capabilities is not moderated by the specific type of capability (see Web Appendix 9), apparently is a powerful lever for all capabilities that an organization may want to develop and leverage.

For the conceptualization of performance with financial, non-financial, or general (i.e., including both financial and non-financial) indicators, three out of five tests show a significant moderating effect. First, the diagnostic control–performance relationship is significantly larger for general than for financial indicators. Second, the interactive control–performance relationship is significantly larger for non-financial and general than for financial indicators. Third, the capabilities–performance relationship is significantly larger for general than for financial indicators. These findings likely follow from the time lag between leading (mainly non-financial) and lagging (mainly financial) indicators. Interactive control is more prone to address strategic aspects that influence leading, non-financial indicators, such as customer satisfaction, whereas diagnostic control is more prone to (also) address short-term, financial indicators, such as return on assets.

Finally, for two out of seven relationships, we find significant industry effects. First, the diagnostic control–performance relationship is significantly larger for manufacturing than for service industry samples. Second, when interactive control is conceptualized as intensity of use, its relationship with performance is significantly larger for service than for manufacturing industry samples. In comparison with the service industries, the clearer measurability and easier availability of critical performance variables in the manufacturing industries appear to be more suitable for effective diagnostic control. Service organizations, for which measuring performance is more challenging due to services' greater intangibility and customer coproduction, seem to be better controlled in a more interactive way, relying on intense and regular exchange between hierarchical levels and face-to-face meetings. As such, industry might partially serve as a proxy for the type of performance measure (i.e., input, process, or output) used. Manufacturing organizations likely rely more strongly on input and process measures that they use diagnostically to ensure quality consistency or employees' safety, whereas service organizations likely rely more strongly on output measures and use them interactively to ensure creative solutions to achieve goals, such as customer satisfaction.

5. Discussion

This study generates important findings for research and practice that help to resolve aspects of the empirical and conceptual ambiguities raised by prior literature and summarized in Table 1. Specifically, we contribute to the literature in three ways. First, we validate theory on the nomological network of the levers and meta-analytically derive stylized facts. Second, we extend theory on the nomological network of the levers by uncovering the need for complementary and competitive theories and by unveiling boundary conditions via our moderator analyses. Third, we contribute to a shift in the perception of survey-based MCS studies—from regarding them as producing idiosyncratic findings resulting from construct reliability and validity issues or small samples to seeing them as part of a robust process in the evolution of science. Finally, we provide robust evidence-based implications for practice. In the following, we interpret the main findings and discuss each contribution in detail.

5.1. Theory validation

Our first contribution is a validation of Simons' reasoning, as we uncover stylized facts for the nomological network of the levers. We show that the combined use of the levers outperforms their isolated use and better describes how organizations use the levers. Following the RBV, the levers have positive total performance effects that are (partially) mediated by capabilities. Specifically, we hypothesized and find support for the argument that the simultaneous use of the levers is associated with increased performance via enhanced capabilities. This supports Simons' (1995) reasoning that control "is achieved by integrating the forces of beliefs control, boundary control, diagnostic control, and interactive control" and that the levers "complement each other when used together" (p. 153). We also hypothesized and find support for the argument that mutually related levers enhance each other as well as performance via increased capabilities. This model mirrors a "continual interplay" (Simons, 1995, p. 30) among the levers and suggests that they "are mutually reinforcing" (p. 161). In contrast to the simultaneous use of the levers, their mutual relations leave room for Simons' (1995) assertion that the levers may be used concomitantly, but for different purposes. Remarkably, and contrary to expectations from prior literature, these findings are robust to influences of different dimensions and conceptualizations of interactive control and hierarchical levels. These stylized facts help to advance research in several ways that we detail below.

First, we provide evidence that neglecting the levers' combined use in research models is not representative of their use in practice and, thus, threatens these models' internal validity. Moreover, our thorough discussion of the conceptual differences among the two models of the levers' combined use and our finding that both models fit the data well empower researchers to consciously choose the model that better fits their level of theoretical abstraction. Researchers interested in the global relationships of the levers' use with other constructs may draw upon a second-order model specification, whereas researchers interested in the levers' distinct relationships with other variables may draw upon a mutual paths model.

Second, the stylized facts may serve MCS researchers as well as reviewers and editors as benchmarks for how knowledge of the LoC framework advances. For example, in line with the RBV, for the interactive control–performance relationship, we find indirect-only mediation via capabilities (and the other levers). For the other levers, however, we find complementary or competitive mediation. Thus, a study that currently replicates Henri's (2006) model and seeks to explain the levers–performance relationship via the RBV and the levers' mutual relationships would be less informative than a study that replicates and supplements Widener's (2007) model, which complements the RBV with a cognitive view of the levers.

Third, our stylized facts support the development of more precise theories. Edwards and Berry (2010) argue that the likelihood of finding significant positive or negative effect sizes increases as methodological rigor increases. Therefore, currently, “developing theories with greater precision, such that their propositions predict something more meaningful than deviations from zero” (Edwards & Berry, 2010, p. 668) is necessary. For example, for exploratory innovation organizations, Bedford (2015) predicts a positive interactive control–performance relationship. Now, this could be refined to predicting a relationship larger than 0.123, the lower bound of the confidence interval for the total effect of the interactive control–performance relationship (Table 4, Panel B). Values between 0.000 and 0.123 should be considered an “indifference zone” that encompasses “values which are essentially equivalent to the null hypothesis” (Binder, 1963, pp. 110–111). With the stylized facts of our meta-analysis at hand, Bedford and Malmi's (2015) actual standardized regression coefficient of 0.323 could, therefore, be considered as theoretically and practically meaningful, in contrast to the (non-hypothesized) 0.106 total effect for the same relationship in Bisbe and Otley (2004).

Fourth, studies that use methods other than surveys also may use our stylized facts. For example, studies based on experiments or analyses of archival data may refer to our effect sizes to anchor their results. Similarly, when using new approaches of data collection such as textual analyses to capture the levers' use, studies may use our effect sizes to triangulate their findings.

Fifth, conceptualizing the levers with a focus on different hierarchical levels or with different dimensions or conceptualizations of interactive control does not yield different results. This finding is particularly notable because these conceptual choices have caused intense debate in the literature but do not materialize in substantial empirical differences. Therefore, researchers can choose either approach without hesitation, based on the conceptualization that they deem most relevant or better understandable in their respective research settings.

Sixth, the stylized facts are important not only for MCS research but also for neighboring disciplines, such as general, strategic, or innovation management. When aiming to explain the antecedents of capabilities or performance and, thus, to develop more complete and practically useful versions of existing theories, researchers may revert to our findings. They also may consider, for example, how the findings of Capon, Farley, and Hoenig's (1990) meta-analysis, which attempts to explain financial performance by environmental, strategic, and organizational variables, would change if a meta-analytic path model would supplement these variables with the levers; or whether the meta-analytical findings on the negative relationship between formalization and market orientation (Kirca, Jayachandran, & Bearden, 2005) or on the insignificant relationship between formalization and organizational innovation (Damanpour, 1991) would change if replications of these two meta-analyses also included the levers in their notion of formalization.

5.2. Theory extension

As a second contribution, we extend theory. In particular, we further clarify the underlying mechanism of the relationships between levers, capabilities, and performance. Moreover, we unravel some of the ambiguities that stem from different conceptual choices.

First, as based on solid empirical evidence, this study is the first to systematically show that, except for the performance effect of interactive control, the RBV logic cannot explain the levers–performance relationships completely. Although the RBV currently is, at least implicitly, widely applied in LoC research, it leaves room for complementary or competitive theories. From a broad theory perspective, for instance, organizational- and individual-level theories rooted in sociology or psychology may suggest mediators that pertain to subordinates' willingness to contribute to organizational goals—mediators that, ceteris paribus, can fruitfully complement the RBV's focus on capabilities. For example, social identity theory (Ashforth & Mael, 1989) can motivate an investigation of whether organizational identification mediates the relationship between the levers' simultaneous use and performance. Moreover, social exchange theory (Homans, 1958) can be used to explain how trust in (senior) management mediates the relationship between the levers' simultaneous use and performance. In a similar vein, self-determination theory (Gagné & Deci, 2005) may suggest to test the satisfaction of employees' need for autonomy as an additional mediator. From a specific theory perspective, for example, to better explain the beliefs control–performance relationship, MCS scholars may refer to the literature on organizations' purpose, identity, or culture to identify additional mediators, such as organizational identification, meaning at work, and ethical work climates (Hollensbe, Wookey, Hickey, & George, 2014; Schneider, Ehrhart, & Macey, 2013). For the boundary control–performance relationship, our RBV-based model does not capture at least one additional negative indirect effect, which might point at time-lagged mediating variables not investigated so far. For example, strong boundaries may imply forgoing opportunities in the short run but also reducing risks and increasing reputation in the longer run. Apart from the time effect, the negative effect may point to undesirable behavioral consequences as another mediating variable and, thus, to a too-much-of-a-good-thing effect (Busse, Mahlendorf, & Bode, 2016). Indeed, a strong emphasis on boundary control might be costly because close adherence to norms and regulations could limit employees' willingness to engage in risky but potentially value-creating activities.

Second, our moderator analyses reveal boundary conditions for the nomological network of the levers. Specifically, we find significant influences of certain choices in regard to the conceptualization of the levers (e.g., diagnostic control, management control practice) and the conceptualization of capabilities and performance as well as industry. These moderating influences change the correlations' strength (not their direction) and, thus, limit the generalizability of the nomological network of the levers. Therefore, they offer ample space for theory refinement. This finding is particularly important if contrasted with the silence about these choices in prior LoC research. Future studies may use our nomological network to advance the understanding of the levers by theoretically and empirically contrasting and extending these conceptual choices. For instance, related to the control problem considered, research could consider how Widener's (2007) findings that beliefs and diagnostic, but not boundary and interactive, control relate to organizational learning change if the levers were not focused on a general control problem, but rather on a specific control problem such as ethics, innovation, or sustainability. One consideration could be whether, for example, boundary control enhances organizational learning if the code of business conduct were focused on ethical behaviors, whereas a focus on restricting certain innovative behaviors could reduce organizational learning. Similarly, related to the industry, research could consider how Widener's (2007) findings would change if theorizing and testing focused only on service industries or on manufacturing industries, instead of a mix of both. For example, one consideration could be whether the relationship between diagnostic control and organizational learning is larger for manufacturing than for service industries. Taking industry differences into account, research could focus on whether distinct nomological networks of the levers exist in different industries or whether only parts of the network change across industries. Even further, within each industry, research could consider how the nomological network of the levers evolves in more digitalized businesses, compared to less digitalized ones.

5.3. Implications for survey-based MCS research

As our third contribution, we make a compelling case for survey-based MCS research. Specifically, our study demonstrates the advantage of individual survey studies as indispensable building blocks of research programs on complex MCS phenomena and, thus, as essential steps in the evolution of our MCS knowledge. Indeed, for LoC research, we show that, after corrections for measurement and sampling error and common method bias as sources of between-study variance, extant survey-based findings, when taken together, are clearer than previously thought. Indeed, they portray a solid and consistent picture of the functioning of MCS processes. At the same time, the diversity of survey studies helps us to understand the boundary conditions for the nomological network of the levers. In general, capturing complex MCS phenomena is demanding in terms of the amount, variety, and depth of the necessary data. Surveys have unique advantages over other methods in satisfying this demand (Speklé & Widener, 2020). Meta-analyses such as this one are then able to capitalize on the often-criticized diversity in conceptual choices by using it to extend existing theories. Thus, we encourage survey researchers to further embrace the investigation of diverse angles of MCS phenomena and meta-analytic researchers to capitalize on the diversity of samples.

5.4. Practical implications

It is our hope that this study will motivate evidence-based management control. In this regard, the findings of our study are meaningful for practice and teaching. Our stylized facts provide solid evidence for managers and future managers (i.e., current students) to use MCSs confidently, in line with Simons' reasoning.

First, although the findings of extant studies appear to be ambiguous and potentially misleading for practice, we find robust positive total performance effects for each lever and for their combined use, in part, mediated by capabilities (and the other levers). Insignificant effects would imply that organizations in the samples have adopted, on average, an optimal level of emphasis on the levers. Thus, our positive total performance effects indicate that organizations in our samples, on average, are below the optimal level of emphasis on the levers (Burkert, Davila, Mehta, & Oyon, 2014; Busse et al., 2016). Therefore, we encourage organizations to invest in the use of the levers. The robustness of the effects across hierarchical levels further indicates that managers at all levels may embrace the levers for their units.

Second, our models suggest that, to benefit from the levers' entire performance potential, organizations should invest more in their combined, instead of isolated, use. The comparison of Models 1 and 2 shows that, so far, organizations in our samples, on average, rely on a more or less loose coupling of the levers but have not yet fully embraced their simultaneous use. Managers may use this finding to refine the way in which they combine the levers' uses.

Third, organizations that wish to increase the use of the levers do not face an easy task. Caution is required because some boundary conditions must be considered. That is, organizations need to be aware of the context in which they apply the levers. For example, managers need to acknowledge differences induced by different control practices (i.e., performance measurement or budgeting), control problems (i.e., specific or general), or the industry environment.

6. Conclusions

In regard to the LoC framework as one of the most influential frameworks in MCS research, this study is the first to meta-analytically identify stylized facts and boundary conditions. We rely on data closer to definitive than had any primary study, as we aggregate 58 samples with 10,374 observations. We summarize the main empirical and conceptual ambiguities in the nomological network of the levers and the respective results from our meta-analysis. First, we find positive relationships among the levers, capabilities, and performance. Moreover, capabilities (partially) mediate the levers–performance relationships. In addition, we show that, compared with models that imply independence of the levers, models that capture their combined use better describe organizations' actual use of the levers. Hence, we clarify the *what*, *how*, and *why* of the nomological network of the levers (Whetten, 1989). Second, we show which ambiguities related to prior studies' conceptual choices are boundary conditions of the nomological network of the levers and, thus, describe its *who*, *where*, and *when* (Whetten, 1989). In particular, different dimensions and conceptualizations of interactive control and hierarchical levels do not condition the relationships' strength. In contrast, their strength, in part, depends on different conceptualizations of diagnostic control, capabilities, and performance, as well as management control practices, general vs. specific control problems, the intention for vs. perception of control, and industry.

As with all research, this study is subject to certain limitations. First, because meta-analyses summarize quantitative findings, we cannot include case studies, despite the important evidence they provide. Second, our analyses are restricted to relationships examined by a sufficient number of primary studies such that, for example, we could not include strategy in our models. This also applies to the moderating variables such that, for example, we cannot include interactive and diagnostic uses of input, process, or output performance measures because prior studies do not make this distinction. Third, due to the nature of primary studies' data, we provide evidence only of linear, unconditional, simultaneous relationships; their interpretation is theory driven. Finally, variance due to correlated omitted variables might distort the findings. Whereas meta-analysis is inherently limited in this regard by the availability of primary studies, we use meta-analytic path models and control for environmental uncertainty and size to address this limitation to the largest possible extent.

Notwithstanding these limitations, we validate and extend theory on the nomological network of the levers. This enables a focused and structured evolution of future research. Moreover, in line with evidence-based management (Rousseau, 2006), the bigger picture of prior research that we present is vital for management control teaching and practice to move forward.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Declaration of competing interest

None.

Data availability

Due to the nature of meta-analysis, data for our analyses stem from the primary studies (see Web Appendix 1). Information regarding the coding of the primary studies can be found in Web Appendix 2.

Acknowledgments

We thank Ranjani Krishnan and two anonymous reviewers for their very helpful comments and suggestions. We are grateful for the support of several authors in providing us with their unpublished correlation analysis results. We thank Amrei Halbsguth for her invaluable support during the research process. For many very helpful suggestions and comments, we thank Albrecht Becker; David Bedford; Josep Bisbe; Martin Hiebl; Anne-Marie Kruis; Stefan Linder; Matthias Mahlendorf; Ricardo Malagueño; Patricia Martyn; Kai Mertens; Matthias Meyer; David Otley; Utz Schäffer; Sergeja Slapničar; Friedrich Sommer; Gerhard Speckbacher; Roland Speklé; Sandra Tillema; Paula van Veen-Dirks; Sally Widener; and participants at the Hamburg Management Accounting Research Seminars 2017 and 2019, Hamburg, Germany; the 14th Annual Conference for Management Accounting Research, Vallendar, Germany; the 40th EAA Annual Congress, Valencia, Spain; the 2017 MCA Conference, Groningen, The Netherlands; the Accounting Research Seminar Series, Groningen, The Netherlands; the Research Seminar of the Faculty of Economics, Ljubljana, Slovenia; the 80th Annual Conference of the VHB—German Academic Association for Business Research, Magdeburg, Germany; the 2018 Empirical Research in Management Accounting & Control Conference, Vienna, Austria; the EIASM 10th Conference on Performance Measurement and Management Control, Nice, France; the Research Seminar of the Financial Management and Accounting Research Group, Tampere, Finland; and the 100 years VHB: Anniversary conference of the German Academic Association of Business Research, Duesseldorf, Germany. A part of the work on this paper was done while Lucia Bellora-Bienengräber was affiliated with the University of Hamburg (Germany), Klaus Derfuss with the University of Dortmund (Germany), and Jan Endrikat with the TU Dresden (Germany) and the University of Hohenheim (Germany).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.aos.2022.101414.

References

Acquaah, M. (2013). Management control systems, business strategy and performance: A comparative analysis of family and non-family businesses in a transition economy in sub-Saharan Africa. *Journal of Family Business Strategy, 4*(2), 131–146.

Argyris, C. (1976). Single-loop and double-loop models in research on decision making. *Administrative Science Quarterly, 21*(3), 363–375.

Ashforth, B. E., & Mael, F. (1989). Social identity theory and the organization. *Academy of Management Review, 14*(1), 20–39.

Barney, J. B. (1991). Firm resources and sustained competitive advantage. *Journal of Management, 17*(1), 99–120.

Barney, J. B. (1995). Looking inside for competitive advantage. *Academy of Management Perspectives, 9*(4), 49–61.

Bedford, D. S. (2015). Management control systems across different modes of innovation: Implications for firm performance. *Management Accounting Research, 28*, 12–30.

Bedford, D. S., Bisbe, J., & Sweeney, B. (2019). Performance measurement systems as generators of cognitive conflict in ambidextrous firms. *Accounting, organizations and Society* (Vol. 72,, 21–37.

Bedford, D. S., & Malmi, T. (2015). Configurations of control: An exploratory analysis. *Management Accounting Research, 27*(June), 2–26.

Bedford, D. S., Malmi, T., & Sandelin, M. (2016). Management control effectiveness and strategy: An empirical analysis of packages and systems. *Accounting, Organizations and Society, 51*, 12–28.

Bellora-Bienengräber, L. (2019). Configurations of control in product development. *Journal of Accounting and Organizational Change, 15*(1), 127–146.

Bellora-Bienengräber, L., Radtke, R. R., & Widener, S. K. (2022). Counterproductive work behaviors and work climate: The role of an ethically focused management control system and peers' self-focused behavior. Accounting. *Organizations and Society, 96*, Article 101275.

Bergh, D. D., Aguinis, H., Heavey, C., Ketchen, D. J., Boyd, B. K., Su, P., … Joo, H. (2016). Using meta-analytic structural equation modeling to advance strategic management research: Guidelines and an empirical illustration via the strategic leadership-performance relationship. *Strategic Management Journal, 37*(3), 477–497.

Binder, A. (1963). Further considerations on testing the null hypothesis and the strategy and tactics of investigating theoretical models. *Psychological Review, 70*(1), 107–115.

Bisbe, J., Batista-Foguet, J.-M., & Chenhall, R. (2007). Defining management accounting constructs: A methodological note on the risks of conceptual misspecification. *Accounting, Organizations and Society, 32*(7–8), 789–820.

Bisbe, J., & Malagueño, R. (2009). The choice of interactive control systems under different innovation management modes. *European Accounting Review, 18*(2), 371–405.

Bisbe, J., & Malagueño, R. (2015). How control systems influence product innovation processes: Examining the role of entrepreneurial orientation. *Accounting and Business Research, 45*(3), 356–386.

Bisbe, J., & Otley, D. (2004). The effects of the interactive use of management control systems on product innovation. *Accounting, Organizations and Society, 29*(8), 709–737.

Blount, S., & Leinwand, P. (2019). Why are we here? *Harvard Business Review, 97*(6), 132–139.

Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen, & J. S. Long (Eds.), *Testing structural equation models* (pp. 136–162). Newbury Park: Sage.

Budianto, R., & Yuliansyha. (2014). An empirical assessment of interactive use of performance measurement system, organizational learning and firm performance. *GSTF Business Review (GBR), 3*(2), 44–53.

Burkert, M., Davila, A., Mehta, K., & Oyon, D. (2014). Relating alternative forms of contingency fit to the appropriate methods to test them. *Management Accounting Research, 25*(1), 6–29.

Busse, C., Mahlendorf, M. D., & Bode, C. (2016). The ABC for studying the too-much-of-a-good-thing effect: A competitive mediation framework linking antecedents, benefits, and costs. *Organizational Research Methods, 19*(1), 131–153.

Capon, N., Farley, J. U., & Hoenig, S. (1990). Determinants of financial performance: A meta-analysis. *Management Science, 36*(10), 1143–1159.

Chenhall, R. H. (2003). Management control systems design within its organizational context: Findings from contingency-based research and directions for the future. *Accounting, Organizations and Society, 28*(2–3), 127–168.

Chenhall, R. H., & Moers, F. (2015). The role of innovation in the evolution of management accounting and its integration into management control. *Accounting, Organizations and Society, 47*, 1–13.

Combs, J. G., Crook, R. T., & Shook, C. L. (2005). The dimensionality of organizational performance and its implications for strategic management research. In D. Ketchen, & D. Bergh (Eds.), *Research methodology in strategy and management* (pp. 259–286). Emerald Group Publishing Limited.

Combs, J. G., Ketchen, J., David, J., Crook, T. R., & Roth, P. L. (2011). Assessing cumulative evidence within 'macro' research: Why meta-analysis should be preferred over vote counting. *Journal of Management Studies, 48*(1), 178–197.

Curtis, E., Lillis, A. M., & Sweeney, B. (2017). Simons' levers of control framework: Commensuration within and of the framework. In *Advances in management accounting* (pp. 87–121). Emerald Publishing Limited.

Dalton, D. R., Daily, C. M., Certo, S. T., & Roengpitya, R. (2003). Meta-analyses of financial performance and equity: Fusion or confusion? *Academy of Management Journal, 46*(1), 13–26.

Damanpour, F. (1991). Organizational innovation: A meta-analysis of effects of determinants and moderators. *Academy of Management Journal, 34*(3), 555–590.

Derfuss, K. (2015). Relating context variables to participative budgeting and evaluative use of performance measures: A meta-analysis. *Abacus, 51*(2), 238–278.

Edwards, J. R. (2001). Multidimensional constructs in organizational behavior

research: An integrative analytical framework. *Organizational Research Methods, 4*(2), 144–192.
Edwards, J. R., & Berry, J. W. (2010). The presence of something or the absence of nothing: Increasing theoretical precision in management research. *Organizational Research Methods, 13*(4), 668–689.
Endrikat, J., Guenther, T. W., & Titus, R. (2020). Consequences of strategic performance measurement systems: A meta-analytic review. *Journal of Management Accounting Research, 32*(1), 103–136.
Franco-Santos, M., Lucianetti, L., & Bourne, M. (2012). Contemporary performance measurement systems: A review of their consequences and a framework for research. *Management Accounting Research, 23*(2), 79–119.
Gagné, M., & Deci, E. L. (2005). Self-determination theory and work motivation. *Journal of Organizational Behavior, 26*(4), 331–362.
Galbraith, J. (1973). *Designing complex organizations*. Boston: Addison Wesley Publishing Company.
Garcia Osma, B., Gomez-Conde, J., & de las Heras, E. (2018). Debt pressure and interactive use of control systems: Effects on cost of debt. *Management Accounting Research, 40*, 27–46.
George, B. A. (2011). Entrepreneurial orientation: A theoretical and empirical examination of the consequences of differing construct representations. *Journal of Management Studies, 48*(6), 1291–1313.
Geyskens, I., Krishnan, R., Steenkamp, J. B. E. M., & Cunha, P. V. (2009). A review and evaluation of meta-analysis practices in management research. *Journal of Management, 35*(2), 393–419.
Gordon, L. A., & Narayanan, V. K. (1984). Management accounting systems, perceived environmental uncertainty and organization structure: An empirical investigation. *Accounting, Organizations and Society, 9*(1), 33–47.
Guenther, T. W., & Heinicke, A. (2019). Relationships among types of use, levels of sophistication, and organizational outcomes of performance measurement systems: The crucial role of design choices. *Management Accounting Research, 42*(2019), 1–25.
Heggen, C. (2019). The role of value systems in translating environmental planning into performance. *The British Accounting Review, 51*, 130–147.
Heine, B.-O., Meyer, M., & Strangfeld, O. (2005). Stylised facts and the contribution of simulation to the economic analysis of budgeting. *The Journal of Artificial Societies and Social Simulation, 8*(4).
Heinicke, A., Guenther, T. W., & Widener, S. K. (2016). An examination of the relationship between the extent of a flexible culture and the levers of control system: The key role of beliefs control. *Management Accounting Research, 33*, 25–41.
Henri, J.-F. (2006). Management control systems and strategy: A resource-based perspective. *Accounting, Organizations and Society, 31*(6), 529–558.
Hofmann, S., Wald, A., & Gleich, R. (2012). Determinants and effects of the diagnostic and interactive use of control systems: An empirical analysis on the use of budgets. *Journal of Management Control, 23*(3), 153–182.
Hogreve, J., Iseke, A., Derfuss, K., & Eller, T. (2017). The service–profit chain: A meta-analytic test of a comprehensive theoretical framework. *Journal of Marketing, 81*(3), 41–61.
Hollensbe, E., Wookey, C., Hickey, L., & George, G. (2014). From the editors: Organizations with purpose. *Academy of Management Journal, 57*(5), 1227–1234.
Homans, G. C. (1958). Social behavior as exchange. *American Journal of Sociology, 63*(6), 597–606.
Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*(1), 1–55.
Hunter, J. E., & Schmidt, F. L. (2000). Fixed effects vs. random effects meta-analysis models: Implications for cumulative research knowledge. *International Journal of Selection and Assessment, 8*(4), 275–292.
Iacobucci, D., Saldanha, N., & Deng, X. (2007). A meditation on mediation: Evidence that structural equations models perform better than regressions. *Journal of Consumer Psychology, 17*(2), 139–153.
Janka, M., & Guenther, T. W. (2018). Management control of new product development and perceived environmental uncertainty: Exploring heterogeneity using a finite mixture approach. *Journal of Management Accounting Research, 30*(2), 131–161.
Journeault, M., De Rongé, Y., & Henri, J.-F. (2016). Levers of eco-control and competitive environmental strategy. *The British Accounting Review, 48*(3), 316–340.
Judge, T. A., & Kammeyer-Mueller, J. D. (2012). General and specific measures in organizational behavior research: Considerations, examples, and recommendations for researchers. *Journal of Organizational Behavior, 33*(2), 161–174.
Kaldor, N. (1961). Capital accumulation and economic growth. In D. C. Hague (Ed.), *The theory of capital* (pp. 177–222). Springer.
Karna, A., Richter, A., & Riesenkampff, E. (2016). Revisiting the role of the environment in the capabilities–financial performance relationship: A meta-analysis. *Strategic Management Journal, 37*(6), 1154–1173.
Kirca, A. H., Jayachandran, S., & Bearden, W. O. (2005). Market orientation: A meta-analytic review and assessment of its antecedents and impact on performance. *Journal of Marketing, 69*(2), 24–41.
Kline, R. B. (2011). *Principles and practices of structural equation modeling* (3rd ed.). New York: The Guilford Press.
Kruis, A.-M., Speklé, R. F., & Widener, S. K. (2016). The levers of control framework: An exploratory analysis of balance. *Management Accounting Research, 32*, 27–44.
Langfield-Smith, K. (2007). A review of quantitative research in management control systems and strategy. In C. S. Chapman, A. G. Hopwood, & M. D. Shields (Eds.), *Handbooks of management accounting research* (Vol. 2, pp. 753–783). Elsevier.
Law, K. S., & Wong, C.-S. (1999). Multidimensional constructs in structural equation analysis: An illustration using the job perception and job satisfaction constructs. *Journal of Management, 25*(2), 143–160.
Libby, R., Bloomfield, R., & Nelson, M. W. (2002). Experimental research in financial accounting. *Accounting, Organizations and Society, 27*(8), 775–810.
Lopez-Valeiras, E., Gonzalez-Sanchez, M. B., & Gomez-Conde, J. (2016). The effects of the interactive use of management control systems on process and organizational innovation. *Review of Managerial Science, 10*(3), 1–24.
Mahama, H., & Cheng, M. M. (2013). The effect of managers' enabling perceptions on costing system use, psychological empowerment, and task performance. *Behavioral Research in Accounting, 25*(1), 89–114.
Martyn, P., Sweeney, B., & Curtis, E. (2016). Strategy and control: 25 years of empirical use of Simons' levers of control framework. *Journal of Accounting and Organizational Change, 12*(3), 281–324.
Merchant, K. A., & Otley, D. (2020). Beyond the systems versus package debate. *Accounting, Organizations and Society, 86*, Article 101185.
Mundy, J. (2010). Creating dynamic tensions through a balanced use of management control systems. *Accounting, Organizations and Society, 35*(5), 499–523.
Naranjo-Gil, D. (2016). Role of management control systems in crafting realized strategies. *Journal of Business Economics and Management, 17*(6), 865–881.
Petter, S., Straub, D., & Rai, A. (2007). Specifying formative constructs in information systems research. *MIS Quarterly*, 623–656.
Podsakoff, P. M., MacKenzie, S. B., Lee, J.-Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology, 88*(5), 879–903.
Podsakoff, P. M., MacKenzie, S. B., & Podsakoff, N. P. (2012). Sources of method bias in social science research and recommendations on how to control it. *Annual Review of Psychology, 63*, 539–569.
Rousseau, D. M. (2006). Is there such a thing as "evidence-based management"? *Academy of Management Review, 31*(2), 256–269.
Schaefer, T., & Guenther, T. (2016). Exploring strategic planning outcomes: The influential role of top versus middle management participation. *Journal of Management Control, 27*(2), 205–249.
Schmidt, F. L., & Hunter, J. E. (2015). *Methods of meta-analysis: Correcting error and bias in research findings*. Newbury Park et al: Sage Publications.
Schneider, B., Ehrhart, M. G., & Macey, W. H. (2013). Organizational climate and culture. *Annual Review of Psychology, 64*, 361–388.
Schreyögg, G., & Kliesch-Eberl, M. (2007). How dynamic can organizational capabilities be? Towards a dual-process model of capability dynamization. *Strategic Management Journal, 28*(9), 913–933.
Schwab, D. P. (2005). In N. J. Mahwah (Ed.), *Research methods for organizational studies* (2nd ed.). Lawrence Erlbaum Associates.
Schweiger, S. A., Stettler, T. R., Baldauf, A., & Zamudio, C. (2019). The complementarity of strategic orientations: A meta-analytic synthesis and theory extension. *Strategic Management Journal, 40*(1), 1822–1851.
Simons, R. (1995). *Levers of control–how managers use innovative control systems to drive strategic renewal*. Boston: Harvard Business School.
Simons, R. (2000). *Performance measurement and control systems for implementing strategy*. Upper Saddle River: Pearson.
Simons, R., & Dávila, A. (2021). How top managers use the entrepreneurial gap to drive strategic change. *European Accounting Review, 30*(4), 583–609.
Speklé, R. F., van Elten, H. J., & Widener, S. K. (2017). Creativity and control: A paradox. Evidence from the levers of control framework. *Behavioral Research in Accounting, 29*(2), 73–96.
Speklé, R. F., & Widener, S. K. (2020). Insights on the use of surveys to study management control systems. *Accounting, Organizations and Society, 86*, Article 101184.
Tanriverdi, H., & Venkatraman, N. (2005). Knowledge relatedness and the performance of multibusiness firms. *Strategic Management Journal, 26*(2), 97–119.
Tessier, S., & Otley, D. (2012). A conceptual development of Simons' levers of control framework. *Management Accounting Research, 23*(3), 171–185.
Tuomela, T.-S. (2005). The interplay of different levers of control: A case study of introducing a new performance measurement system. *Management Accounting Research, 16*(3), 293–320.
Whetten, D. A. (1989). What constitutes a theoretical contribution? *Academy of Management Review, 14*(4), 490–495.
Widener, S. K. (2007). An empirical analysis of the levers of control framework. *Accounting, Organizations and Society, 32*(7/8), 757–788.
Wood, J. (2008). Methodology for dealing with duplicate study effects in a meta-analysis. *Organizational Research Methods, 11*(1), 79–95.
Zhao, X., Lynch, J. G., & Chen, Q. (2010). Reconsidering Baron and Kenny: Myths and truths about mediation analysis. *Journal of Consumer Research, 37*(2), 197–206.