The Implications of Feedback Frequency for Employee Creativity

Dissertation

zur Erlangung des akademischen Grades
Doktor der Wirtschafts- und Sozialwissenschaften (Dr. rer. pol.)

Vorgelegt im
Fachbereich 07 – Wirtschaftswissenschaften
der Universität Kassel

von Sören Abel

Kassel, den 11.11.2021

Disputation am 25.03.2022

Dedicated to my entire family,

both here and in the beyond.

Marianna and Juliana, I will always be grateful to you

for accompanying and carrying me on this journey.

I adore you

Declaration

Declaration in accordance with § 8 of the General Provisions for Doctoral Degrees at the University of Kassel dated 14.07.2021:

1. I herewith give assurance that the submitted dissertation "*The Implications of Feedback Frequency for Employee Creativity*" is the product of my work alone.
2. I affirm that I have not used any sources or aids other than those explicitly cited. All passages that are drawn from published or unpublished writings, either word-for-word or in paraphrase, have been clearly identified as such.
3. The dissertation or parts of it
   - has/have not been used in another doctoral or tenure process in Germany or abroad.
   - ~~has/have been submitted in another doctoral or tenure process in Germany or abroad.~~
4. Both, the submitted digital version and the submitted written copies are identical.
5. Third parties were not involved in the drafting of the content of this dissertation; most specifically I did not employ the assistance of a commercial dissertation advisor.
6. In case of a cumulative dissertation: I have documented the participation of co-authors by attaching a written statement signed by all of the co-authors outlining which parts of jointly written contributions are attributed to which co-authors.
7. I confirm the correctness of the above mentioned statements.

Sören Abel

Overview of Contents

Table of Contents

Table of Figures

1. Introduction

We witness times of increasingly frequent change in which firms rely on the creativity of their employees to remain competitive (Anderson, Potocnik, and Zhou 2014). As they do so, firms are more than ever confronted with the threat of outdated information that renders employees ineffective. Quick feedback turnover times overcome this issue by preventing the information from becoming obsolete or superseded by disruptive events, ensuring a seamless flow of information along the links of the information chain within an organization, and providing guidance to employees (Crowder 2014; Ozkok et al. 2019; Thornock 2016).

At the same time, quantum leaps in technological advancements have improved firms' ability to collect, analyze and exploit data. While "immediate feedback provision until now has either been impractical or not possible" (Lechermeier and Fassnacht 2018, p. 146), evolving technology nullifies seemingly everlasting limitations and mitigates some of the costs involved in the provision of feedback (Goomas and Ludwig 2017), making it "easier for companies to provide frequent feedback" (Holderness, Olsen, and Thornock 2020, p. 156). Correspondingly, practitioners report substantial upcoming changes in firms' feedback systems (PricewaterhouseCoopers 2015). Finding themselves at the epicenter of this transformation, managers are advised to recognize feedback frequency as "a critical feature of the new performance management paradigm" (Deloitte Global Human Capital Trends 2017, p. 71) and to provide employees with "more frequent feedback sessions" (PricewaterhouseCoopers 2015, p. 5), since the inflexible and rigid "end-of-year evaluation will be gone, replaced by continuous feedback" (Deloitte 2017, p. 71). In parallel, employees, who are at the forefront of the increasingly fast-paced business world, are demanding "more regular feedback" from their supervisors (Deloitte 2017, p. 66). Contrary to employees' presumably ever-present cry for more frequent feedback, their demands today are not only the result of facing a more volatile work environment, but are also exacerbated by their experiences with feedback outside of their work, such as on social media (Deloitte 2017).

These global business trends are redefining the imperatives and opportunities in designing feedback systems in terms of feedback frequency (i.e., how often feedback is provided), such that the question of "how to give people feedback is one of the hottest topics in business today" (Buckingham and Goodall 2019, p. 101).

Also from a research perspective, feedback and its frequency have long been recognized as "a central component of an organization's management accounting and control system" and are "central to the discussion of system design […] in the management accounting literature, especially in the context of management control systems" (Luckett and Eggleton 1991, p. 371). Similarly, in the feedback literature (e.g., Ilgen, Fisher, and Taylor 1979; Taylor, Fisher, and Ilgen 1984), the frequency of feedback is considered a major component of the feedback stimulus that determines the feedback recipient's response. Thus, it is not surprising that feedback frequency has recently regained attention in academic discussion and has been shown to play a pivotal role in the effectiveness of feedback (e.g., Casas-Arce, Lourenco, and Martinez-Jerez 2017; Holderness, Olsen, and Thornock 2020). However, knowledge about the effects of feedback frequency is still scarce.

Particularly when firms seek to foster employee creativity, they face an as yet unsolved dilemma: the nature of the ideation process benefits from frequent feedback providing employees with multiple stimuli that may serve as starting points for novel ideas and enabling them to reflect on their current creative attempts, as the desired creative outcome cannot be determined ex ante (Davila and Ditillo 2017; Sijbom et al. 2018); yet, frequent feedback can erode creative germs by limiting flexibility, creating a sense of being controlled, and thus undermining intrinsic motivation (Amabile, Goldfarb, and Brackfield 1990; Speckbacher 2017). In addition, feedback frequency has been linked to myopia and information overload (Holderness, Olsen, and Thornock 2020; Lurie and Swaminathan 2009), which might adversely affect employees' cognitive resources and engagement in generating creative ideas. Consequently, the current demands placed on the feedback systems' frequency could prove to be a fatal trap for creativity[1]. Despite this threat, the impact of feedback frequency on employee creativity constitutes a substantial gap in research.

Fostering employee creativity has become a hot topic of research in many strands of literature, such as service research (e.g., Agnihotri et al. 2014a) or leadership research (e.g., Hughes et al. 2018). Notably, in management control system research, a "paradigm shift away from the traditional focus on established objectives and stable environments" (Adler and Chen 2011, p. 63) has become an impetus for exploring the "paradox" of what has long been considered incompatible: control (e.g., feedback) and employee creativity (e.g., Speklé, van Elten, and Widener 2017; Davila and Ditillo 2017). Creative ideas initiate and are prerequisite for organizational innovation[2]; thus, "creativity is widely seen as an important source of competitive advantage and business success" (Speckbacher 2017, p. 49). Viewing employees as a source of creative ideas allows firms to incorporate relevant perspectives, perceptions, and experiences into organizational development. For instance, employees are crucial for firms in uncovering customer needs, and they recognize firsthand when scripted behaviors are not enough to meet them and creative ideas are

---

[1] This dissertation uses the terms "creativity" and "employee creativity" interchangeably.

[2] While creativity refers to the process of idea generation, innovation encompasses subsequent processes for idea implementation (Amabile 1988; Hughes et al. 2018). Therefore, while the product of creativity is an idea, the product of innovation is an implemented idea (Hughes et al. 2018).

required (Agnihotri et al. 2014a). Hence, "to survive, adapt, and gain competitive advantage, organizations need to fully take advantage of their employees' innate creative potential" (Zhou and George 2003). Nonetheless, managing employee creativity is not straightforward since creativity displays an intrapersonal process that – despite being largely cognitive – is also influenced by motivational, affective, or social-relational factors (Hughes et al. 2018).

In addition to the complex nature of creativity, previous literature describes multiple functions of feedback, such as information, motivation, controlling, and explanation functions, as well as several factors influencing feedback effectiveness, such as feedback source (e.g., supervisor or customer) or recipient characteristics (e.g., experience) (cf. Agnihotri et al. 2014a; Coelho and Augusto 2010; Deci, Connell, and Ryan 1989; Ilgen, Fisher, and Taylor 1979; Mullins, Agnihotri, and Hall 2020), which potentially have contrasting effects on the impact of feedback frequency on employee creativity. Even more, employees are constantly exposed to the influence of multiple feedback sources[3] (e.g., supervisor or customer), and novel "technology is being utilized to support frequent and multi-source feedback" (PricewaterhouseCoopers 2015, p. 5). Thus, managers are still floundering when they have to decide on the optimal level of feedback frequency; either when they consider themselves as a source of feedback, but also when they shape the feedback environment of their employees, e.g., by allowing more (or less) frequent customer feedback. Furthermore, given that feedback is a time-consuming task, managers might be reluctant to provide feedback frequently, foregoing a potentially powerful lever for increasing employee creativity. Ultimately, clarifying the ambiguities about the

[3] From a psychological perspective, even the task and "one's own self" represent important sources of feedback for the individual (e.g., Greller and Herold 1975). Thus, regardless of the employee's job description or position, he or she is always confronted with feedback from more than one source.

implications of feedback frequency on employee creativity is essential since feedback frequency is an inevitable managerial decision – even no feedback has a frequency.

Consequently, a major goal of this dissertation is to fill this research gap and examine the effects of feedback frequency on employee creativity. In doing so, I make important contributions to various streams of literature, such as management control system research, the feedback literature, service research, and gender research. As an empirical basis, three distinct studies are conducted, two surveys (n = 385; n = 400) and one field experiment (n = 105), and the data is analyzed with state-of-the-art procedures, e.g., PLS-POS, FIMIX-PLS, or high-creativity idea identification (cf. Hair et al. 2016; Kachelmeier, Wang, and Williamson 2019; Matthews et al. 2016). Indeed, my results reveal that the frequency of feedback is a powerful lever for managers to enhance employee creativity. However, the findings underline the imperative of a thoughtful decision about the frequency of the provided feedback. Furthermore, I discuss how these insights can be exploited by managers to design superior feedback strategies.

More specifically, by drawing on multiple theories (see chapter 2), I am able to uncover critical factors that moderate the impact of feedback frequency on employee creativity, including recipient characteristics, feedback source[4], feedback source characteristics, and use of feedback. Besides, I demonstrate how feedback frequencies from multiple sources interact. Moreover, I open the black box of intrapersonal processes that mediate the effects of feedback frequency on recipient

[4] I follow current research on feedback sources (Andiola 2014; Lechermeier and Fassnacht 2018) by distinguishing between source (i.e., conceptual "original" source of information about work performance, such as customers or supervisor) and source characteristics (i.e., the source attributes, such as credibility or gender). Yet, I acknowledge that these two cannot always be completely separated, "since certain characteristics are ascribed to certain sources and therefore build their underlying basis" (Lechermeier and Fassnacht 2018, p. 150).

(i.e., the employee) creativity and provide insights into the complicated conversion process of feedback frequency corresponding to various, sometimes antagonistic, feedback functions. Furthermore, I corroborate the impact of feedback frequency on employee creativity by ruling out the effects of other feedback dimensions that are often confounded with feedback frequency in practice (and in some of previous research) that might interfere with this relationship, namely feedback valence, feedback timing, and feedback quantity. Complementarily, I identify creative requirements on the feedback recipient (i.e., the employee) as an important driver for supervisors in their decision of the level of feedback frequency and how they use feedback. Lastly, I investigate the implications of feedback frequency on the quality of employee creativity, since in practice only a fraction of all generated ideas can be implemented in organizations due to resource constraints.

The remainder of the dissertation is organized as follows. In *chapter 2*, I outline my research approach by providing a brief overview of the theories drawn upon to explain the creative effects of feedback frequency and elaborating on the empirical methodology employed. Corresponding to the methods used in the studies, this chapter includes sections for survey research with partial least-squares structural equation modeling and for experimental research. Next, *chapter 3* presents the first survey-based study. This study observes service providers who are particularly susceptible to the influence of multiple feedback sources due to their boundary-spanning position. The results demonstrate that the frequency of both customer and supervisor feedback influences employee creativity. Furthermore, it is shown how the feedback stimulus is transformed into a creative response and how various feedback functions apply to feedback frequency in this process. Within *chapter 4*, an experimental field study provides evidence for the effectiveness of feedback frequency by controlling for related feedback dimensions. Besides, the influences of gender of the feedback source, gender of the feedback recipient, and gender (in)congruence on feedback frequency effectiveness are examined, and a more nuanced understanding of the mechanisms that transform feedback frequency into creativity is provided. Then, in the second survey-based study feedback frequency is considered from a management control system perspective in *chapter 5,* exhibiting the importance of control use for its creative effects. *Chapter 6* summarizes the results of the dissertation.

2. Research Approach

The epistemology of this dissertation follows a positivistic approach to research (Amaratunga et al. 2002; Collis and Hussey 2013; Easterby-Smith, Thorpe and Jackson 2012). In a positivistic approach, hypotheses are derived from established theories and verified (or falsified) by (quantitative) empirical methods in a specific context (deduction). In other words, the empirical methods link (and confront) theories with reality. Hence, the empirical results generalize (and may update) a theory, which represents a consistent set of statements that form a homogenous edifice of ideas. Accordingly, Engel, Blackwell, and Miniard (1995, p. G11) describe positivism as the approach to research "in which rigorous empirical techniques are used to discover generalized explanations and laws". In contrast, a phenomenological approach to research strives to understand a particular phenomenon and to develop "theories through explanatory methods" (Amaratunga et al. 2002, p. 19), such as qualitative and naturalistic empiricism (induction). Nevertheless, since my particular aim is to "look for causality and fundamental laws" (Amaratunga et al. 2002, p. 19) regarding the implications of feedback frequency on employee creativity, I pursue a positivistic approach to achieve my research objectives.

Before appropriate theories can be selected to explain the impact of feedback frequency on employee creativity, a common understanding of feedback frequency and employee creativity needs to be obtained. In this dissertation feedback is defined as communication between a sender (i.e., feedback source) and a recipient (i.e., the employee) about the recipient's job-related behaviors or performance (Agnihotri et al. 2014a; Mullins, Agnihotri, and Hall, 2020). The underlying assumptions are that feedback is interpersonal (i.e., while the feedback not necessarily needs to be provided in person, the sender and receiver of the performance information are both human, which does not apply to impersonal performance information processes alike managerial reporting), dyadic (i.e., there is one employee receiving feedback from one source[5]), and one-sided (i.e., the feedback source is the initiator of the feedback exchange[6]). Thus, this feedback understanding separates from notions of upward employee feedback (Kim and Kim 2020; Jhun, Bae, and Rhee 2012), feedback seeking behavior (Auh et al. 2019; Anseel et al. 2015; Sijbom et al. 2018), or team feedback (Hoever, Zhou, and van Knippenberg 2018).

Feedback frequency, in turn, determines how often this communication occurs (within a given time period). Hence, formal feedback activities that routinely provide performance information to employees, such as performance appraisal systems (Kuvaas 2011), may be part of feedback frequency (if they conform to the above-mentioned criteria), but feedback frequency is not limited to these.

When it comes to employee creativity, the widely used definition of Shalley, Zhou, and Oldham (2004, p. 934) is drawn upon here, which refers to creativity as “the development of ideas about product, practices, services, or procedures that are (a) novel and (b) potentially useful to the organization”. However, while the authors consider ideas as novel “if they are unique relative to other ideas currently available in the organization” (p. 934), an individual perspective regarding the novelty of ideas (i.e., thinking outside “one’s own” box) is adopted in this dissertation, since it conceptualizes creativity as a process that takes place within an individual (in contrast to, for instance, team creativity research, e.g., Adler and Chen 2011; Davila and Ditillo 2017; Klein and

[5] This does not mean that the employee can only receive feedback from one source, but that feedback from a different source is considered as a different feedback (e.g., customer vs. supervisor feedback), nor that a feedback source consists of only one person (e.g., customer feedback) (e.g., Greller and Herold 1975; Lechermeier and Fassnacht 2018).
[6] This does not imply that the feedback source mechanistically elaborates on his or her feedback issues, but that the feedback conversation centers around the employee’s performance.

Speckbacher 2020). Therefore, an employee's creativity does not depend on ideas generated by other organizational members and unknow to the employee, but an idea is considered creative relative to the employee's prior information, experiences, or (cognitive) resources. This approach is consistent with the examined research questions of whether, how and when an employee is stimulated to develop creative ideas by the frequency of feedback.

Similar to this notion, it is acknowledged that not only a set of employees (e.g., members of the research and development department) can be creative. Considering that creativity is the "first step that is necessary for subsequent innovation", which refers to the implementation of ideas in the organization (Shalley, Zhou, and Oldham 2004, p. 934), it is particular important for organizations to include different perspectives and local expert knowledge in the ideation processes to increase their innovative potential – for example, the janitor may be more apt to come up with an idea to save water in the irrigation of the firm's green spaces than the CEO. Or as Zhou and George (2003, p. 547 f.) put it, "while a Pulitzer prize-winning author and a designer of a functioning prosthetic limb are creative at work, so too is a secretary who designs a new scheduling system to reduce errors and a nurse who comes with a way to reallocate nursing responsibilities in a unit to increase the quality of patient care". Consequently, creativity is understood as component of the overall employee performance[7] (Ng and Feldman 2008), but one that has specific characteristics, such as being heuristic, since the intended outcome cannot be determined ex ante (Speckbacher 2017). At an abstract level, this dissertation therefore joins a creativity literature stream that examines contextual factors (i.e., feedback frequency) that influence creativity, and that has emerged from creativity research that focuses on

[7] Thus, feedback can address the employee's creative attempts, but it does not have to.

personal characteristics[8] on the premise that "individuals with certain personality characteristics may be especially effective at recognizing problems or at combining new information" (Shalley, Zhou, and Oldham 2004, p. 935; Coelho and Augusto 2010).

Building on this consistent understanding of feedback frequency and employee creativity in this dissertation and reflecting on the research questions underlying each of the three distinct studies, I identified theories established in international research that are suitable to explain the creative effects of feedback frequency and serve as starting points for deriving hypotheses (positivistic deduction). Consequently, these theories have a common ground in that they, for instance, consider feedback frequency as a lever to influence employee creativity, refer to an individual level and are employee-specific (e.g., by emphasizing employee perceptions), are applicable to employees across hierarchical levels and regardless of their departmental affiliation, or describe an interpersonal dyadic relationship[9] (e.g., addressor and addressee of management controls, feedback provider and recipient, leader and subordinate, or role sender and focal person[10]).

More specifically, *Study 1* aims to identify the intrapersonal processes that transform feedback frequency from multiple sources into a creative response. Hence, it draws on the dynamic Stimulus-Organism-Response (S-O-R) framework (Jacoby 2002), which explains how an external stimulus and intertemporal linkages between

---

[8] While personal characteristics are included in the analyses in Study 1 and in Study 2, namely employee experience and the gender of the feedback recipient, these variables are considered only as determinants of the effectiveness of feedback frequency and not as predictors of employee creativity. The latter is also confirmed by the empirical results.

[9] While the Stimulus-Organism-Response framework is not limited to interpersonal triggers, in Study 1 a feedback stimulus is assumed to initiate the S-O-R process. Hence, the general S-O-R framework is specified to reflect an interpersonal dyadic relationship as well.

[10] Thus, these terms are used synonymously in this dissertation to describe both actors in the feedback dyad, i.e., when referring to the feedback recipient, it always means an employee.

multiple stimuli (e.g., feedback frequency) trigger behavioral responses (e.g., the generation of creative ideas) by recognizing affective and cognitive processes within the individual (organism). These intrapersonal processes (organism) are conceptualized by role theory (Katz and Kahn 1978), which states that feedback sources communicate their perceived congruence of a recipient's behavior with their (role) expectations through feedback to recipients. Role theory is particularly well suited for observing the impact of feedback frequency from multiple sources as it specifies the employee's perceived ambiguities about these expectations with respect to each of the sources (e.g., customer role ambiguity and supervisor role ambiguity) and accounts for potential perceived conflicts between the expectations of these sources (i.e., inter-sender role conflict). The joint choice of role theory and dynamic S-O-R is further substantiated as it allows to identify several distinct feedback frequency functions.

In *Study 2* the implications of gender for the creative effectiveness of feedback frequency are examined. For this research objective, feedback theory (Ilgen, Fisher, and Taylor 1979) is selected, which describes how feedback frequency affects the recipient's internal processing of the feedback stimulus. In contrast to other theories that focus on providing a framework for feedback effects, e.g., feedback intervention theory (Kluger and DeNisi 1996) or control theory (Carver and Scheier 1981), the seminal work of Ilgen, Fisher, and Taylor (1979) particularly regards feedback frequency as a pivotal component of the feedback stimulus and provides specific propositions about which stage of the recipient's internal processing of the feedback stimulus is affected by feedback frequency. Besides, although this theory clearly distinguishes between different feedback dimensions (e.g., feedback frequency, feedback valence, or feedback timing), it acknowledges that these dimensions may also have interactive effects, which supports the study's pursuit of identifying potential confounding variables and establishing an interfere-free relationship between feedback frequency and employee creativity. Furthermore, this feedback theory guides me in incorporating gender of the feedback source (source characteristics) and gender of the feedback recipient (recipient characteristics) into the research model. To refine the research model, gender role theory (Eagly and Karau 1991; Eagly and Wood 2012; Lanaj and Hollenbeck 2015) is employed to specify the influence of gender for the effectiveness on feedback frequency. Gender role theory posits that gender-specific stereotypical norms of leadership behaviors, such as the provision of feedback (frequency), determine perceptions of (and thus responses to) that behaviors as a function of the gender of the leader and the observer (i.e., the feedback recipient). Not only does gender role theory offer explicit arguments for the potential influence of the gender of the feedback source, the gender of the feedback recipient, and gender (in)congruence on feedback frequency effectiveness, but it is also consistent with Ilgen, Fisher, and Taylor (1979) in focusing on the importance of recipient perceptions in determining (creative) outcomes.

Finally, *Study 3* aims to investigate the relationship between feedback frequency and employee creativity from a management control perspective. This study draws on Tessier and Otley's (2012) conceptualization of management control systems as it emphasizes the importance of control presentation (e.g., its frequency) for the effectiveness of controls (e.g., feedback). Since this framework represents a revised version of Simons' (1995) seminal Levers of Control framework, the underlying issues of how to manage the tension "between top-down direction and bottom-up creativity" (Simons 1995, p. 4) and "how managers can combine innovation and control" (Adler and Chen 2011, p. 63) remain fundamental. Thus, this framework not only allows me to locate feedback frequency in management control research (following Tessier and Otley (2012), it refers to how the control "feedback" is communicated), but also considers employee creativity as a particular control outcome. In addition, this framework was chosen because it corresponds to the underlying individual perspective of feedback (control) and creativity (individual control outcome). Because this framework posits that control outcomes are determined by the interplay of control presentation (e.g., feedback frequency) and managerial intentions (e.g., control use), this guides me in identifying the moderating effects of interactive feedback use and diagnostic feedback use for the implications of feedback frequency on employee creativity.

As the underlying theories and their explanations of the creative effects of feedback frequency are expediently applied in more detail in the respective studies (particularly in the hypotheses), the focus of the remainder of this chapter is to provide fundamental insights into the empirical methodologies employed in the three studies, namely survey research with partial least squares structural equation modelling and experimental research, utilized to test the theory-derived hypotheses positivistically.

2.1 Survey Research with Partial Least Squares Structural Equation Modelling

2.1.1 Introducing PLS-SEM

Structural Equation Modeling (SEM hereafter) is a multivariate analysis and especially advantageous for social science as it enables to assess relationships between unobservable constructs (i.e., latent variables). Many conceptual constructs in the field of management accounting or controlling are not directly observable, for example, individual’s perceptions, attitudes, or cognitions. Therefore, scales have been developed to operationalize latent constructs through measurable variables (Döring and Bortz 2016; Hair, Hult et al. 2017). In contrast to so-called “first-generation techniques", such as regression-based approaches, SEM, as a "second-generation technique", takes into account the distinct but interrelated multi-level model structure resulting from latent variable characteristics: the inner model (latent variable level), which displays the relationships between the latent variables, and the outer model (measurement level), that illustrates the operationalization of the latent variable through observable manifest variables (Fornell 1982; Fornell and Larcker 1987). A complete path model therefore consists of one inner model and multiple outer models (one for each latent variable).

In comparison to first-generation techniques that utilize summed-scored scales or mean-scored scales, respectively, which might at first glance address unobservability issues of latent variables (Henseler et al. 2014), SEM systematically accounts for measurement errors (Chin 1998). The complexity of the measurement model reduces measurement errors (e.g., arising from poorly worded questions, misunderstood scales, or validity issues) and thus increases the quality of measurement (Hair, Hult et al. 2017; Thiele, Sarstedt, and Ringle 2015). Furthermore, SEM enables to test more complex inner models, for instance, by allowing constructs to be endogenous and exogenous simultaneously (i.e., mediation) and thereby to assess indirect effects (Döring and Bortz 2016). Consequently, SEM cannot only provide evidence of the statistical relationship between two constructs, but also test entire models (Döring and Bortz 2016). As a major goal of this dissertation is to dissect the intrapersonal processes stimulated by feedback frequency, which requires measuring constructs that are imperceptible from the outside (e.g., role stress) and complex construct relationships (e.g., mediation or moderation), data of both survey-based studies (see chapter 3 and chapter 5) are analyzed using SEM.

Partial Least Squares (PLS) has become the dominant method for analyzing SEM (Hair et al. 2019). Whereas early SEM-research largely used covariance based (CB) SEM (LISREL and AMOS are among the most frequently executed software), considerable discussions (cf. Goodhue, Lewis, and Thompson 2012; Hair, Matthews et al. 2017; Hair, Ringle, and Sarstedt 2011; Henseler et al. 2014; Khan et al. 2019; Marcoulides, Chin, and Saunders 2012; Marcoulides and Saunders 2006; Rigdon 2012, 2014a, 2016) about when to use CB-SEM or PLS-SEM have since led to established decision criteria (for a more extensive overview see Hair et al. 2019): goal of the analysis, construct measurement, model complexity, and sample characteristics. Firstly, PLS-SEM is a variance-based approach and strives to maximize the explained variance (R^2) of endogenous constructs (Hair, Ringle, and Sarstedt 2011). Thus, "PLS-SEM is a causal-predictive approach to SEM that emphasizes prediction in estimating statistical models, whose structures are designed to provide causal explanations" (Hair et al. 2019, p. 3; Sarstadt, Ringle, and Hair 2017; Wold 1982). Secondly, there are no difficulties in PLS-SEM regarding the measurement of latent variables with formative, reflective, or with single-item measures (MacCallum and Browne 1993). Thirdly, PLS-SEM allows for higher structural model complexity, which increases with the number of constructs (latent variables), number of paths, and number of items (manifest variables). Finally, PLS-SEM makes no distributional assumptions about the underlying data (Cassel, Hackl, and Westlund 1999). Wold (1982) hereby refers to, what he calls, "soft model basic design", but "it is not the concepts nor the models nor the estimation techniques which are 'soft', only the distributional assumptions" (Lohmöller 1989, p. 64). In addition, PLS-SEM provides advanced analysis features, such as PLS-POS or FIMIX-PLS, which will be discussed in more detail later in this chapter. Following these considerations, PLS-SEM is employed for data analysis in both survey-based studies of this dissertation.

2.1.2 The Inner Model – Designing Relationships between Latent Variables

In the inner model (also referred to as structural model), the sequence of the latent variables and their relationships is determined. Traditionally, the latent variables are arranged from left to right, such that the exogenous latent variables (predictors) are placed on the left and the endogenous latent variables (outcome) are placed on the right. Hence, a logical structure emerges when the inner model is "read": Latent variables on the left side affect latent variables on the right side (Sarstadt, Ringle, and Hair 2017). Figure 2.1 illustrates such a path model. Here, the exogenous variables Y_1 and Y_2 affect the endogenous variables Y_3 and Y_4. More specifically, Y_3 displays a variable that serves simultaneously as an independent variable and as a dependent variable. Nevertheless, each dependent variable is considered endogenous. Where a latent variable is located in the inner model, for example, as an exogenous or endogenous variable, more so whether it is excluded or included in the model, and what relationships are drawn with other latent variables, depends on theoretical considerations, such as the underlying theory, logic, or practical experience of the researcher (Falk and Miller 1992).

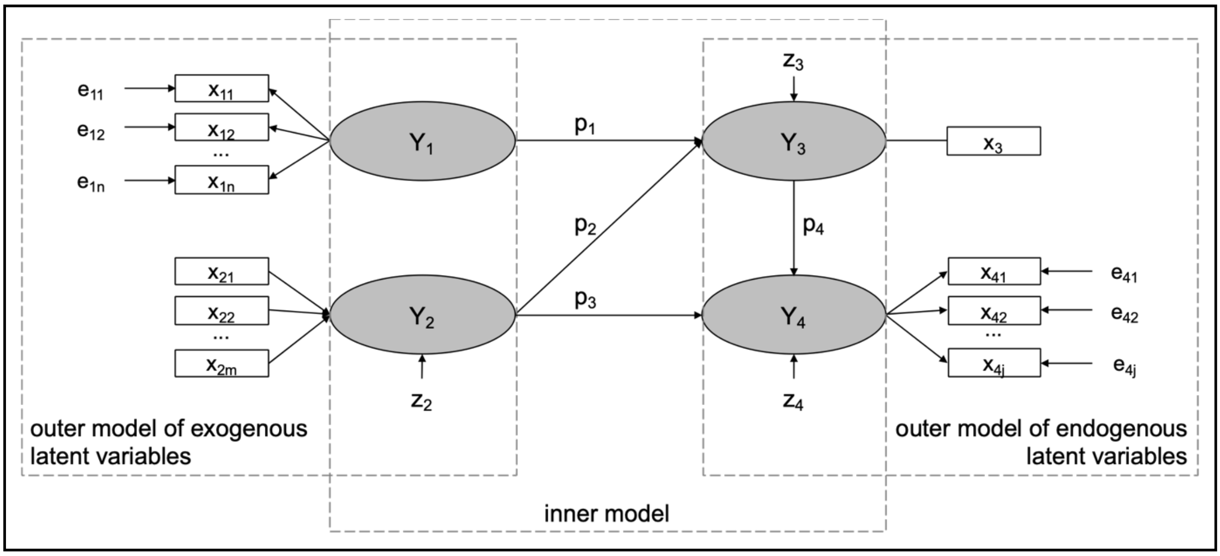


Figure 2.1: Scheme of a Path Model[11]

2.1.2.1 Direct Relationships

The relationship between two latent variables is determined by path coefficients (p_1, p_2, p_3, and p_4 in Figure 2.1), which indicate the strength of the respective relationship on a scale[12] from -1.0 to 1.0, whereby values close to zero indicate a (statistical) independence. These path coefficients are displayed by arrows in the inner models and point in the direction of the dependent variable. The size of the path coefficient is calculated by means of partial regressions in which the endogenous latent variable serves a dependent variable (e.g., Y_3 in Figure 2.1) and all direct predictor variables of this latent variable (e.g., Y_1 and Y_2 in Figure 2.1) are independent variables of the regression. This means that there is a separate partial regression model for each endogenous latent variable. The (standardized) regression weights indicate the respective path coefficients (e.g., p_1 for the relationship between Y_1 and Y_3 in Figure 2.1). Thus, the path coefficients have to be interpreted as linear estimates of their

[11] Adapted from Hair, Hult et al. (2017, p. 12).
[12] Path coefficients outside this range are technically possible, but might indicate collinearity issues (Sarstedt, Ringle, and Hair 2017).

respective relationships. Consequently, the explained variance (R^2) and error terms can be determined for each endogenous latent variable (e.g., Z_3 and Z_4 in Figure 2.1).

In the simplest form, a dependent variable is affected by only one other latent variable. In this case, the path coefficient can be interpreted as the correlation between these latent variables. However, in addition to the number of predictor variables in a direct path relationship (which simply determines the number of factors in the partial regression), more complex relationships between latent variables can be designed, such as mediated relationships or moderated relationships. The underlying research questions (respectively the theory-derived hypotheses) in the survey-based studies necessitate these more complex relationships, as they aim, for example, to uncover the intrapersonal processes linking feedback frequency to employee creativity (requiring a mediated relationship; see Study 1) or the influence of feedback use on the effectiveness of feedback frequency on employee creativity (requiring a moderated relationship; see Study 3). Therefore, mediated relationships and moderated relationships between latent variables are detailed in the next sections.

2.1.2.2 Mediated Relationships

A mediated relationship implies that a third (mediating) latent variable interrelates two latent variables (predictor and outcome). Consequently, the mediating latent variables serves as both an exogenous and an endogenous latent variable. Importantly, the effect of the predictor variable (e.g., Y_2 in Figure 2.1) onto the outcome variable (e.g., Y_4 in Figure 2.1) is channeled through the mediating variable (e.g., Y_3 in Figure 2.1). To illustrate this, the first study reveals (see chapter 3) that feedback frequency (predictor) leads to less recipient role ambiguity (mediator), which in turn influences recipient's creativity (outcome). Thus, feedback frequency has an indirect effect on recipient (i.e. employee) creativity via recipient role ambiguity. The size of the indirect effect is the product of the respective path coefficients (e.g., p_2 and p_4 of the indirect effect of Y_2 on Y_4 in Figure 2.1). Consequently, the sign of the indirect effect is also determined by these path coefficients: If both path coefficients have the same sign, the indirect effect is positive (as in the example of feedback frequency, role ambiguity, and creativity). Otherwise, the indirect effect is negative.

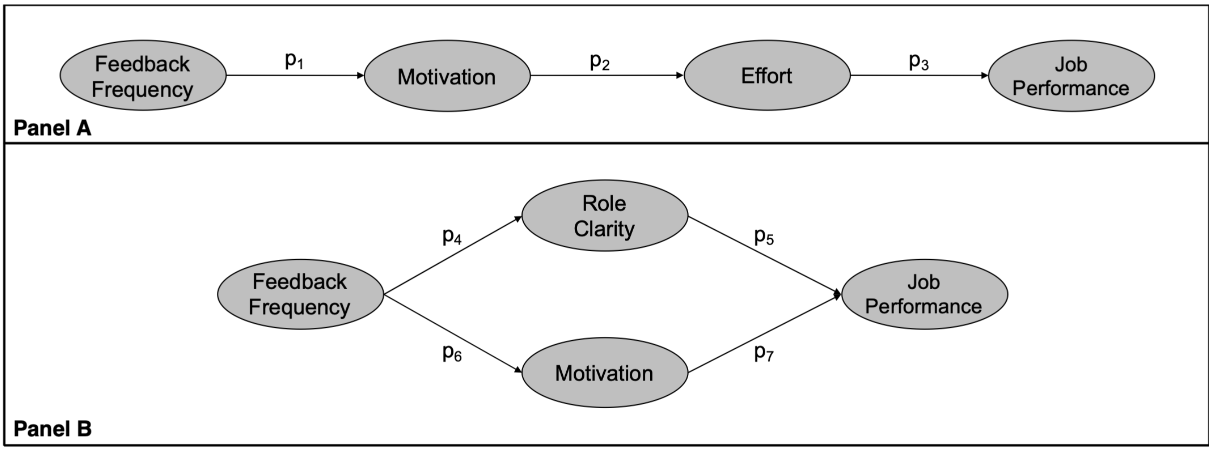


Figure 2.2: Illustrative Examples of more complex mediated Relationships

In the inner model, various sequences of mediating effects can be designed (Cepeda, Nitzl, and Roldán 2017; Nitzl, Roldán, and Cepeda 2016), such that a relationship can be transmitted through multiple mediators. For instance, as depicted in Panel A of Figure 2.2, feedback frequency (predictor) may lead to more motivation (mediator 1), which in turn leads to more effort (mediator 2), which ultimately affects job performance (outcome). In this case of a “serial” mediation, the indirect effect of feedback frequency on job performance is the product of p_1, p_2, and p_3. Besides, Panel B of Figure 2.2 displays an example of a “parallel” mediation, such that feedback frequency (predictor) directly affects recipient’s role clarity (mediator 1.1) and recipient motivation (mediator 1.2), both of which, and independently, impact recipient job performance. Here, the indirect effect (c) of feedback frequency on job performance is the sum of each independent mediations, i.e., c = p_4 x p_5 + p_6 x p_7. More generally, combinations of serial and parallel mediations are possible (e.g., in Panel B of Figure 2.2, an additional relationship could be drawn from role clarity to motivation). However, the rationale that the underlying theory determines the model design (Falk and Miller 1992) also applies to the selection (and structure) of specifically relevant mediators, such that different model design specifications (or restrictions) result from different theories (e.g., role theory or control theory).

In addition to the indirect path relationship, a direct path relationship between predictor and outcome variable is still possible (e.g., p_3 in Figure 2.1) but not necessary (e.g., the mediated relationship between Y_1 and Y_4 through Y_3 in Figure 2.1). As a result, one predictor may influence another variable through multiple pathways. The overall effect of one predictor variable on another variable is referred to as the total effect. The total effect of the predictor variable is composed of its direct and (potentially multiple) indirect path relationships on the respective outcome variable (Cepeda, Nitzl, and Roldán 2017; Nitzl, Roldán, and Cepeda 2016). Thus, there is a unique total effect between any two latent variables. In the extreme case, the total effect is equal to the direct effect (if there is no mediation).

Mediations within the inner model allow for a more sophisticated causal chain between predictor variable and outcome variable (Henseler, Hubona, and Ray 2016; Iacobucci, Saldanha, and Deng 2007; Shmueli et al. 2016). Moreover, the analysis of mediators indicates the likelihood of omitted variables (Zhao, Lynch, and Chen 2010). In other words, ignoring (potential) indirect effects would “bias the interpretation of the results” (Nitzl, Roldán, and Cepeda 2016, p. 1849). This could mean that the inclusion of a mediator reveals opposing effects that would otherwise remain hidden. To provide a practical example: For decades, research largely assumed detrimental effects on employee creativity arising from role conflict. However, when Bettencourt and Brown (2003) included job satisfaction as a mediator in their model, the results showed that role conflict has a positive (direct) effect for employee creativity.[13] According to these results, managers should not reduce employees' role conflict but, on the contrary, foster it to enhance employees' creativity while controlling job satisfaction respectively compensating for the adverse effects of role conflict on job satisfaction, since a reduction of job satisfaction would impair employees' creativity (i.e., the indirect effect of role conflict).

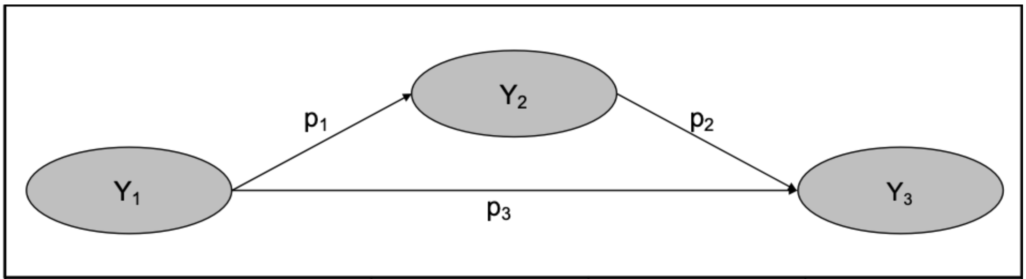
This practical example underlines the meaningfulness of a typology of mediations. The most commonly used framework to categorize mediation effects is that of Baron and Kenny (1986). Their approach is essentially based on the size of the indirect effect relative to the total effect, i.e., the variance accounted for (VAF).[14] According to their taxonomy, full mediation (VAF > 80%), partial mediation (80% ≥ VAF ≥ 20%), and no mediation (VAF < 20%) can be distinguished. Besides, they define two necessary prerequisites for a mediation: a significant indirect effect (e.g., p_1 x p_2 in Figure 2.3) and a significant direct effect, but with exclusion of the mediator from the inner model ("X-Y test"[15]).

However, more recent research has queried Baron and Kenny's (1986) approach due to conceptual and methodological flaws (cf. Hayes 2017; Preacher and Hayes 2004, 2008; Shrout and Bolger 2002; Zhao, Lynch, and Chen 2010). For instance, the rationale of examining only a significant path relationship between latent

---

[13] According to Schepers, Nijssen, and van der Heijden (2016) and Lages and Piercy (2012) "internal influence behaviors" (Bettencourt and Brown 2003; Bettencourt, Brown, and MacKenzie 2005) can be referred to as employee creativity.
[14] Considering the connotations in the mediation model shown in Figure 2.3, VAF = (p_1 x p_2) / (p_1 x p_2 + p_3).
[15] "X-Y test" refers to X and Y as traditional connotation for the independent and dependent variables, respectively. Thus, this test refers to the assessment of the condition whether the "effect to be mediated" is significant (Preacher and Hayes 2004; Zhao, Lynch, and Chen 2010).

variables for a potential mediation effect falls short in the case of competitive mediation, since the (hidden) indirect path may cancel out the direct effect of the predictor variable, causing the “X-Y test” to fail. Therefore, competitive mediations that are “of equal theoretical interest a priori” (Zhao, Lynch, and Chen 2010, p. 1999) suffer systematic underrepresentation. Besides, the use of the VAF as measure of the degree of mediation has raised concerns because “the strength of mediation should be measured by the size of the indirect effect, not by the lack of the direct effect” (Zhao, Lynch, and Chen 2010, p. 198). Moreover, the stepwise approach of ex-/including a mediator in a model biases the PLS path estimates (Nitzl, Roldán, and Cepeda 2016), so that falsely assumed (in)significances might be derived from the “X-Y test”. Lastly, Baron and Kenny (1986) suggest using the Sobel z-test (1982) for analyzing the significance of the indirect path. However, the product of two (assumed) normally distributed path coefficients (e.g., p_1 and p_2 in Figure 2.3) representing the indirect effect does not meet the parametric conditions (i.e., normal distribution) for the use of the Sobel z-test (Preacher and Hayes 2004, 2008).



Type of mediation	**$p_1 \times p_2 \neq 0$ (indirect effect)**	**$p_3 \neq 0$ (direct effect)**	**$p_1 \times p_2 \times p_3 > 0$**
indirect-only mediation	sig	ns	
complementary mediation	sig	sig	yes
competitive mediation	sig	sig	no
direct-only nonmediation	ns	sig	
no-effect nonmediation	ns	ns	

Note: notations in the table header refer to the mediation model shown above; shaded fields are irrelevant for the typology; sig = significant, ns = non-significant.

Figure 2.3: Typology of Mediations according to Zhao, Lynch, and Chen (2010)

Addressing this criticism, Zhao, Lynch, and Chen (2010) developed a novel classification of mediation types. In their decision tree, the significance of the indirect effect, the significance of the direct effect, and a comparison of the direction of both effects are examined in three stages. Figure 2.3 summarizes the results of the decision tree of Zhao, Lynch, and Chen (2010) in a table. From that, five types of mediations can be distinguished: First, the *indirect-only mediation* refers to a case where the direct effect (e.g., p_3 in Figure 2.3) is not significant whereas the indirect effect (e.g., $p_1 \times p_2$ in Figure 2.3) is significant. Hence, the effects of the predictor variable (e.g., Y_1 in Figure 2.3) on the outcome variable (e.g., Y_3 in Figure 2.3) are completely channeled through the mediating variable (e.g., Y_2 in Figure 2.3). In other words, the mediating variable "completely absorbs the positive or negative effect of [the predictor variable,] […] it can completely pass an effect or it can completely hinder the effect" (Nitzl, Roldán, and Cepeda 2016, p. 1855). At a more abstract level, an only indirect mediation indicates a good fit of the theoretical model and that omitted variables are unlikely (Zhao, Lynch, and Chen 2010).

Next, *complementary mediation* and *competitive mediation* are considered. They are characterized by the significance of both the indirect effect and the direct effect. Consequently, the predictor variable “still explains a portion of [the outcome variable] that is independent of [the mediator]” (Nitzl, Roldán, and Cepeda 2016, p. 1856). Zhao, Lynch, and Chen (2010) further distinguish between the homogeneity of direction (positive or negative) of the indirect effect and the direct effect. In a complementary (or consistent) mediation, both effects have the same direction. In a competitive (or inconsistent) mediation, both effects are opposing each other. Thus, the above-mentioned mediation in Bettencourt and Brown’s (2003) model is a competitive mediation. A priori, complementary mediation and competitive mediation are equally likely and “both identify an unexplained direct effect and guide future research to look for alternative mediators” (Zhao, Lynch, and Chen 2010, p. 1999; MacKinnon, Fairchild, and Fritz 2007).

Finally, two types of non-mediation (i.e., there is a non-significant indirect effect) can be specified. On the one hand, a *direct-only nonmediation* involves a significant direct effect between predictor variable and outcome variable. Even if the chosen mediator does not affect their relationship, an unrecognized mediator (or multiple) could still exist (Nitzl, Roldán, and Cepeda 2016; Shrout and Bolger 2002). Such a finding suggests that the underlying theoretical basis should be reconsidered (Zhao, Lynch, and Chen 2010). On the other hand, *no-effect nonmediation* is determined by both (indirect and direct) effects being non-significant. However, this nonmediation does not necessarily imply that the total effect of the predictor variable on the outcome variable is non-significant (Zhao, Lynch, and Chen 2010).

Despite the previously mentioned shortcomings, Baron and Kenny’s (1986) categorization and procedure is still used by some scholars (e.g., Nitzl and Hirsch 2016). Nonetheless, there is considerable common ground with Zhao, Lynch, and Chen’s (2010) typology in interpreting mediation analysis results: “Our complementary mediation overlaps with Baron and Kenny’s partial mediation; our indirect-only mediation overlaps with their full mediation” (Zhao, Lynch, and Chen 2010, p. 200), so there are different degrees of how mediation (the indirect path) explains the outcome variable. At the conceptual level, the biggest difference is how the direct path is handled. While for Baron and Kenny (1986) the significance of the direct path (i.e., when mediation is excluded from the model) is a prerequisite for mediation analysis, Zhao, Lynch, and Chen (2010) see the direct effect as a distinct dimension and thus as a source for further research or model optimization: Baron and Kenny’s (1986) “full-partial-no scale assumes one dimension [, however,] proper interpretation of one’s data requires two dimensions for the indirect path and the direct path” (Zhao, Lynch, and Chen 2010, p. 200), so that “unexplained direct effects may turn from irritation to inspiration” (Zhao, Lynch, and Chen 2010, p. 205).

Conclusively, the inclusion of mediators in the inner model (structural model) has the potential to contribute substantially to model sophistication and interpretation. Nonetheless, the research question and underlying theory preliminary drive the model design; and hence the design of mediators. Similarly, conducting an analysis of mediators depends on whether “corresponding hypotheses have been formulated” (Hair, Ringle, and Sarstedt 2013, p. 4). This is reflected in current publication practices in management accounting journals (including The Accounting Review, Journal of Management Accounting Research, Journal of Accounting Research, Contemporary Accounting Research, and Accounting, Organizations and Society) that utilize PLS-

SEM (Nitzl 2016): Although these studies regularly use complex models (on average, more than six latent variables and eleven path relationships) – suggesting that mediations are commonly part of model design – only less than one third (32.4%) conduct an explicit mediation analysis. This also applies to publications in other research fields such as information management (23.1% of studies that utilize PLS-SEM in Management Information Systems Quarterly (MISQ) perform a mediation analysis; Ringle, Sarstedt, and Straub 2012), marketing (34.2% of such studies in the Journal of Marketing, Journal of the Academy of Marketing Science, and Journal of Marketing Research; Ringle, Sarstedt, and Straub 2012), or strategic management (Long Range Planning; Hair, Ringle, and Sarstedt 2013).

Although no explicit mediation hypotheses are formulated in either survey-based study in this dissertation, mediation analyses are employed providing assurance about the robustness of the findings. For instance, by drawing on the typology of Zhao, Lynch, and Chen (2010) Study 3 reveals that higher creative requirements on employees translate fully into higher employee creativity only if supervisors respond to this change by increasing the frequency of their feedback (complementary mediation).

2.1.2.3 Moderated Relationships

Moderated relationships indicate that there is no consistent linear relationship between the exogeneous and the endogenous variable, but the strength (or direction) of the relationship is determined by the level of a third variable, the moderator (Baron and Kenny 1986). To provide a practical example for such a relationship, Study 1 shows that supervisor feedback frequency (exogeneous variable) has different effects on employee creativity (endogenous variable) depending on employee’s professional experience (moderator). Similar to other facets of SEM model design, moderated relationships need to be hypothesized a priori (Hair, Hult et al. 2017). Thus, the underlying theory does not only determine the selection of the moderator, but also whether a single path relationship, multiple path relationships, or all path relationships within the model are moderated. In PLS-SEM, both observable (e.g., years of professional experience) and unobservable (e.g., job satisfaction) variables can be employed as moderators.

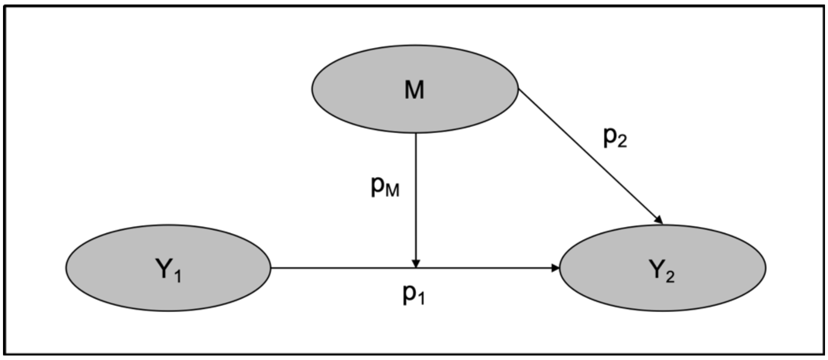


Figure 2.4: Conceptual Illustration of a moderated Path Relationship[16]

Figure 2.4 visualizes a moderated path relationship: The moderator variable (M) affects (p_M) the path relationship (p_1) between two latent variables (Y_1, Y_2). The moderating effect is depicted by the arrow pointing to the moderated relationship. The mathematical formulation of this moderating relationship[17] helps to distinguish multiple effects:

$$Y_2 = (p_1 + p_M \times M) \times Y_1 + p_2 \times M.$$

First, the relationship between Y_1 and Y_2 remains linear. However, the slope of this linear relationship depends on the size of the moderating effect (p_M) at a given level of the moderator. Second, the simple effect (p_1) indicates the strength of the relationship between the exogenous and the endogenous variable if the moderator is zero (due to standardized values in PLS-SEM, this means that the moderator equals

[16] Adapted from Henseler and Fassott (2010, p. 717).
[17] To improve readability, the intercept and error term are omitted in the following paragraphs.

to its mean value). In other words, “if the moderator variable is one, i.e., one standard deviation higher than its mean, the exogenous variable’s influence on the endogenous variable is $p_1 + p_M$” [18] (Henseler and Fassott 2010, p. 747). Hair, Ringle, and Sarstedt (2013) point out the distinction between the main effect and the simple effect. While the main effect indicates the strength of the relationship between exogenous and endogenous variable if “all other independent variables remain constant” (Hair, Ringle, and Sarstedt 2013, p. 3), i.e., there is only a direct path relationship with no moderation effect, the simple effect indicates the strength of the similar relationship if the moderator “has a value of zero, and all other independent variables remain constant”[19] (Hair, Ringle, and Sarstedt 2013, p. 3). Third, the direct effect (p_2) of the moderator (M) on the endogenous variable (Y_2) contributes only to the intercept of the partial regression and therefore does not affect the relationship between the exogenous and endogenous variable.[20] However, excluding this path relationship, often referred to as “reduced model”, does not adequately reflect the moderating effect as it “would overestimate the size of the moderating effect” (Henseler and Fassott 2010, p. 719; Carte and Russell 2003; Cohen 1978; Cronbach 1987; Irwin and McClelland 2001).

Furthermore, by rearranging the equation for the moderated relationship, an interaction term can be established:

$$Y_2 = p_1 \times Y_1 + p_M \times (M \times Y_1) + p_2 \times M.$$

The interaction term consists of the product of the exogeneous variable and the moderator variable $M \times Y_1$. Most importantly, the interaction term allows to

[18] In the nomenclature of Figure 2.4.
[19] As mentioned earlier in this paragraph due to standardized values in PLS-SEM, “zero” means in this case that the moderator equals to its mean value.
[20] While this is true at the conceptual level, the PLS-algorithm statistically accounts for this path relationship similarly to other direct path relationships in model estimation.

operationalize the conceptual moderation into a PLS-SEM, since only direct path relationships between latent variables are technically possible (Hair, Hult et al. 2017; Henseler and Fassott 2010). Consequently, also interactive effects can be designed in PLS-SEM[21]: As displayed in Figure 2.5, an auxiliary latent variable is added to the inner model reflecting the interactive effects of the exogenous variable and moderator. So, from a mathematical (statistical) point of view, it is irrelevant which variable represents the moderator. However, this should not be misinterpreted to imply that interaction and moderation, or moderator and exogenous variable are conceptually interchangeable: "Although the statistical technique for interaction and moderation is identical […], there is a subtle difference in the intention or interpretation" (Wu and Zumbo 2008, p. 381).

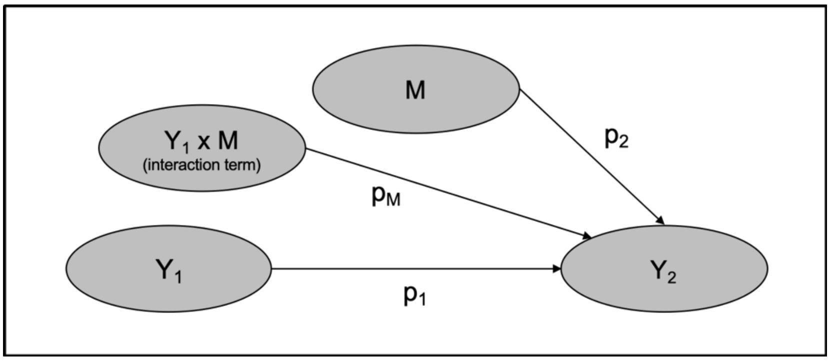


Figure 2.5: Implementation of a Moderation in the Inner Model[22]

As mentioned earlier, a moderation describes a relationship between an exogenous variable and an endogenous variable whose causation is contingent on a third variable (the moderator). A moderation, therefore, "seeks to understand *when* X

[21] In other words, designing an interaction or a moderation is the same in PLS-SEM.
[22] Adapted from Hair, Hult et al. (2017, p. 248).

influences Y or *for whom* the relationship exists or varies in strength or sign"[23] (Jollineau and Bowen 2021, p. 2).

In contrast, an interaction describes a joint effect of two exogenous variables on an endogenous variable. Thus, in addition to the unique effects of each exogenous variable, they also have a combined effect on the endogenous variable. For example, financial incentives and feedback frequency (exogenous variables) may both have a positive impact on employee creativity (endogenous variable), but their interaction could be negative due to crowding out their favoring effects on motivation. Accordingly, feedback frequency would be particularly effective for employee creativity when financial incentives are low, while the same would be true for financial incentives when feedback frequency is low. In other words, "*how much* X_1 affects Y is conditional on the value of X_2 and *how much* X_2 affects Y is conditional on the value of X_1"[24] (Luft and Shields 2003, p. 174; Hartmann and Moers 1999). Consequently, an "interaction can be viewed as a moderator effect and either independent variable can be viewed as the moderator" (McClelland and Judd 1993, p. 377).

Unlike the added (interacting) exogenous variable, the moderator variable conceptually has "no influence on Y in the absence of X_1, as well as no influence on X_1: its influence operates only by changing the effect of X_1 on Y"[25] (Luft and Shields 2003, p. 174). For example, Study 1 reveals that recipients' job experience influences (moderates) the effect of feedback frequency on employee creativity. Since a direct influence of employee experience on employee creativity (and feedback frequency) is neither supported by empirical research (Ng and Feldman 2008, 2013) nor by feedback

[23] Italics and nomenclature were retained from the original text, X denotes the exogenous variables, Y the endogenous variable.
[24] Italics and nomenclature were retained from the original text, X_1 and X_2 denote exogenous variables, Y the endogenous variable.
[25] Nomenclature was retained from the original text, X_1 denotes the exogenous variables, Y the endogenous variable.

theory, this clearly argues for interpreting this relationship as a "causal interaction effect" (Wu and Zumbo 2008), i.e., as a moderation.

In general, there is still ambiguity about the "true" moderator and the "true" exogenous variable in moderated relationships, especially when the empirical results also indicate a significant direct effect of the moderator variable on the endogenous variable. Thus, research should strive to "theoretically rule out the reverse interaction in which the independent variable is moderating the relationship between the moderating variable and the dependent variable"[26] (Andersson, Cuervo-Cazurra, and Nielsen 2014, p. 1067). Although only "the focus of the intended interpretation is somewhat different" (Wu and Zumbo 2008, p. 381), I follow the extensive PLS-SEM research by consistently using the terms "moderation effect" as well as "interaction term" when referring to its representation within the inner (structural) model.

The choice of the moderator variable is usually associated with the type of moderation (Henseler and Chin 2010; Rigdon, Ringle, and Sarstedt 2010). Continuous moderation is often a reasonable design choice for moderators measured with an interval scale.[27] In case of the previous example, continuous moderation would mean that with each year[28] of professional experience, the effect of supervisor feedback frequency on recipients' job motivation diminishes. Depending on the effect size of the moderation and the range of values of the moderator variable[29], there is a threshold

[26] Nomenclature was eliminated.
[27] An interval scale is characterized by similar differences between its units, for example the Celsius scale. Despite the fact that the Likert-scale is frequently used as an interval scale (including in PLS-SEM), it is actually an ordinal scale (cf. Wu and Leung 2017). However, I follow the widely accepted notion that the Likert-scales are a sufficiently good approximation of an interval scale and apply practical recommendations in survey design in this regard, such as equidistant scale points (cf. Hair, Hult et al. 2017).
[28] Strictly speaking, this principle applies to any unit of time.
[29] Mathematically, a continuous moderation always has a root, as it represents a linear function with a nonzero slope. In research practice, however, the value for the moderation variable at the root may be unrealistic or impractical, such as a professional experience of 100 years or values above 7 on a 7-point Likert-scale.

value of the continuous moderation at which the effect of the exogeneous variable on the endogenous variable is zero respectively changes its sign. Nonetheless, a moderation that 'solely' changes the magnitude of the exogenous variable's effect on the endogenous variable does not necessarily have a smaller contribution to the existing literature.

A categorical moderator is usually employed if the moderator variable is nominal (e.g., the industry of the feedback recipient's firm). Due to the properties of the variable, an incremental change of the moderator variable is not possible. However, this is a prerequisite for the application of continuous moderation. Instead, the drawn sample is split into two or more subsamples depending on the number of unique values of the moderator variable. The next step is to compare the strength of the relationships between exogenous and endogenous variables across all subsamples. If substantial differences arise, the moderation is established. In PLS-SEM, parametric tests, such as the Welch-Satterthwaite t-test (Satterthwaite 1946; Welch 1947), or nonparametric tests, such as PLS-Multigroup Analysis (PLS-MGA) (Henseler 2012; Henseler, Ringle, and Sinkovics 2009) and the permutation test (Chin 2003; Chin and Dibbern 2010), can be used to assess the significance of differences in path relationships.[30] The results can be read independently or comparatively. For instance, feedback frequency may have a non-significant effect on employee creativity for employees in banks (first subsample), while this effect may be positive (independent) / larger (comparative) for employees in marketing firms (second subsample).

[30] Since parametric tests rely on distributional assumptions that conflict with PLS-SEM principles and are generally more lenient, nonparametric tests should be preferred when conducting multigroup analyses (Sarstedt, Henseler, and Ringle 2011).

Consider a path relationship Y1 ➔ Y2, such as displayed in Figure 2.4, moderated by the variable "annual sales" as a proxy for firm size. Usually, one would consider to design a continuous moderation, since “sales” satisfies the properties of an interval scaled[31] variable. Nevertheless, an application of categorical moderation is possible by defining value segments into which the continuous variable is subdivided. In case of an observable variable such as sales, practical considerations may determine the definition of segments (e.g., the formal classification of micro, small, and medium-sized enterprises in terms by the European Commission[32] based on annual sales figures). Otherwise, statistical criteria can also define segment boundaries, such as a mean-split, a median-split, or a quartile-split. For unobservable variables that can equally serve as moderators, the latter segment definitions are most common. Similarly, ordinal and nominal (e.g., dichotomous) scaled variables can be used in a continuous moderation. Usually, this is done by means of dummy variables whose coding reflects the variable values (e.g., 0 = unlisted company; 1 = listed company) (Hair, Hult et al. 2017).

Although the properties of the moderator variable play an important role in deciding whether the moderation is designed to be continuous or categorical, there are further factors that influence the model design choice. First, theory might suggest that the relationship $Y_1 \rightarrow Y_2$ is substantially stronger for small enterprises (i.e., firm size moderates $Y_1 \rightarrow Y_2$). In order to reflect or test the theory in the path model, a categorical moderation analysis is required in which the sample is divided into small enterprises and other (non-small) enterprises. Second, the underlying data distribution

---

[31] Sales numbers also fulfill the properties of a ratio scaled variable. In addition to the properties of an interval scaled variable, a ratio scaled variable has a non-arbitrary zero value and thus has meaningful ratios (Stevens 1946). However, with regard to the application in PLS-SEM, a further distinction between ratio scaled and interval scaled variables is not necessary.
[32] See EU recommendation 2003/361, https://ec.europa.eu/growth/smes/sme-definition_en [retrieved 03-21].

might not be sufficiently well represented by a continuously (linear) moderated path relationship. For instance, there might be certain thresholds for sales values at which the relationship Y_1 ➔ Y_2 substantially alters its size or direction (i.e., firm size moderates Y_1 ➔ Y_2), such as is the case with piecewise constant functions (e.g., a step function). Splitting the sample according to these sales thresholds (i.e., converting sales into a categorical moderator) allows for a better representation of the moderating effect or is even necessary for its detection. Third, it may be helpful for understanding the results and "getting the message through" if distinct results of the relationship Y_1 ➔ Y_2 are given for clearly delineated revenue ranges. Practitioners, in particular, are likely to be more interested in accessible and seemingly vibrant groupings of firms such as micro, small, medium-sized, and large enterprises with which they can identify than in incremental changes in the relationship Y_1 ➔ Y_2 with each variation in the moderator variable relative to its standard deviation. In addition, when interpreting complex[33] path models, also researchers can benefit from converting a continuous moderation into a categorical moderation.

These considerations underlie the design decisions (i.e., categorical vs. continuous) of the moderations in the SEMs of Study 1 and Study 3. For example, the moderator employee experience (although measured on a continuous scale) is included in Study 1 as a categorical moderator in the SEM to reduce complexity in the inner (moderated-moderated-mediated) model.

[33] More complex models could include moderated mediations, parallel mediations so that different effect pathways emerge, moderation of multiple paths, multiple moderations of individuals paths, or interaction effects between moderators, which increases the complexity of model analysis.

At first glance, it may seem that information is lost when categorical moderation is used due to the simplification of reducing the variety of variable values.[34] This impression is erroneous, however, as categorical moderation is occasionally helpful in providing a more sophisticated model design – both from a theoretical perspective and in terms of results. These considerations are reflected in current PLS-SEM publication practices in the field of management accounting (16.2% included a categorical moderator and 8.1% included a continuous moderator, Nitzl 2016), information management (24.6% vs. 12.3%, Ringle, Sarstedt, and Straub 2012), or marketing (23.0% vs. 7.4%, Hair et al. 2012).[35] However, researchers have to be aware that results are sensitive to the moderation design (e.g., the segmentation) and that "there is little guidance on choosing the best procedure" (Rigdon, Ringle, and Sarstedt 2010, p. 265; Jedidi, Jagpal, and DeSarbo 1997; Squillacciotti 2010).

Consequently, as the SEMs in this dissertation involve both categorical and continuous moderations, the robustness of these design choices is tested (e.g., by altering the modes of moderation) to avoid artificial effects resulting from these choices. Nonetheless, the identified moderating effects are also indicated when other approaches are applied.

2.1.2.4 Unobserved Heterogeneity

Since novel empirical techniques for identifying unobserved heterogeneity are extensively used in Study 1 (Hair et al. 2016; Matthews et al. 2016), this chapter provides an overview of conceptual and fundamental methodological considerations

---

[34] Only if the number of segments equals the number of unique variable values, e.g. if two segments are obtained from a dichotomous variable, there is no loss of information.

[35] Surprisingly, these proportions/ratios are nearly reversed in the leading marketing journals (i.e., Journal of Marketing, Journal of Marketing Research, and Journal of the Academy of Marketing Science), as 12.2% of all PLS-SEMs included a categorical moderator and 17.1% of all PLS-SEMs included a continuous moderator (Ringle, Sarstedt, and Straub 2012).

for identifying unobserved heterogeneity. As the procedures are particularly applicable to unveil heterogenous relationships among latent variables and their (ultimate) implementation in the SEM is similar to moderated relationships, this chapter builds on and expands the preceding elaboration on moderations (i.e., observed heterogeneity).

PLS-SEM implicitly assumes that “the data stem from a single population” (Hair et al. 2012, p. 427; Jedidi, Jagpal, and DeSarbo 1997) and pools observations when estimating parameters (Becker et al. 2013). However, “in many real-world applications, this assumption of homogeneity is unrealistic” (Rigdon, Ringle, and Sarstedt 2010, p. 256). Instead, the observed participants are heterogenous, for instance regarding prior experiences, perceptions, attitudes, or attributes. When analyzing data, there is an eligible concern that underlying heterogeneity will bias the results. With respect to the inner model, these differences may yield that a path relationship has opposite signs in heterogenous subgroups, that it has the same sign but different size, or that it is (non-)significant only for a fraction of these subgroups (Becker et al. 2013). If this heterogeneity remains uncontrolled, it reduces the predictive power of the model and can “lead to Type I and Type II errors and invalid inferences” (Becker et al. 2013, p. 669), such as the overgeneralizing of a significant (Type I error) or a non-significant (Type II error) path relationship in the overall sample that exists, however, only in a fraction of the heterogenous subgroups (i.e., the Type I and Type II errors occur in the respective non-/significant subgroups). Thus, heterogeneity poses a threat to validity.

The principle of controlling for heterogeneity in SEM is based on assigning heterogenous observations into homogenous subgroups (Rigdon, Ringle, and Sarstedt 2010). Therefore, identifying adequate subgroups is of utmost importance. Observed heterogeneity is referred to when “existing theory that incorporates moderators or contextual factors” (Becker et al. 2013, p. 668) determines the definition of subgroups. Then, a priori considered variables are expected to be causal and used to test for heterogeneity, such as a known third variable (moderating effect) (Rigdon, Ringle, and Sarstedt 2010; Sarstedt, Henseler, and Ringle 2011).[36] Often, these variables must be integrated during the development of the model, which adds another facet to the a priori deliberations.

Complementarily, unobserved heterogeneity occurs “when theory does not assume heterogeneity even though it exists or when theory indicates heterogeneity but the specified group variables do not sufficiently capture it in the population” (Becker et al. 2013, p. 668). Then, heterogeneity “masks group-specific effects” (Hair, Sarstedt et al. 2017, p. 175) and “its true sources are unknown” (Rigdon, Ringle, and Sarstedt 2010, p. 269). Ex-post analysis can be used in an attempt to uncover unobserved heterogeneity. However, despite its ease of use, traditional exploratory segmentation (e.g., k-means clustering) is conceptually flawed since it does not account for relationships between the inner (path) and outer (measurement) models (Hair, Sarstedt et al. 2017; Sarstedt and Ringle 2010).

In PLS-SEM, a highly sophisticated approach has been developed to overcome the methodological weaknesses of traditional clustering and to provide a systematic test for and uncover unobserved heterogeneity. For these purposes, a four-step[37] procedure is recommended (Becker et al. 2013; Hair, Sarstedt et al. 2017; Matthews et al. 2016; cf. Figure 2.6):

[36] The conceptualization of moderating effects has been described in previous sections and will not be revisited here.
[37] I divide the entire recommended procedure into four steps.

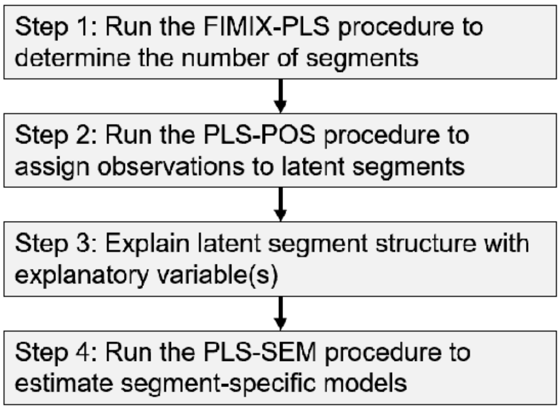


Figure 2.6: Systematic Procedure of jointly applying FIMIX-PLS and PLS-POS[38]

First, finite mixture partial least squares (FIMIX-PLS; Hahn et al. 2002; Sarstedt, Becker, Ringle, and Schwaiger 2011) is applied to determine the number of underlying segments. FIMIX-PLS tries to “disentangle the overall mixture distribution and estimate parameters (e.g., the path coefficients) of each group in a regression framework (i.e., mixture regression)” (Hair, Sarstedt et al. 2017, p. 66). FIMIX-PLS’s strengths lie in capturing heterogeneity in the inner model (structural model) and in providing indicators for the optimal number of segments (Becker et al. 2013). This is crucial for dealing with unobserved heterogeneity, as under- or over-segmentation biases the results and yields invalid inferences (Becker et al. 2013).

FIMIX-PLS offers several model evaluation criteria, such as the Akaike’s Information Criteria (AIC, Akaike 1973) or Bayesian Information Criterion (BIC, Schwarz 1978), which are computed for each predetermined number of segments. Generally, lower values of these criteria indicate a better segment solution. In particular, Sarstedt, Becker, Ringle, and Schwaiger (2011) tested the performance and robustness of these criteria and recommend to jointly use the Modified AIC with factor 3 (AIC3, Bozdogan 1994) and the Consistent AIC (CAIC, Bozdogan 1987). In addition,

[38] Adapted from Hair, Hult et al. (2017, p. 179).

the value of the normed entropy statistic (EN, Ramaswamy et al. 1993) should exceed 0.50, indicating that the segments are sufficiently clear-cut (Ringle, Sarstedt, and Mooi 2010). It is important to note that these criteria not necessarily coincide and may provide different recommendations for the optimal number of segments, so that "a purely data-driven approach provides only rough guidance regarding the number of segments to select" (Hair, Sarstedt et al. 2017, p. 185).

In addition to the statistical model-specific heuristics, the adequate number of segments is thus co-determined by three necessary segment attributes (Becker et al. 2013, p. 685 ff.). One, the segments need to be substantial. Segments that contain only a small proportion of all observations often cannot be analyzed for reliability issues and are generally prone to "represent data idiosyncrasies (e.g., outliers and bad respondents)" (Becker et al. 2013, p. 685). Irrelevant segments are excluded from further analysis. Nevertheless, care must be taken when deciding whether to exclude segments to avoid confusing a niche segment worthy for further analyses with an irrelevant segment or a manifestation of statistical artifacts. Two, substantial segments need to be differentiable. If there are no significant differences in path relationships among multiple segments, a smaller number of segments should be given preferential consideration to avoid the risk of over-segmentation. Since homogenization by FIMIX-PLS segmentation does not affect the results, the conclusion is that the observations are relatively homogeneous with respect to the estimated model and there should be little concern about validity due to unobserved heterogeneity. Three, differentiable segments need to be plausible. Segment plausibility comprises segment-specific characteristics, conceptual differences between the segments, and theoretical or managerial relevance. An implausible segment may indicate theoretical limitations or its irrelevance (i.e., the segment was incorrectly considered to be substantial). It is important that implausible segments do not become part of a combined sample, as it is these segments in particular that compromise the validity of the conclusions drawn from model estimation and hypotheses evaluation.

If the results of the FIMIX-PLS procedure indicate that there is only one underlying segment, even after other segments have been excluded, the test is complete and provides confidence that the results of the PLS-SEM are robust to biases arising from unobserved heterogeneity. Otherwise, more steps need to be taken, since FIMIX-PLS's ability to correctly assign observations into groups is limited (Becker et al. 2013).

Second, PLS prediction-oriented segmentation (PLS-POS) (Becker et al. 2013; Matthews et al. 2016) is applied to assign observations to the predetermined number of segments. In this process, PLS-POS aims to maximize the explained variance (R^2) of all or a single endogenous variable and thus also the predictive power of the path model. As a result, each observation is assigned to one specific latent segment. Therefore, PLS-POS paves the way for assessing whether the plausible segments are also accessible (Becker et al. 2013).

Third, the latent segment structure solution derived from applying PLS-POS is converted into a segment-specific model whose segmentation is defined "in terms of managerially meaningful variables" (Rigdon, Ringle, and Sarstedt 2010, p. 278; Becker et al. 2013; Hair, Sarstedt et al. 2017; Matthews et al. 2016). Thus, based on statistical properties, theoretical and practical considerations, latent segmentation is attempted to be explained by one or more explanatory variable(s).

Fourth, it is not the PLS-POS results but the identified explanatory variable(s) that determine(s) the ultimate segments used as distinct (sub)samples for model estimation with PLS-SEM. Importantly, the explanatory variable(s) shall reflect the latent segment structure derived from PLS-POS. Therefore, it is expected that the effect sizes in the segment-specific models will be lower than when latent segmentation is used for model estimation. However, the PLS-SEM results always display average reflections of the relationships between the latent variables, so the variances in the path relationships are incorporated into the variable estimation. Moreover, only a division into tangible segments allows to draw understandable and actionable conclusions, and only by identifying the explanatory variable can adaptions be made in the theory and unobserved heterogeneity be "turned into observed heterogeneity for future studies" (Becker et al. 2013, p. 668).

Still, the search for the explanatory variable(s) is rather exploratory and does not necessarily lead to satisfactory results since either a complex variable combination or variables that have not been surveyed may reflect the underlying segmentation. Nonetheless, the number of latent segments (derived from FIMIX-PLS) and the segmentation assignments (derived from PLS-POS) provide guidance in the search for the explanatory variable(s).

Due to its novelty and complexity, it is not surprising that FIMIX-PLS has not been applied in many publications (cf. Hair et al. 2012; Ringle, Sarstedt, and Straub 2012). However, an emerging trend in publication practice indicates that finite mixture techniques are used in various ways[39]: First, FIMIX-PLS is run to check the robustness of the PLS-SEM results. For instance, Swoboda, Puchert, and Morschett (2016, p. 464) "believe that the likelihood of unobserved heterogeneity is reduced in this study" since the indicators suggest a one-segment solution. Second, the FIMIX-PLS segment assignments are characterized by their differences in the inner model. Ringle, Wende,

[39] From a methodological perspective, many of these publications utilizing FIMIX-PLS are flawed because they use segmentations derived from FIMIX-PLS to estimate model variables that do not account for heterogeneity in the outer model (Becker et al. 2013). In addition, the use of variables to only describe, but not determine, the different segments does not make the results actionable: "It should not be a process in which the best discriminating "left-over" variable in the dataset (that is not part of the model) is used to explain segment differences" (Becker et al. 2013, p. 687), but rather a process in which an explanatory variable(s) ultimately "determine[s] to which segment responses belong" (Becker et al. 2013, p. 687).

and Will (2010) identify price-oriented customers ("this segment is characterized by a strong relationship between Price[40] and Satisfaction and weak relationship between Quality and Satisfaction", Ringle, Wende, and Will 2010, p. 201) and quality-oriented customers ("this segment is characterized by a strong relationship between Quality[41] and Satisfaction and a weak relationship between Price and Satisfaction", Ringle, Wende, and Will, p. 201) based on the segment differences in the relationships between latent variables in the model. Third, additional data can be used to describe the segments derived from finite mixture results (Janka and Guenther 2018; Kumar, Dass, and Topaloglu 2014; Mancha et al. 2014). In the article by Janka and Guenther (2018) that introduced FIMIX-PLS to management accounting research, a two-segment solution uncovers complementary responses of firms to perceived environmental uncertainty in the context of new product development: "Whereas the first group of firms increases all management controls when perceived environmental uncertainty increases, the second group responds with a loosening of management control" (Janka and Guenther 2018, p. 131). The authors subsequently test for differences in the descriptive data and find that "the first group consists of slightly larger and older firms than the second group, with relatively emergent innovation strategies, greater innovation capabilities, and weaker perceived complexity of their environment" (Janka and Guenther 2018, p. 131). These attributes not only help to make the segments more tangible, but also to increase plausibility and to integrate the results with the existing literature. Similarly, Mancha et al. (2014) use descriptive data to provide further insights into the composition of their four-segment solution of auction bidders derived by applying FIMIX-PLS. However, the authors go beyond a trivial focus on significant differences in the values of individual variables and even derive a

[40] Construct names are in italics in the original, their capitalization has been retained.
[41] Construct names are in italics in the original, their capitalization has been retained.

taxonomy of bidders by interpreting descriptive segment data holistically ("hedonic bidders", "participators", "vigilant bidders", and "naïve bidders") which offers a tangible picture of "average" segment members in presence of more complex segment structures. Fourth, some publications even use finite mixture to test for hypothesized heterogeneity (Grewal et al. 2013; Ratzmann, Gudergan, and Bouncken 2016; Wilden and Gudergan 2015). Wilden and Gudergan (2015), for instance, use this ex ante approach to account for environmental turbulence in their model, which they assess with different facets of turbulence, such as technological turbulence, market turbulence, and competitor turbulence. After FIMIX-PLS indicated a two-segment solution, the authors note that "Subgroup 1 contained firms acting in environments with high technological and competitor turbulence; Subgroup 2 firms functioned in relatively stable environments" (Wilden and Gudergan 2015, p. 193). Here, the FIMIX-PLS procedure enables the specification of facets of environmental turbulence that cause different relationships between dynamic capabilities and the operational marketing and technological capabilities of the firms.

Reflecting these approaches in previous literature, the identification and treatment of unobserved heterogeneity in Study 1 is, to the best of my knowledge, the first to undergo all the procedures described in the methodological literature and fully complies with the methodological recommendation (Becker et al. 2013; Hair et al. 2016; Matthews et al. 2016).

2.1.3 The Outer Model – Measuring Latent Variables

In PLS-SEM, latent variables are constructed with indicators (i.e., manifest variables or items). While the previous chapter elaborated the relationships between the latent variables (inner model), this section deals with the relationships between the latent variable and its measurement (outer model, see Figure 2.1). The first step in defining the measurement of a latent variable is to select adequate indicators that operationalize the latent variable.[42] In selecting indicators, researchers often rely on established and prevalidated scales (and make only minor adjustments, if any). However, if material changes are made to existing scales or if a suitable set of indicators is not available because, for instance, the latent variable was not surveyed in previous publications or there is a discrepancy between the measurement model (the outer model) and the construct definition[43], a new set of measures can also be created. In particular, I follow established guidelines (Diamantopoulos and Winklhofer 2001; MacKenzie, Podsakoff, and Podsakoff 2011) when I create a novel measurement for control use (i.e., interactive use and diagnostic use of feedback) (Tessier and Otley 2012).

An overview of PLS-SEM publication practice shows (see Figure 2.7) that, while a variety of measurement models are considered – consistent with the subsequent descriptions in this section, reflective measurement models are most commonly relied on when operationalizing constructs. The reviews also indicate that the outer model designs of the survey-based studies in this dissertation consonant with publication practice.

[42] Indicator selection and the mode of measurement determine the outer model ex ante. However, before the measurement is used to assess the inner model, it must satisfy several statistical properties ex-post, the non-fulfillment of which could necessitate further adjustments in the outer model. These are discussed in later sections.
[43] One example is the frequently discussed question of whether employee creativity should be measured by external observation or by means of self-assessment (Farmer, Tierney, and Kung-McIntyre 2003; Lages and Piercy 2012; Speklé, van Elten, and Widener 2017; Tierney, Farmer, and Graen 1999; Zhou and George 2001). In this case, the construct definition (e.g., whether creativity must necessarily be observable by others) drives not only the selection of items but also the data source (e.g., supervisor or employee).

Strand of literature	**Management Accounting** *(Nitzl 2016)*	**Information Management** *(Ringle, Sarstedt, and Straub 2012)*	**Marketing** *(Hair et al. 2012)*	**Top journals in Marketing** *(Ringle, Sarstedt, and Straub 2012)*
Share of models with only reflective measures	56.8%	42.2%	42.1%	30.0%
Share of models with only formative measures	0.0%	1.8%	6.4%	1.7%
Share of models with formative and reflective measures	21.6%	30.3%	39.6%	53.3%
Mean number of indicators per reflective measurement	3.8	3.6	4.0	3.6
Mean number of indicators per formative measurement	4.4	3.0	4.6	4.1
Share of models with single-item constructs	32.4%	47.7%	46.3%	51.2%
Mean total number of indicators in model	not reported	27.4	29.6	34.6

Figure 2.7: Overview of Publication Practices regarding Outer Model Design in different Strands of Literature

As multiple measurement models are utilized in the SEMs of Study 1 and Study 3, I encounter the need to specify the concepts of reflective measurement models, formative measurement models, single-item measurement models, hierarchical component measurement models, and the measurement of auxiliary latent variables (interaction terms). In addition, due to the threat of measurement model misspecification for validity (Bisbe, Batista-Foguet, and Chenhall 2007; Chenhall 2012; Jarvis, MacKenzie, and Podsakoff 2003) and to avoid (still) common mistakes in research practice (Hair et al. 2012; Hair, Ringle, and Sarstedt 2013; Nitzl 2016; Ringle, Sarstedt, and Straub 2012), I devote a section to the differentiation of reflective and formative measurement models.

2.1.3.1 Reflective Measurement Models

Once the indicators have been selected, the nature of the relationship between the latent variable and its indicator(s) has to be defined. In a reflective measurement model, the latent variable is considered to influence its multiple indicators (Hair, Sarstedt et al. 2017). A reflective measurement model is visualized by arrows pointing from the latent variable to its indicators (see Figure 2.8) and can be expressed by the following regression equations:

$$x_1 = l_1 \text{ x } Y_i + e_1$$

$$x_2 = l_2 \text{ x } Y_i + e_2$$

$$...$$

$$x_n = l_n \text{ x } Y_i + e_n$$

In a reflective measurement model, each indicator is estimated by one regression. The random measurement error (e_j) therefore occurs at the indicator level and is associated with the respective indicator. The indicator loading (l_j) is the regression coefficient and hence displays the strength of the relationship between the latent variable (Y_i) and one of its indicators (x_j). Consequently, a change in the value of the latent variable causes *all* indicators to change simultaneously according to their respective loading. In other words, reflective indicators “can be viewed as a representative sample of all the possible items available in the conceptual domain of the construct” (Hair, Sarstedt et al. 2017, p. 9; Nunnally and Bernstein 1994). For this reason, the reflective indicators of a latent variable should be highly correlated and also interchangeable, so that the ex-/inclusion of an indicator does not alter the latent variable's meaning.[44]

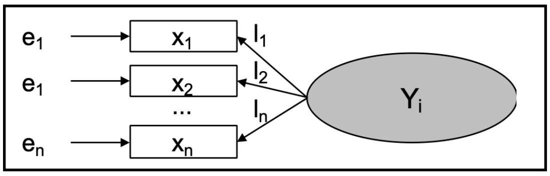


Figure 2.8: Latent Variable with Reflective Measurement Model

For instance, items such as "my supervisor often lets me know how well I am doing my job", "I always receive information from my supervisor about the results of my work", "anytime, I have the opportunity to get feedback from my supervisor about my job performance", and "my supervisor often lets me know about my job performance" all reflect the construct of supervisor feedback frequency. If a supervisor increases his or her feedback frequency, the respondent (the feedback recipient) is expected to rate all items with higher values (e.g., on a 7-point Likert-type scale ranging from "does not apply at all" to "fully applies").

2.1.3.2 Formative Measurement Models

Complementarily, a formative measurement model assumes that the indicators influence the latent variable (Hair, Sarstedt et al. 2017). A formative measurement model is visualized by arrows pointing from the indicators to the latent variable (see Figure 2.9) and can be expressed by the following linear combination regression equation:

$$Y_i = w_1 \text{ x } x_1 + w_2 \text{ x } x_2 + \ldots + w_n \text{ x } x_n + z_i$$

[44] Provided that the statistical properties (e.g., construct reliability) remain unaffected.

In a formative measurement model, the latent variable is estimated by regression. Although the random measurement error (z_i) occurs at the latent variable level, the character of the formative indicators conceptually determines its existence (Bollen 2011; Bollen and Bauldry 2011). On the one hand, *causal indicators* display the causes of the underlying construct. As it is expected that the indicators do not capture all the causes of the latent variable, the error term is considered to be non-zero (see left part of Figure 2.9). On the other hand, the linear combination of *composite indicators* fully represents the latent variable. Therefore, the error term is considered to be zero (see right part of Figure 2.9). Composite indicators form rather than cause the latent variable. Whether formative indicators are considered to be causal indicators or composite indicators, respectively whether the error term is assumed to be (non-)zero, has substantial implications for model estimation (Henseler et al. 2014) and is determined by the content of the indicators (Bollen 2011; Henseler, Hubona, and Ray 2016). In either case, the formative indicators are considered to be error-free (Diamantopoulos 2011).

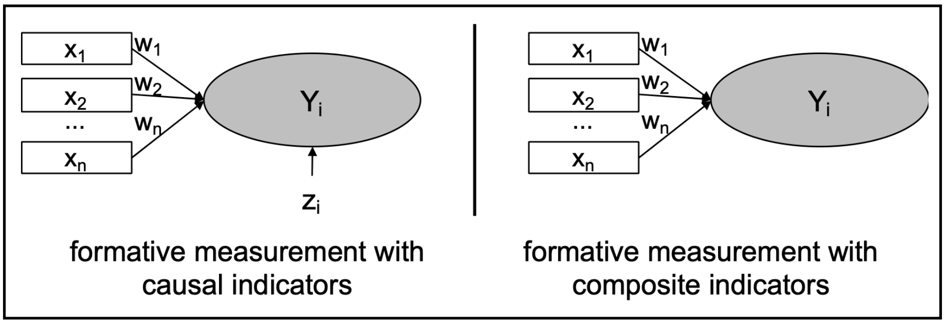


Figure 2.9: Latent Variables with Formative Measurement Model

The indicator weight (w_j) can be interpreted as the regression coefficient and hence represents the influence of one indicator (x_j) on the latent variable (Y_i). Consequently, a change in *one* indicator is associated with a change in the value of the latent variable according to indicator’s regression weight. As one indicator displays a distinct facet of all latent variable's causes, formative indicators are not interchangeable and not necessarily correlated. Latent variables with formative measurement models are therefore "inextricably tied to their measures" (Hair, Sarstedt et al. 2017, p. 10). Accordingly, the selection of formative indicators is of particular importance (Diamantopoulos and Winkelhofer 2001) and should cover all relevant facets of the latent variable (Nunnally and Bernstein 1994).[45]

Thereby, formative measurement models allow the relative importance of indicators to be assessed by comparing the size of their respective weights (Hulland 1999). Identifying specific success drivers can offer valuable practical and theoretical contributions to the literature, including management accounting research (Nitzl 2016). Moreover, formative measurement models are needed to integrate archival data into the SEM (Gefen, Rigdon, and Straub 2011; Rodgers and Guiral 2011).

To provide an illustrative example of a formative measurement, items such as "my supervisor frequently lets me know how well I am doing in acquiring new customers", "my supervisor frequently lets me know how well I am doing in retaining customers relationships", "my supervisor frequently lets me know, how well I am doing in documenting customer interactions", and "my supervisor frequently lets me know how well I am doing in working with my colleagues", all display distinct, rather non-redundant subjects of supervisor feedback frequency[46], which together form the latent variable. If a supervisor increases his or her feedback frequency on cooperating with colleagues, the respondent (the feedback recipient) is expected to rate the latter item

[45] This is referred to as content validity (Hair, Hult et al. 2017).
[46] Although it is conceptually possible, the operationalization of feedback frequency used in Study 1 and Study 3 adheres to previous survey-based measurement practices of feedback (e.g., Auh et al. 2019; Hackman and Oldham 1975; Kuvaas 2011) and is designed reflectively. Also, since the heterogeneous work environment complicate identifying and covering all non-redundant (formative) facets of feedback frequency that are also applicable to all employees.

with higher values and to perceive supervisor's feedback more frequently overall (which is reflected by a higher latent variable value), but not necessarily to rate *all* items with higher values (e.g., the items concerning the feedback frequency on customer-related issues). Furthermore, comparing the indicator weights allows to identify the facet of supervisor feedback frequency that contributes most to the overall latent variable.

2.1.3.3 Differentiating Reflective and Formative Measurement Models

Measurement model misspecification can bias the results and therefore poses a threat to the validity of the results (Jarvis, MacKenzie, and Podsakoff 2003). Particularly in management accounting research, the question of whether measurement is reflective or formative must be carefully considered (Bisbe, Batista-Foguet, and Chenhall 2007; Chenhall 2012). In management accounting research, models frequently include newly created constructs (Nitzl 2016) whose measurements, logically, have not yet been widely tested. Nevertheless, re-testing and refinements are important steps in the development of a measurement model to substantiate and improve its reliability and validity (Churchill 1979). Furthermore, research in management accounting faces the problem that the rather exploratory use of novel constructs is accompanied by a lack of strong theories (Nitzl 2016). Strong theories are in turn necessary for the development and assessment of indicators (Bagozzi 2011). A review of prominent accounting journals[47] substantiates these conceptual

[47] The article reviewed the following journals: Accounting, Organization and Society (AOS), The Accounting Review (TAR), Contemporary Accounting Research (CAR), Auditing: A Journal of Practice & Theory (AJAPT), Accounting, Auditing, & Accountability Journal (AAAJ), Behavioral Research in Accounting (BRIA), The International Journal of Accounting (TIJA), Journal of Accounting and Public Policy (JAPP), Management Accounting Research (MAR), Abacus, and Journal of Management Accounting Research (JMAR). The whole sample comprised of 66 SEM publications. There is no discernible pattern in the share of potentially misspecified models with respect to certain fields in accounting or accounting journals.

considerations by showing that 79% of all SEM publications potentially exhibit model misspecifications (Rodgers and Guiral 2011). Beyond the scope of management accounting research, methodological reviews across multiple literature strands consistently identify misspecified measurement models and stress the importance of avoiding them in future research (cf. Hair et al. 2012; Hair, Ringle, and Sarstedt 2013; Nitzl 2016; Ringle, Sarstedt, and Straub 2012). This involves not only the correct definition of measurement models, but also the application of adequate criteria[48] in their empirical evaluation. Still, criteria for reflective measurement models are often mistakenly used when assessing formative measurement models (cf. Hair et al. 2012; Hair, Ringle, and Sarstedt 2013; Nitzl 2016; Ringle, Sarstedt, and Straub 2012), which similarly poses a threat to result validity.

Primarily, the decision on the mode of measurement model should be based on theoretical considerations made ex ante (Diamantopoulos and Winklhofer 2001; Hair et al. 2019; Jarvis, MacKenzie, and Podsakoff 2003; Rossiter 2002). In practice, however, the boundary between formative and reflective measurement models are blurred as it depends on construct conceptualization. Therefore, Jarvis, MacKenzie, and Podsakoff (2003) created a guideline (see Figure 2.10) that summarizes the fragmentary advice from previous considerations (e.g., Bagozzi 1984; Bollen 1989; Fornell and Bookstein 1982; MacCallum and Browne 1993) and coincides with recommendations of Rossiter (2002) and Chin (1998).

[48] The criteria for evaluating reflective and formative measurement models will be introduced in a later section.

Criterion	Differentiation
1. Direction of causality	**Reflective**: *from latent variable to indicators* - indicators are manifestations of latent variable ("consequences") - change in indicator(s) does not cause change in latent variable - change in latent variable causes change in indicator(s) **Formative**: *from indicators to construct* - indicators are defining characteristics of latent variable ("causes") - change in indicator(s) causes change in latent variable - change in latent variable does not cause change in indicator(s)
2. Interchange-ability of indicators	**Reflective**: *indicators should be interchangeable* - indicators should have same/similar content - indicators should share a common theme - Dropping indicator should not alter latent variable's conceptual domain **Formative**: *indicators are not necessarily interchangeable* - indicators not need to have same/similar content - indicators not need to share a common theme - Dropping indicator may alter conceptual domain of latent variable
3. Covariation among indicators	**Reflective**: *indicators expected to covary* - change in one indicator should be associated with change in other indicators **Formative**: *indicators not necessarily covary* - change in one indicator needs not to be associated with change in other indicators
4. Nomological net of indicators	**Reflective**: *nomological net for indicators should not differ* - indicators required to have same antecedents and consequences **Formative**: *nomological net for indicators may differ* - indicators not required to have same antecedents and consequences

Figure 2.10: Guidelines for differentiating Reflective and Formative Measurement Models[49]

[49] Adapted from Jarvis, MacKenzie, and Podsakoff (2003, p. 203).

Furthermore, confirmatory tetrad analysis for PLS-SEM (CTA-PLS; Gudergan et al. 2008) provides ex-post empirical evidence of the adequateness of the chosen measurement model (i.e., reflective or formative). This analysis is based on the underlying common domain of indicators in a reflective measurement model, which should empirically be reflected by zero tetrads (τ). Tetrads display differences of a pair of complementary products of two indicators' covariances (thus, a minimum of four indicators are required for this analysis). Consider a measurement model with four indicators. Six unique covariations can be derived between the indicators (σ_{12}, σ_{13}, σ_{14}, σ_{23}, σ_{24}, and σ_{34}). These are combined to three complementary products ($\sigma_{12} \times \sigma_{34}$, $\sigma_{13} \times \sigma_{24}$, and $\sigma_{14} \times \sigma_{23}$), which subsequently form three[50] tetrads[51]:

$$\tau_{1234} = \sigma_{12} \times \sigma_{34} - \sigma_{13} \times \sigma_{24}$$

$$\tau_{1342} = \sigma_{13} \times \sigma_{24} - \sigma_{14} \times \sigma_{23}$$

$$\tau_{1423} = \sigma_{14} \times \sigma_{23} - \sigma_{12} \times \sigma_{34}$$

Next, it is tested whether one of the tetrads is significantly different from zero. If all tetrads are not significantly different from zero[52], the so-called "vanishing tetrads" support the assumption of a reflective measurement model. Otherwise, if one or more tetrads are significantly different from zero, the results indicate that the assumption of a reflective measurement model has to be rejected and a formative measurement should eventually be considered. Therefore, the CTA-PLS method can also be used in a confirmatory manner when a formative measurement model has been designed ex ante. Nonetheless, CTA-PLS results should not be followed mechanistically nor attach less importance to the theoretical reasoning (Hair, Hult et al. 2017), but "routinely

[50] Since the sign of the tetrads is irrelevant for the further analysis, redundant tetrads are not displayed (e.g., $\tau_{1324} = \sigma_{13} \times \sigma_{24} - \sigma_{12} \times \sigma_{34} = -\tau_{1234}$).

[51] Note that the number of unique covariations between indicators and the number of resulting tetrads potentially grows with the number of indicators.

[52] In a later section, the non-parametric bootstrapping procedure used in PLS-SEM to test for significance will be described.

employ this technique" (Hair et al. 2012, p. 423) to provide empirical support for the measurement model chosen ex ante.

To counter this threat in the SEMs of the dissertation, indicators in all measurement models are carefully selected and (eventually) adjusted reflecting the guidelines of Jarvis, MacKenzie, and Podsakoff (2003) (Figure 2.10). Furthermore, the recommendations of Hair et al. (2012) and Hair et al. (2019) for assessing outer models are thoroughly followed. Although the empirical tests and results clearly indicate that the measurement models do not suffer from measurement misspecification, I explicitly apply CTA-PLS in Study 3 (see chapter 5.4.3) to ensure that the intended reflective measurement model design of the newly created operationalizations of diagnostic use and interactive use is empirically confirmed.

2.1.3.4 Single-Item Measurement Models

The measurement of a latent variable[53] can also consist of only one indicator, referred to as a "single-item construct" (Churchill 1979). A single-item measurement model is visualized with a line between the single indicator and the associated latent variable (see Y_3 in Figure 2.1). It has been argued that a single-item measurement model is ideal in settings when "a construct's scope is narrow, unidimensional, and unambiguous for the respondents" (Hair et al. 2012, p. 423). In addition, single-item measures offer several practical advantages over multi-item measures in that, ceteris paribus, shortening the length of the survey. In particular, lengthy surveys likely "overload respondents, lead to a decrease in response rates, break-offs, and contain more missing values" (Fuchs and Diamantopoulos 2009, p. 196). Shorter surveys thus

[53] Similarly, observable constructs (e.g., professional experience, revenue, or number of correct answers in an experimental setting) can be measured with a single-item. In this case, the disadvantages of using single-items described in this paragraph are less likely to occur.

reduce the risk of sampling bias (Moore et al. 2002) and response bias (Drolet and Morrison 2001). However, several studies have shown that multi-item measures outperform single-item measures regarding their psychometric properties (e.g., Diamantopoulos et al. 2012; Gardner et al. 1998; Sarstedt and Wilczynski 2009), such that a single-item measurement does not offer "more for less" (Sarstedt and Wilczynski 2009). Some authors therefore even recommend avoiding single-item measurement models (Sarstedt, Diamantopoulos, and Salzberger 2016; Sarstedt, Diamantopoulos, Salzberger, and Baumgartner 2016). In PLS-SEM in particular, the use of single-item measurement models should be cautiously considered due to their "tendency to bias estimates (i.e., an overestimation of the measurement model relations and an underestimation of the structural model relations) when the number of indicators and/or the number of observations increase (i.e., consistency at large)" (Ringle, Sarstedt, and Straub 2012, p. 7; Hair et al. 2012; Lohmöller 1989; Wold 1982).

Hence, researchers face a trade-off when utilizing single-item measurement models. Diamantopoulos et al. (2012) provide guidance on this ambiguity and suggest considering a single-item measurement model in cases where there is a small sample ($n < 50$), expected low effect sizes (< 0.3), and when there are exceptionally homogenous (Cronbach's alpha > 0.90) and semantically redundant items in the multi-item scale from which it is derived. Otherwise, a multi-item measurement model should be preferred. The authors conclude that "opting for single-item measures in most empirical settings is a risky decision as the set of circumstances that would favor their use is unlikely to be frequently encountered in practice" (Diamantopoulos et al. 2012, p. 446).

By following these recommendations, in the two survey-based studies of this dissertation, all latent variables of the core research model are measured with multi-item scales.[54] Nevertheless, observable single-item variables are also collected, e.g., employee experience (in years) in Study 1 or an objective measure of feedback frequency in Study 3, in order to conduct a redundancy analysis of the reflective multi-item operationalization of feedback frequency employed in the SEM.

2.1.3.5 Hierarchical Component Measurement Models

So far, the measurement models have been considered as first-order (unidimensional) models, i.e., indicators are directly attached to the latent variable. In some instances, multidimensional hierarchical order component models (HCMs) represent valuable alternatives for measuring latent variables to reduce model complexity, to circumvent collinearity issues, or to adequately reflect theory suggesting distinct subdimensions (Hair, Sarstedt et al. 2017; Wetzels, Odekerken-Schröder, and Van Oppen 2009). A hierarchical component model consists of two layers of latent variables[55]: A higher-order component (HOC) which captures the abstract construct to be measured, and multiple lower-order components (LOCs) that represent the subdimensions of the HOC and which the indicators are directly assigned. Thereby, the number of indicators per LOC should not differ substantially, as this distorts the relationship between the LOCs and the HOC (Becker, Klein, and Wetzels 2012).

Similar to first-order measurement models, the relationship between the HOC and its LOCs can be reflective (paths point from HOC to LOCs) or formative (paths point from LOCs to HOC). Hence, from the relationships between HOC and its LOCs and the measurement model of the LOCs, a taxonomy of four HCM-types can be

[54] For research efficiency reasons, I measure one control variable (i.e., output measurability) with a single-item in Study 3.

[55] From a methodological point of view, any number of layers could be designed (Wetzels, Odekerken-Schröder, and Van Oppen 2009). In practice, however, two layers are most common (Hair, Sarstedt et al. 2017).

derived (Jarvis, MacKenzie, and Podsakoff 2003): a reflective-reflective HCM, a reflective-formative HCM, a formative-reflective HCM, and a formative-formative HCM. Exactly as in the decision on first-order measurement models[56], the relationship between HOC and its LOCs is determined; it is primarily driven by conceptual considerations and can be tested ex-post with CTA-PLS.

To draw on the example of feedback frequency, research may be interested in the implications for the recipient arising from the overall feedback frequency he or she experiences, which is composed of feedback frequency from multiple feedback sources. Instead of directly linking supervisor feedback frequency, customer feedback frequency, and co-worker feedback frequency directly with, for example, four endogenous latent variables that depict consequences of feedback frequency, a HOC is constituted ("overall feedback frequency") that is formatively measured with the aforementioned three feedback frequency sources (LOCs) as distinct subdimensions. In turn, these three latent variables are measured with indicators in a regular first-order measurement model (e.g., reflectively). In this example, designing a HCM reduces the number of path coefficients from 12 (= 3 exogenous variables x 4 endogenous variables) to 7 (= 3 LOCs + 4 endogenous variables) and offers insights into an overarching construct of feedback frequency. Although the HCM opens up new analysis possibilities, such as the identification of success factors for feedback frequency intensity (cf. section on formative measurement models), one has to be aware that implementing a HCM is always accompanied by a loss of information, since the previous direct effects between a LOC and the endogenous latent variable(s) are now diluted though the mediation of the HOC (Hair, Sarstedt et al. 2017).

---

[56] For more details, see the previous section on the distinction between reflective and formative measurement models.

Ringle, Sarstedt, and Straub (2012) and Wetzels, Odekerken-Schröder, and Van Oppen (2009) recommend the two-stage approach to measure HCMs, which is similar to the two-stage approach to measuring the interaction term in moderations (Henseler and Chin 2010) and builds upon the results of the repeated indicator approach (Lohmüller 1989; Wold 1982). The latter approach alone leads to difficulties when the HOC serves as an endogenous variable and the relationships between the HOC and its LOCs are formative. Then, there is a crowding-out of the HOC's explained variance (R^2) from the LOCs, such that the path relationships of the HOC's antecedents (i.e., latent variables that point at the HOC) will be close to zero and insignificant, potentially leading to a Type II error (Ringle, Sarstedt, and Straub 2012).[57]

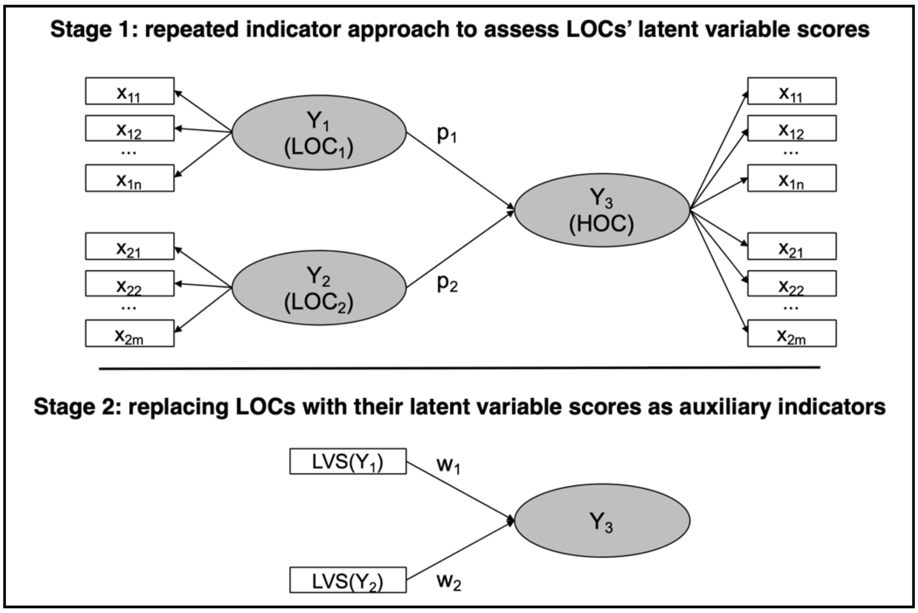


Figure 2.11: Visualization of the Two-stage Approach to assess a Reflective-Formative HCM[58]

[57] Alternatively, Becker, Klein, and Wetzels (2012) propose a hybrid approach that connects the antecedents not only to the HOC but also to all (formative) LOCs to account for the indirect effects of the antecedents on the LOCs. The total effect, i.e., the sum of the direct effect of the antecedent on the HOC and the indirect effects of the antecedent on the HOC via the LOCs, then represents the true effect of the antecedent on the HOC. Since the two-stage approach is applicable to all types of HCMs and simplifies measurement model evaluation, whereas the hybrid approach requires manual recalculations and substantially increases model complexity (which counteracts the benefits of HCMs), only the two-stage approach is further considered for HCM measurement.

[58] Adapted from Ringle, Sarstedt, and Straub (2012, p. 7).

In the first stage of the two-stage approach (see Figure 2.11), the latent variable scores (LVS) of the LOCs are determined using the repeated indicator approach: The indicators are assigned to the LOCs and repeatedly also to the HOC according to the measurement model of the LOCs (e.g., reflective in Figure 2.11). Thus, the HOC (for itself) is considered an auxiliary latent variable. While at this stage the measurement models of the LOCs are evaluated similarly to ordinary first-order measurement models, the measurement model of the HOC should not be analyzed (Hair, Sarstedt et al. 2017). In the second stage, the LOCs are replaced with their latent variable scores, which serve as auxiliary indicators for the HOC. Importantly, the path relationships between the HOC and its LOCs correspond to the relationships between the HOC and its auxiliary indicators: weights for formative first-order measurement model (as in Figure 2.11) or loadings for reflective first-order measurement model. Again, the measurement model (now of the HOC) is evaluated analogously to regular reflective or formative first-order measurement models.

Employing a HCM in the SEMs in this dissertation was considered, for instance, when developing the novel measurement for control use (Study 3) because I could not draw on an established measure that either treats diagnostic use and interactive use as subdimensions (LOCs) of an overall use construct (HOC) or does not (two first-order latent variables).[59] However, as the empirical analysis underscores theoretical considerations about the distinctiveness of diagnostic and interactive use (Tessier and Otley 2012) and the need for reducing model complexity is not compelling, I refrain from operationalizing control use in a HCM.

[59] Similarly, the indicators operationalizing both uses of control could have formed subdimensions, which was also not supported by the empirical results.

2.1.3.6 Measurement Models of the Interaction Term

The operationalization of the interaction term in a moderated relationship differs from other measurement models since no specific indicators are collected for this auxiliary latent variable. In contrast, the interaction term is assessed with a combination of indicators reused from the latent variables that co-determine the moderation (i.e., the moderator variable and the exogenous variable). Generally, there are three approaches for creating the measurement model of the auxiliary latent variables: the repeated indicator approach, the orthogonalization approach, and the two-stage approach (e.g., Henseler and Chin 2010; Henseler and Fassott 2010; Rigdon, Ringle, and Sarstedt 2010). A simulation study of Henseler and Chin (2010) showed that the two-stage approach should be preferred when the aim of the study is to test the significance of the interaction term. Furthermore, the orthogonalization approach and the repeated indicator approach are unable to account for formative measurements, such that the usage of the two-stage approach is generally recommended (Hair, Hult et al. 2017; Hair, Sarstedt et al. 2017).

In the first stage of the two-stage approach (see Figure 2.12), the latent variable scores (LVSs) are determined in a main effects model (i.e., the interaction term is excluded). In the second stage, the indicators of the latent variables are replaced by their latent variable scores. Besides, the interaction term is added to the model and measured with the product of the LVS of the exogeneous variable (Y_1 in Figure 2.12) and the LVS of the moderator (M in Figure 2.12). Whereas the measurement models of the latent variables in the main effects model (first stage) must satisfy regular measurement criteria, e.g., for reflective or for formative measurement models (see next section), no quality checks are necessary for the measurement of the interaction term (second stage) since it displays an auxiliary measure (Hair, Hult et al. 2017; Hair, Sarstedt et al. 2017).

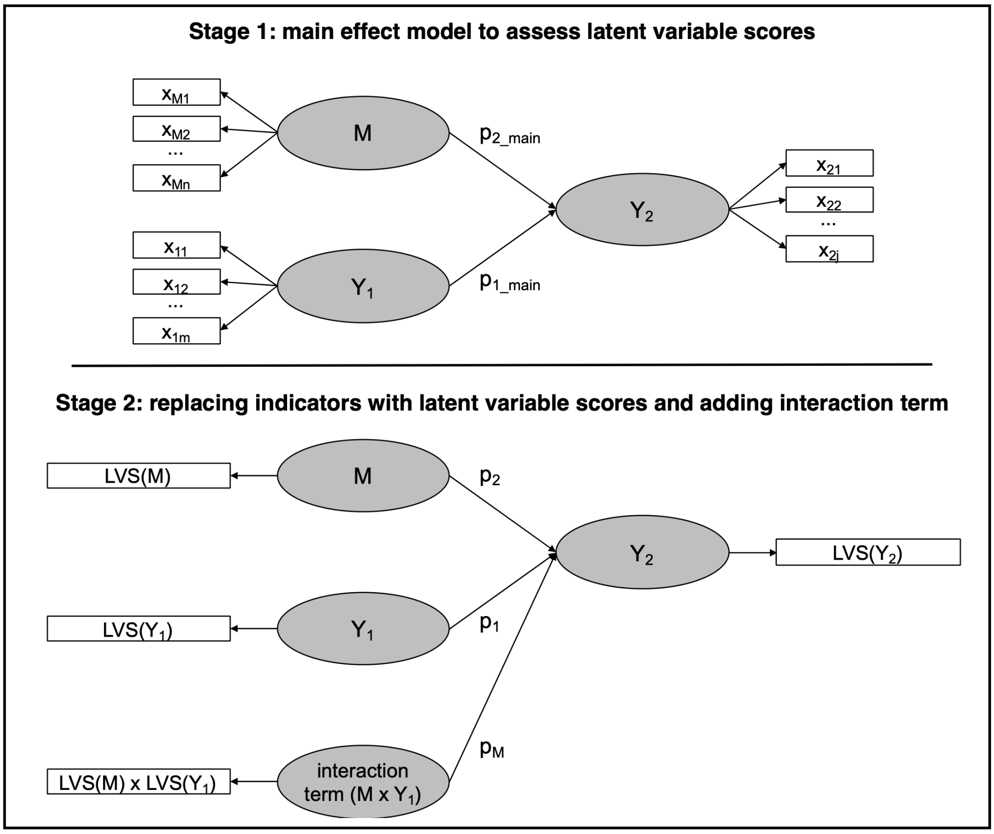


Figure 2.12: Visualization of the Two-stage Approach to assess the Measurement of an Interaction Term[60]

2.1.4 Analyzing the Structural Equation Model

In the previous steps, relationships between latent variables and relationships between latent variables and their respective indicators have been determined. In other words, the SEM has been set. Subsequently, the SEM is analyzed by means of empirical data and the hypothesized relationships are tested. More specifically, the analysis consists of the following steps: (1) application of the PLS-algorithm to calculate the latent variables scores, (2) evaluation of the outer model, and (3)

[60] Adapted from Rigdon, Ringle, and Sarstedt (2010, p. 263).

evaluation of the inner model. Afterwards, the results[61] are interpreted and conclusions are drawn. The structure of the following sections is based on the sequence of PLS-SEM analysis (1-3).

2.1.4.1 The PLS-Algorithm

The PLS-SEM algorithm (Wold 1982; Lohmöller 1989) converts the observable empirical data (i.e., the indicators used in the outer model) into latent variable scores. In turn, the latent variables scores are used to estimate the relationships between the latent variables (i.e., the path coefficients in the inner model) and the relationships between the latent variables and their respective indicators (i.e., the indicator loadings for reflective measurement models and the indicator weights for formative measurement models in the outer model). Generally, the algorithm strives to optimize the multiple partial regression results (i.e., to minimize the sum of the squared residual error terms) by iteratively approximating the ideal values for the path coefficients, the indicator loadings, and the indicator weights.

[61] Facultatively, additional analyses (such as multigroup analysis, FIMIX-PLS or PLS-POS) can be conducted. These have been described in previous sections since they conceptually address the relationships between latent variables in the inner model and are not particularly relevant to the evaluation of the main structural equation model.

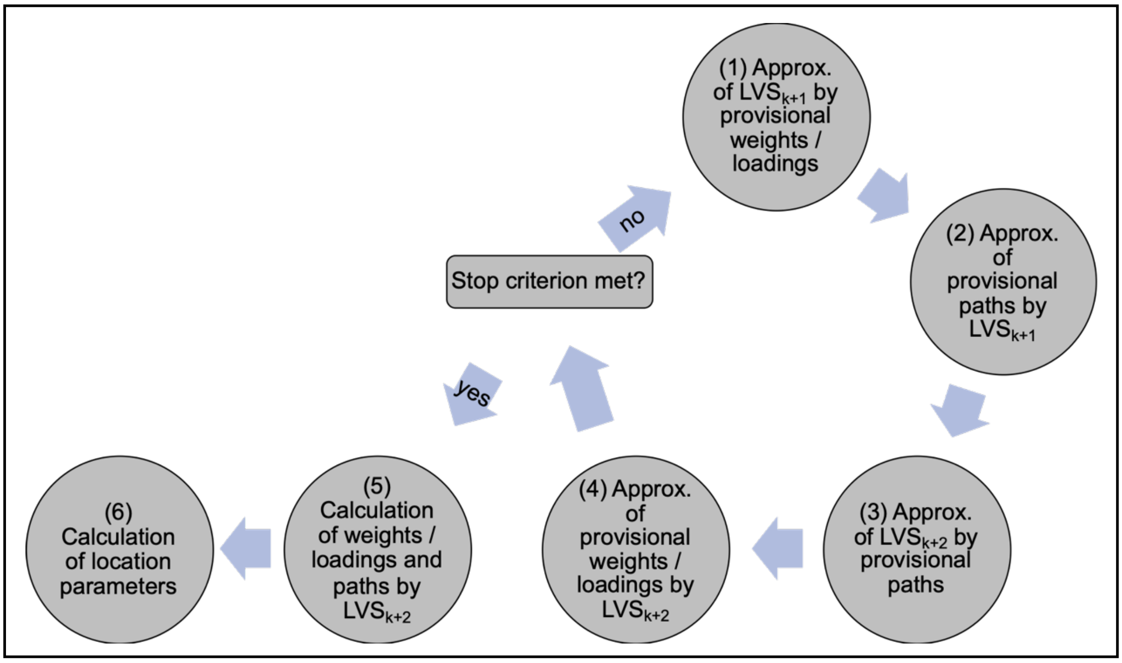


Figure 2.13: Visualization of the iterative PLS-SEM Algorithm

Drawing on Lohmöller (1989) and Henseler, Ringle, and Sinkovics (2009), the PLS-SEM algorithm comprises the following steps (see Figure 2.13): (1) Latent variable scores (LVS_{k+1}) are approximated by a linear combination of the latent variable's indicators and the provisional indicator weights[62] (and loadings[63]). In the first iteration, the initial weights (and loadings) should be set to 1.0 (Henseler 2010). Otherwise, the estimated (provisional) indicator weights (and loadings) from the previous iteration are used. (2) Provisional path coefficients (inner model) are approximated by partial regressions using the previously estimated latent variable scores (LVS_{k+1}).[64] (3) Novel latent variable scores (LVS_{k+2}) are approximated by a linear combination of the previously estimated (provisional) path coefficients and latent

[62] By means of the regression equation: $Y_i = w_1 \times x_1 + w_2 \times x_2 + \ldots + w_n \times x_n + e_i$, where n denotes the number of the latent variable's indicators (see previous section on formative measurement models).
[63] By means of the regression equations: $x_1 = l_1 \times Y_i + e_1$, $x_2 = l_2 \times Y_i + e_2$, ... $x_n = l_n \times Y_i + e_n$, where n denotes the number of the latent variable's indicators (see previous section on reflective measurement models).
[64] By means of the regression equation for each endogenous latent variable: $Y_i = p_1 \times Y_1 + p_2 \times Y_2 + \ldots + p_n \times Y_n + e_i$, where n denotes the number of the exogenous latent variable (Y_1, $Y_2 \ldots Y_n$) pointing directly at the endogenous latent variable (see previous section on direct relationships in the inner model).

variable scores (LVS_{k+1}). (4) Novel provisional indicator weights (and loadings) are approximated by a linear combination of the latent variable scores (LVS_{k+2}) and its indicators. If the difference between these provisional weights (and loadings) and the provisional weights (and loadings) used in the first step of this iteration is below a certain threshold (regularly, the sum of all changes should not exceed 10^{-5}, Wold 1982), the iterative algorithm terminates and subsequently (5) the 'final' indicator weights, indicator loadings, and path coefficients as well as the (6) location parameters (e.g., means or intercept of unstandardized latent variables) are calculated. If the stop criterion is not met, the iteration continues with its initial step, (1) by using the previously estimated provisional weights (and loadings) (from step 4) to approximate the latent variable scores (LVS_{k+3}). Thus, the iteration reflects an interplay between optimizations within the inner model (steps 2 and 3) and the outer model (steps 1 and 4).

Significant tests are not an inherent part of the PLS-algorithm. Nonetheless, significant tests are required repeatedly during inner and outer model evaluation and build on the multiple application of the PLS-algorithm. Therefore, this section describes the principles and procedure of significant tests in PLS-SEM.

Due to its leniency to distributional assumptions, parametric significant tests cannot be applied in PLS-SEM. Thus, PLS-SEM utilizes the non-parametric bootstrapping procedure to estimate the significance of its parameters (Davison and Hinkley 1997; Efron and Tibshirani 1986, 1994). In this procedure, multiple samples (referred to as bootstrapping samples) are drawn from the original sample. Each bootstrapping sample equals the original sample in its number of observations (the so-called bootstrapping cases). However, their composition differs since the observations are drawn from the original sample with replacement. Hence, it is to be expected that a bootstrapping sample will contain some observations more than once, while some observations in the original sample will not be included at all. In another bootstrapping sample, the composition may be just about reversed. As a minimum, the number of bootstrapping samples must exceed the number of observations, but 5,000 bootstrapping samples are generally recommended (Hair, Ringle, and Sarstedt 2011; Hair et al. 2012).[65] In a subsequent step, the previously designed PLS-SEM is calculated for each bootstrapping sample. Thus, 5,000 estimated values are available for each parameter, e.g., a specific path coefficient, the weight of a specific formative indicator, but also more complex parameters such as the heterotrait-monotrait ratio (HTMT).[66] These values form of bootstrap distribution that "can be viewed as reasonable approximation of an estimated coefficient's distribution in the population, and its standard deviation can be used as a proxy for the parameter's standard error in the population" (Hair, Hult et al. 2017, p. 151). In turn, this standard deviation can be used in a Student's t-test[67] to estimate whether the respective parameter is significantly different from zero. The resulting t-value indicates the significance level of the respective parameter. Usually, the thresholds[68] are 1.65, 1.96, and 2.57, corresponding to two-tailed significance levels of 1%, 5%, and 10%, respectively.

2.1.4.2 Criteria for the Evaluation of the Outer Model

Generally, the assessment of the outer model is based on the measurement model's reliability and validity (Hair, Hult et al. 2017). Measurement models with a high reliability and a high validity provide assurance that the measured value, e.g., of a

---

[65] I consistently follow this recommendation when the bootstrapping procedure is employed.
[66] Further details to the HTMT will be given in a later section.
[67] The Student's t-test is defined as follows: $t = \frac{x_i}{se_{xBTi}}$, where se_{xBTi} denotes the standard deviation in the bootstrap distribution of the respective variable x_i, whose value is derived from the PLS-SEM model results based on the original sample.
[68] For more than 30 observations, the t-distribution approximates the Gaussian normal distribution sufficiently well to allow Gaussian normal quantiles to be used for converting the t-values to significance levels (Hair, Hult et al. 2017).

latent variable score, comes close to the actual value. However, the measured value is also influenced by measurement errors, that should be minimized. The relationship between the measured value (x_M) and the actual value (x_T) can be displayed by the following equation:

$$x_M = x_T + e_s + e_r$$

While reliability of a measurement refers to the size of random error (e_r), validity is associated with the size of systematic error (e_s). A random measurement error can arise from situational influences, such as the mood of an individual respondent when participating in the study, and cannot be replicated. Hence, reliable measurement models (i.e., $e_r \approx 0$) lead to similar results when the measurement is repeated; they are accurate. Complementarily, systematic measurement errors still occur when the measurement is replicated. In other words, a measurement is valid if it systematically measures the "right" object (i.e., $e_s \approx 0$). Potential cause of systematic errors is poorly worded survey items. Only a reliable measurement can be considered to be valid, otherwise the systematic error cannot be distinguished from the random measurement error (Sarstedt and Mooi 2014). Conversely, a reliable measurement is not necessarily valid.

More specifically, *reflective measurement models* are evaluated based on their internal consistency reliability, their convergent validity, and their discriminant validity in PLS-SEM (Hair et al. 2019; Hair et al. 2012). First, reliability has to be established as a precondition for validity (Sarstedt and Mooi 2014). By definition, testing a reliable measurement involves assessing the degree of congruence of results when the measurement is repeated. In surveys, however, a repeated measurement is inherently difficult to perform. On the one side, given that "space on a questionnaire is limited and therefore very valuable” (Fuchs and Diamantopoulos 2009, p. 196), repeating entire measurement models would represent a relative waste of space just to establish measurement reliability. On the other side, repeated measurements at different points in time increase the risks of interim events[69], learning effects[70], or selection effects[71], and also have research economic disadvantages. Thus, in PLS-SEM reliability is usually established by the internal consistency (Hair, Hult et al. 2017). This rather implicit approach assesses the interrelation of one indicator with the other indicators of the latent variable’s measurement model. Importantly, internal consistency is only relevant for assessing reflective measurement models, since indicators in a formative measurement do not necessarily share a common theme or interrelate (cf. Figure 2.10) (Diamantopoulos 2006; Diamantopoulos and Winklhofer 2001). Traditionally, the Cronbach’s alpha[72] is used to establish internal consistency reliability. It measures the intercorrelations of the latent variable’s indicators on a scale ranging from zero to one, with higher values indicating higher reliability. Since the Cronbach’s alpha assumes that all indicators are equally reliable (i.e., all indicators have equally sized loadings) and is sensitive to the number of indicators, it has long been criticized as a criterion for internal consistency reliability in PLS-SEM (e.g., Hair et al. 2012). In contrast,

---

[69] Interim events might change respondent’s answers in the questionnaire. In order to limit the adverse effect of interim events on the assessment of reliability, a short time gap between the initial test and the subsequent retest is preferred (Döring and Bortz 2016).

[70] Learning effects largely refer to tests that require participants to perform experimental tasks, such as simple mathematical calculations. If these tasks are performed repeatedly, participants will have better skills than they had when they initially started with the first tasks (Döring and Bortz 2016). However, learning effects may also be present in survey studies, as respondents may ruminate on the intentions of the surveys and adapt their responses accordingly, referred to as implicit theory bias (Podsakoff et al. 2003).

[71] In a retest approach, respondents must be motivated to participate in two surveys. Besides the additional effort in data collection to compensate for dropouts in general, certain respondents may be particularly susceptible to dropping out, which also leads to selection biases.

[72] The Cronbach’s alpha is defined as follows: $\alpha = \left(\frac{n}{n-1}\right) * \left(1 - \frac{\sum_{i=1}^{n} s_i^2}{s_t^2}\right)$, where s_i^2 denotes the variance of the respective indicator i, n the number of indicators of the latent variable, and s_t^2 the variance of the sum of all n indicators.

composite reliability[73] (Jöreskog 1971), which compares loadings with the measurement error of their respective indicators, accounts for different loading sizes and is insensitive to the number of indicators. Similarly, higher values on its scale from zero to one imply higher reliability. More recently, however, it has been recommended to consider both Cronbach's alpha and composite reliability to assess internal consistency reliability (e.g., Hair et al. 2019). More specifically, both indices should exceed values of 0.7 (Bagozzi and Yi 1988; Nunnally and Bernstein 1994), with the Cronbach's alpha being the more conservative measure and the composite reliability tending to overestimate internal consistency reliability (Hair et al. 2019). However, values above 0.9 are not desirable since they suggest indicator redundancy, which increases error term correlations and thus reduces reliability (Drolet and Morrison 2001), or the presence of straight-lining response patterns (Hair et al. 2019).

Second, convergent validity expects the measure to correlate with other measures of the same construct (Hair, Hult et al. 2017). Due to their common conceptual domain (cf. Figure 2.10), its reflective indicators display alternative measures of the latent variable. Thus, the indicators are expected to correlate with the latent variable score. In particular, the indicator loading[74] reflects the extent to which it correlates with the latent variable. In order to attain convergent validity, the indicator loadings should be larger than 0.7 (Hulland 1999).[75] In turn, the communality of an

---

[73] The composite reliability is defined as follows: $p_c = \frac{(\sum_{i=1}^{n} l_i)^2}{(\sum_{i=1}^{n} l_i)^2 + \sum_{i=1}^{n} var(e_i)}$, where l_i denotes the loading of the respective indicator i, n the number of indicators of the latent variable, and $var(e_i)$the variance of the measurement error of the respective indicator i ($var(e_i) = 1 - l_i^2$).

[74] Since the loadings (l_i) represent regression coefficients in a univariate regression: $x_1 = l_1 \times Y_i + e_1$, $x_2 = l_2 \times Y_i + e_2$, ... $x_n = l_n \times Y_i + e_n$, where x_i denotes the indicator, l_i its respective loading, e_i its respective error term, and Y_i the latent variable. The size of the loading is often referred to as indicator reliability (Hair, Hult et al. 2017).

[75] The underlying assumption is that the indicators point in the same direction. Sometimes reverse coded items are included in a measurement scale. However, these items are usually reversed before being included in PLS-SEM. Otherwise, indicator loadings of size -0.7 and larger absolute values would also be acceptable if their reversed polarity is conceptually reasonable. Generally, the range of the standardized loadings is similar to a correlation (i.e., from -1 to 1).

item – the proportion of variance that is explained by the latent variable (recall that the arrows point from latent variable to the indicators) – will be larger than 50% (0.7 x 0.7 ≈ 0.5).[76] A communality of this size indicates that the shared variance of latent variable and its indicator is larger than the measurement error's variance. Another approach to assess the latent variable's convergent validity is its average variance extracted (AVE).[77] The AVE is defined as the mean of the squared loadings of all latent variable's indicators and is equivalent to the construct's communality (Hair, Hult et al. 2017). Similarly, an AVE above 0.5 is desired (Bagozzi and Yi 1988), which means that the latent variable explains on average more than 50% of its indicators' variance. According to Bagozzi, Yi, and Phillips (1991) and Hair, Ringle, and Sarstedt (2011), to improve the latent variable's convergent validity, indicators with low loadings (i.e., below 0.4) should be removed from the measurement model, while indicators with loadings slightly below the threshold of 0.7 (i.e., between 0.4 and 0.7) should be removed from the measurement model only if doing so improves the latent variable's internal consistency reliability or average variance extracted.

Third, discriminant validity describes the extent to which measures are statistically different from measures of other latent variables (Hair, Hult et al. 2017). If discriminant validity is established, the latent variable is unique and its domain is not captured by other latent variables in the model. Cross-loadings, i.e., correlations of an indicator with latent variables to which it is *not* attached, are used to test discriminant validity at the indicator level (Chin 1998). More specifically, the cross-loadings of an indicator should not exceed its loading (i.e., the correlation to the latent variable to which it is attached to). Otherwise, this would pose an issue of discriminant validity.

---

[76] Some researchers therefore recommend a threshold value of 0.708 for the loadings, since 0.708^2 just exceeds 0.5 (Hair et al. 2019).

[77] The average variance is defined as follows: $AVE = \frac{\sum_{i=1}^{n} l_i^2}{n}$, where l_i denotes the loading of the respective indicator i and n the number of indicators of the latent variable.

The Fornell-Larcker criterion (Fornell and Larcker 1981) has long been used to test discriminant validity at the latent variable level. According to this criterion, the square root of the latent variable’s AVE should be larger than the latent variable’s correlation with any other latent variable in the model. Frequently, the check for the Fornell-Larcker criterion is presented with a correlation matrix of the latent variables, in which the diagonal values[78] are the square root of the latent variables’ AVEs (e.g., Janka and Guenther 2018). However, since the AVE is only meaningful for latent variables with reflective measurement models, diagonal values of latent variables with formative measurement model or single-items should be crossed out.

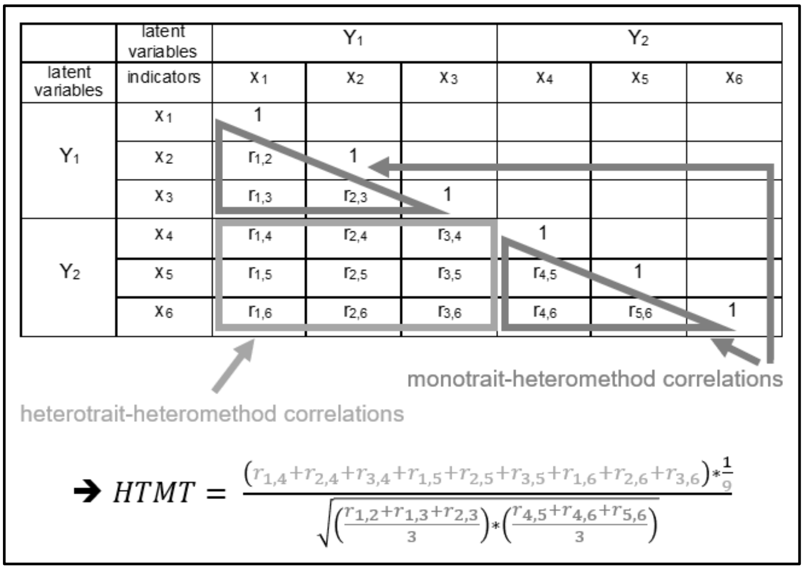


Figure 2.14: Illustration of the HTMT by a Correlation Matrix of the Indicators[79]

More recently, Henseler, Ringle, and Sarstedt (2015) and Rönkkö and Evermann (2013) detected flaws in the previous criteria for establishing discriminant validity. In particular, they fail to adequately evaluate discriminant validity if two latent variables are perfectly correlated (cross-loadings) or if the loadings of the latent

[78] By definition, diagonal values in a correlation matrix indicate the correlation of a latent variable with itself (i.e., all are equal to 1) and therefore convey no additional information.
[79] Adapted from Henseler, Ringle, and Sarstedt (2015, p. 122).

variable vary little (Fornell-Larcker criterion). To overcome these shortcomings, Henseler, Ringle, and Sarstedt (2015) introduced the heterotrait-monotrait ratio (HTMT), which was also found to generally perform better in evaluating discriminant validity (Franke and Sarstedt 2019; Voorhees et al. 2016). The HTMT is used to assess whether two latent variables are statistically unique. In this approach, the average (arithmetic mean) correlation of each combination of indicators from two different latent variables (heterotrait-heteromethod correlations) is compared to the geometric mean of the average correlations between the indicators of each latent variable (monotrait-heteromethod correlations). The resulting ratio (the HTMT), also referred to as disattenuated correlation, estimates the true correlation between these two constructs. Figure 2.14 visualizes the HTMT approach with a correlation matrix of the indicators (two latent variables with three indicators each) and states the formula that would apply to this trivial example. An HTMT close to 1.0 indicates a lack of discriminant validity[80], whereas a value below 0.85 (Henseler, Ringle, and Sarstedt 2015; Voorhees et al. 2016) suggests that discriminant validity is established.[81] Figure 2.15 summarizes the previous elaboration on statistical criteria for reflective measurement models that will be referred to in both survey-based studies in this dissertation.

[80] Consistent with the previous discussion on the direction of correlations and the sign of measurement evaluation criteria, Henseler, Ringle, and Sarstedt (2015) point out that “strictly speaking, one should assess the absolute value of the HTMT, because a correlation of -1 implies a lack of discriminant validity, too” (p. 122).

[81] The HTMT can also be used for a statistical test of discriminant validity with a bootstrapping procedure (see previous section) (Henseler, Ringle, and Sarstedt 2015). Then, not the true indicator correlations are used to determine the HTMT (see Figure 2.14), but multiple bootstrapping samples form a confidence interval for the HTMT. If the confidence interval for the HTMT does not contain 1.0, discriminant validity is established.

Criterion	Indicator	Threshold Value
Internal Consistency	Cronbach's alpha	> 0.70
	Composite reliability	> 0.70
Convergent Validity	Indicator loadings	> 0.70
	AVE	> 0.50
Discriminant Validity	HTMT	< 0.85

Figure 2.15: Overview of Criteria for Reflective Measurement Model Assessment with respective Threshold Values

The evaluation of *formative measurement model* requires different criteria because of the different relationship between the latent variable and its indicators (cf. Figure 2.10). Even before data collection, establishing content validity, i.e., "ensuring that the formative indicators capture all (or at least major) facets on the construct" (Hair, Hult et al. 2017, p. 139), through extensive literature reviews, expert interviews, and theoretical grounding, ensures that no important aspect of the construct are omitted.[82] When it comes to empirically test formative measurement models, convergent validity, lack collinearity, and possess relevant and significant indicators are required. First, convergent validity, in which the measurement is expected to correlate with other measurements of the same construct (Hair, Hult et al. 2017), can be assessed by performing a redundancy analysis (Chin 1998).[83] In this approach, the latent variable with the formative measurement model under test serves as the single exogeneous

[82] While the congruence between the conceptual construct and its measurement is also relevant to the validity of reflective measurement models, their approach of indicator interchangeability makes them less susceptible to measurement error issues when an indicator is removed/left out or when not all facets of the construct are represented by indicators.
[83] The redundancy analysis is similar to a multiple indicators and multiple causes (MIMIC) model (Jöreskog and Goldberger 1984; Diamantopoulos and Winklhofer 2001), which cannot be performed in PLS-SEM since it requires indicators to be cause and effect simultaneously (Fornell and Bookstein 1982). However, this principle of MIMIC models is reflected at the latent variable level: The exogeneous latent variable causes / affects the endogenous latent variable.

variable for an endogenous latent variable that is conceptually equal but is operationalized with a different measurement. After applying the PLS-SEM algorithm, the estimated path coefficient[84] should be 0.7 or higher, reflecting that more than (approximately) 50% of the latent variables' variance overlaps. Otherwise, this would indicate convergent validity issues. Note that unlike the assessment of convergent validity for reflective measurement models, the establishment of convergent validity for formative measurement models thus needs to be considered prior to data collection, as additional indicators need to be gathered. Despite Chin (1998) proposed reflective measurement models for the endogenous latent variable in the redundancy analysis, more recent research suggest the use of a single-item measurement model as an alternative (Cheah et al. 2018; Sarstedt, Wilczynski, and Melewar 2013). The general reservations[85] against single-item measurement models do not bear fruit as "their role in redundancy analyses is different because single-items only serve as proxy for the constructs under consideration [and] the aim is no to fully capture the content domain of the construct but only to consider its salient elements" (Hair, Hult et al. 2017, p. 141). Instead, they offer a practical solution (Fuchs and Diamantopoulos 2009) that limits the drawbacks of testing convergent validity of formative measurement models with redundancy analysis.

Second, multicollinearity[86] among formative indicators has to be avoided. Since formative indicators capture different facet of the underlying construct (cf. Figure 2.10), they should not highly correlate, implying that their facets are not unique. In other words, formative indicators then steal each other's thunder when causing their latent

[84] Since the redundancy analysis requires that the inner model contains only these two latent variables, in other words, the endogenous latent variable has no other antecedents, the path coefficient can be interpreted as correlation between the two conceptually identical latent variables.
[85] See previous section on single-item measurement models.
[86] Actually, multicollinearity refers to collinearity of three or more variables. Thus, the term collinearity, which describes a high correlation between two variables, is further used to correctly address the simplest case of (multi-)collinearity.

variable. Besides conceptual considerations, collinearity among indicators leads to biased estimations of indicator weights and their significances since their effect on the latent variable cannot be isolated (Hair, Hult et al. 2017). The tolerance (TOL) or its inverse, the variance inflation factor (VIF), are used to assess the degree of collinearity for a formative indicator (Hair, Ringle, and Sarstedt 2011). The tolerance indicates the share of the indicator's variance that is *not* explained by the remaining indicators in the formative measurement model. To calculate the TOL, the formative indicator is approximated by a linear regression, with all other formative indicators in the latent variable's measurement model being independent variables. Next, the explained variance of the regressed indicator is subtracted from 1. The resulting difference has a range from 0 to 1 and can be interpreted as the proportion of the indicator's uniqueness that is not reflected by the remaining indicators of the formative measurement model. In PLS-SEM, the TOL of each formative indicator should exceed 0.2 (Hair, Ringle, and Sarstedt 2011). Since collinearity issues may be present at even higher values, the TOL ideally is higher than 0.33 (Becker et al. 2015; Hair et al. 2019). Similarly, the VIF[87] should be lower than 5 or 3, respectively. The VIF – entailing the same information as the TOL – can be interpreted as the factor by which the standard error of the indicator increases in a hypothetical situation in which the indicator would have zero correlation with any other indicator. Dealing with collinearity issues among formative indicators is a crucial step before estimating the significance and relevance of the formative indicators. Eliminating problematic indicators requires careful conceptual and theoretical considerations of whether the remaining formative indicators still sufficiently capture the content of the construct. Alternatively, the

[87] The variance influence factor is defined as follows: $VIF_{xi} = \frac{1}{1-R_{xi}^2}$, where R_{xi}^2 denotes the explained variance of the respective formative indicator x_i by the remaining formative indicators of the latent variable's measurement model.

indicators can be combined into a new composite indicator, or designing a formative-formative hierarchical component measurement model could solve collinearity issues – if, again, this is reflected by the theory (Hair, Hult et al. 2017).

Lastly, the formative indicators' significance and relevance should be assessed (Hair et al. 2019; Hair et al. 2012). Both are evaluated by the weights of the indicators. Since the weights are standardized, the relative contribution of the respective indicator to the latent variable can be read directly from them. Nonetheless, the question that needs to be answered is whether the relative contribution is truly different from zero. Consequently, a bootstrapping procedure is applied to test the significance of the indicator weights. However, the number of indicators in the formative measurement model has "important implications for the statistical significance and the magnitude of each indicators' weight" (Cenfetelli and Bassiellier 2009, p. 694) as the explained variance of the latent variable is divided up among the number of indicators, such that if more indicators are attached to the latent variable, the average share each indicator can explain reduces, potentially leading to biased significance values. More specifically[88], the maximum average weight of two uncorrelated formative indicators is 0.707, whereas for ten indicators this number shrinks to 0.316. Thus, Cenfetelli and Bassiellier (2009) recommend to further consider the absolute contribution of the formative indicator if its weight (i.e., its relative contribution) is non-significant. The absolute contribution is the weight of the indicator in a simple regression and as such

---

[88] In general, the maximum possible average weight is calculated by $w_{\emptyset max} = \frac{1}{\sqrt{n}}$, where n denotes the number of formative indicators. More specifically, for two uncorrelated indicators the explained variance would be split half, i.e., $R^2 = 100\% = 2 \times 0.5 \approx 2 \times (0.707)^2$, while for ten indicators, the explained variance would be split into ten parts, i.e., $R^2 = 100\% = 10 \times 0.1 \approx 10 \times (0.316)^2$. Note that these calculations represent ideal settings in which there is zero correlation between the indicators and zero measurement error (i.e., the indicators fully explain the latent variable). In practice, the significance of indicator weights is more severely affected by the number of indicators, since their relative contribution is expected to vary (hence lower weights are present which are more likely to be indicated as non-significant) and the latent variable is not fully explained by its indicators (hence less than 100% is to be "distributed" among the weights).

is statistically equivalent to the correlation between the indicator and the latent variable, and to the indicator's loading on the construct. Hair et al. (2019) specify this threshold: For a formative indicator with non-significant weight, its loading should exceed 0.5, indicating absolute relevance. Otherwise, indicator removal should be strongly considered based on the conceptual value of the indicator and its potential overlap with other indicators in the measurement model. Cenfetelli and Bassiellier (2009) even recommend to eliminate the indicator if also its loading is insignificant, calling into question its theoretical relevance. Still, an elimination of such an indicator has "almost no effect on the parameter estimates [; in general,] formative indicators should never be discarded simply on the basis of statistical outcomes" (Hair, Hult et al. 2017, p. 148 f.). Figure 2.16 summarizes the statistical evaluators for formative measurement models that will be utilized for assessing the formatively measured latent variable in Study 1.

Criterion	**Indicator**	**Threshold Value**
Convergent Validity	Path coefficient in redundancy analysis	> 0.70
Collinearity	VIF / TOL	< 3.00 / > 0.33
Significance and relevance of indicators	Indicator weight	significant (or indicator loading > 0.50)

Figure 2.16: Overview of Criteria for Formative Measurement Model Assessment with respective Threshold Values

As previously mentioned, the measurement models of *auxiliary measurement models* (such as HOCs in a HCM or interaction terms) should generally not be assessed. Similarly, the criteria for measurement model evaluation in PLS-SEM do not apply to *single-item measurement models* since “the single indicator and the latent variable have identical values” (Hair, Hult et al. 2017, p. 109). However, criterion validity, i.e., testing the correlation of the single-item with a multi-item scale of the same construct, comparable to redundancy analysis (Chin 1998), could be used to assess its validity (Diamantopoulos et al. 2012).

Consistent with the PLS-algorithm, the evaluation of the outer model and the inner model are related in that adequate measurement models are a prerequisite for the evaluation of path coefficients. If adjustments are to be made in the outer model to meet relevant criteria, for instance, by eliminating an indicator, the PLS-algorithm must be reapplied for parameter estimation due to the interplay between the outer and the inner model.[89]

2.1.4.3 Criteria for the Evaluation of the Inner Model

The assessment of the inner model involves examining the predictive power of the model and the relationships between the latent variables. In the following, several criteria for evaluating the inner model are presented. First, collinearity among latent variables has to be assessed (Hair, Hult et al. 2017; Hair et al. 2019). The relationships between latent variables are determined by multiple regressions in which the endogenous variable is approximated by the exogenous latent variables that are directly linked with it. Thus, similar to the structure of formative measurement models, collinearity among latent variables may bias the size of the path coefficients and their estimated significances. Therefore, the same criterion applies for assessing collinearity

[89] Consequently, the measurement models of the latent variables are not independent from each other either.

among latent variables[90], such that any pair of predictors for a selected endogenous latent variable should have a tolerance value (TOL) above 0.33, respectively a variance inflation factor (VIF) below 3. Analogously, "eliminating constructs, merging predictors into a single construct, or creating higher-order constructs" (Hair, Hult et al. 2017, p. 194) display valid options to overcome collinearity issues in the inner model.

Second, the significance and relevance of the path coefficients representing the hypothesized relationships are evaluated (Hair, Hult et al. 2017; Hair et al. 2019). Since the calculation of path coefficients is based on multiple regressions, a comparison with formative measurement models can also be made here. The path coefficients, similar to indicator weights, indicate the strength (absolute value) and direction (sign) of the relationship between the latent variables. A bootstrapping procedure is employed to estimate the significance of the path coefficient, indicating whether it is truly different from zero.[91] Due to the standardization of the path coefficients, an increase in the exogenous latent variable by one[92] standard deviation leads, ceteris paribus, to an increase in the endogenous latent variable by the product of its standard deviation and the path coefficient (Hair et al. 2010). Furthermore, the standardization also allows for a simple comparison of the relative importance of multiple predictors (exogenous latent variables): If one path coefficient is larger than another, the importance (contribution) of the associated latent variable for the endogenous variable is also larger than the other latent variable.

Third, the coefficient of determination (R^2) is used to evaluate the model's predictive power (Hair et al. 2019; Hair et al. 2012). More precisely, it represents the

[90] Please refer to the previous section regarding formative measurement model evaluation for more details on collinearity issues in multiple regressions, as well as about the TOL or the VIF.
[91] Please refer to a previous section for more details to the bootstrapping procedure and its evaluation.
[92] This trivial numerical example is given for illustrative purposes. However, any change in the exogenous latent variable (in terms of its standard deviation) can be directly referred to a change in the endogenous latent variable (in terms of its standard deviation) with the same proportion of the path coefficient.

in-sample predictive power (Rigdon 2012, Sarstedt et al. 2014). The R^2 expresses the share of the endogenous latent variable's variance that can be explained by its predictors.[93] Thus, it ranges from 0 to 1, with higher values indicating higher predictive power. General thresholds for R^2 are hard to define since the goodness of predictive power depends not only on the research discipline (Hair et al. 2019), model complexity (Hair, Hult et al. 2017), but also on the specifically observed constructs (Homburg and Baumgartner 1995). For instance, one would expect substantially higher predictive power of the level of CEO feedback frequency for a particular employee's satisfaction than for the resulting impact on the firm's overall performance (e.g., as measured by return on assets), so that different levels of explained variance (R^2) would be considered sufficient. Nonetheless, both relationships could be significant and therefore offer important contributions to the existing literature. Even more, Hair et al. (2019) caution against excessively high R^2-values (i.e., > 0.9), as "they are typical indicative of overfit" (p. 15).

Because the R^2 is sensitive by mathematical definition to the number of predictors (even if they have nonsignificant effects on the latent variable), the adjusted coefficient of determination[94] (R^2_{adj}) takes into account the number of exogenous latent variables and can otherwise be interpreted analogously to the R^2. Therefore, it can be used to compare models with different number of exogenous latent variables.

Fourth, while the R^2 and the R^2_{adj} refer to the outcome of the explained variance of the endogenous construct, the exogenous latent variable's effect size (f^2) defines its

---

[93] The coefficient of determination is defined as follows: $R^2 = \sum_{i=1}^{n}(p_i^{\,2})$, where p_i denotes the path coefficient of the endogenous latent variable's predictor i and n the number of exogenous variables that are associated with the endogenous latent variable. Note that the contribution of a path coefficient to the R^2 is independent from its sign.

[94] The adjusted coefficient of determination is defined as follows: $R^2_{adj} = 1 - \left(1 - R^2\right) * \frac{n-1}{n-k-1}$, where R^2 denotes the (unadjusted) coefficient of determination, n the sample size and k the number of exogenous latent variables.

genuine contribution to the explanation of the exogenous latent variable.[95] Technically, two PLS path models are estimated, one in which the exogenous latent variable is included in the model and one in which it is omitted from the model. The relative change in R^2 of the endogenous latent variable then refers to the effect size of the exogenous latent variable. Thus, the f^2 always refers to a specific relationship between two latent variables and ranges from 0 to 1. Cohen (1988) defines 0.02, 0.15, and 0.35 as thresholds for the effect size, which translates into small, moderate, and large effects, respectively. Again, note that also the f^2 is sensitive to the number of predictors (since it is based on a difference of two R^2 values), so more leniency must be given to these thresholds when more predictors are present.

Fifth, the Stone-Geisser's Q^2 value (Geisser 1974; Stone 1974) can be used for the inner model evaluation if the endogenous variable is measured reflectively or with a single-item measurement model (Hair et al. 2019; Hair et al. 2012). In contrast to the coefficient of determination, which refers to the accuracy of the prediction, the Q^2 value indicates the predictive relevance of the model and combines aspects of out-of-sample prediction and in-sample prediction (Hair et al. 2019; Sarstedt et al. 2018; Shmueli et al. 2016). In a blindfolding procedure (Chin 1998; Henseler, Ringle, and Sinkovics 2009; Tenenhaus et al. 2005), in which single data points in the endogenous latent variable's indicators are eliminated and replaced with mean values (i.e., they are treated as missing values), the PLS-algorithm is employed to estimate the replaced values. The difference between the original and the estimated values are then incorporated for Q^2 value calculation. Positive Q^2 values indicate that there is little

---

[95] The f^2 effect size of an exogenous latent variable is defined as follows: $f^2 = \frac{R_{incl}^2 - R_{excl}^2}{1 - R_{incl}^2}$, where R_{incl}^2 denotes the (unadjusted) coefficient of determination of a specific endogenous variable when the exogenous latent variable is included in the model, and R_{excl}^2 denotes the (unadjusted) coefficient of determination of the same endogenous variable when the exogenous latent variable is omitted from the model.

difference between the original and estimated values; hence, the model has predictive relevance for the endogenous latent variable, while complementarily negative values indicate a lack of predictive relevance.

Similar to the f^2 effect size, a q^2 effect size can also be calculated, indicating the change in Q^2 if the exogenous latent variable is omitted from the model.[96] Analogously, the values 0.02, 0.15, and 0.35 correspond to small, medium, and large effect sizes, respectively (Hair et al. 2012).

Finally, in the special case of an auxiliary latent variables for the interaction term,[97] only a selection of the previously introduced criteria must be used for their evaluation in the inner model: significance and relevance of path coefficients and effect sizes. More specifically, the significance of their path coefficient indicates whether a moderating effect is present, in other words, whether it is different from zero. Likewise, the path coefficient's sign and size indicate the direction and strength of this effect. The only difference to the assessment of ordinary latent variables (except for the number of criteria) refers to the thresholds of their effect sizes (f^2), which can be calculated analogously. Based on reviews of effect sizes of moderations, such as Aguinis et al. (2005), Kenny (2013) suggests more lenient thresholds for small, moderate, and large effect sizes of interaction terms (namely, 0.005, 0.01, and 0.025, respectively); because in research practice, Cohen's (1988) suggested thresholds for effect sizes are too optimistic to indicate the relevance of these auxiliary latent variables. This

[96] The q^2 effect size of an exogenous latent variable is defined as follows: $q^2 = \frac{Q_{incl}^2 - Q_{excl}^2}{1 - Q_{incl}^2}$, where Q_{incl}^2 denotes the Stone-Geisser's Q^2 value of a specific endogenous variable when the exogenous latent variable is included in the model, and Q_{excl}^2 denotes the Stone-Geisser's Q^2 value of the same endogenous variable when the exogenous latent variable is omitted from the model.

[97] HOCs are evaluated similarly to regular latent variables in that after applying the two-stage procedure, the LVSs of the LOCs are treated as if they were regular indicators of a first-order measurement model, so there are no differences from regular latent variables (except for the origin of the indicators).

specification is already reflected in current publication practices (e.g., Kuegler, Smolnik, and Kane 2015).

In contrast to CB-SEM, there are no established criteria in PLS-SEM to assess the overall model, such as a goodness-of-fit index (GoF). Despite attempts to create adequate model fit measures (cf. Dijkstra and Henseler 2015; Henseler et al. 2014; Tenenhaus, Amato, and Esposito Vinzi 2004; Tenenhaus et al. 2005), there are empirical and conceptual concerns (cf. Hair, Hult et al. 2017; Henseler and Sarstedt 2013; Rigdon 2012, 2014b; Shmueli 2010) as to whether model validation for exploratory modeling (as focused by CB-SEM) is applicable to PLS-SEM, whose approach is to maximize prediction. Therefore, the current recommendation (e.g., Hair et al. 2019; Hair et al. 2012) is not to test for an overall model fit.

2.2 Experimental Research

2.2.1 The Experiment as an Empirical Methodology in Social Science

An experiment is “a scientific investigation in which independent variables are manipulated and their effects on other dependent variables are observed” (Sprinkle 2003, p. 289). This empirical methodology was employed in Study 2. Its underlying principle in is that any systematic change in the dependent variable can be attributed to the (intentional) change in the independent variable (referred to as “condition” or “treatment”). Thus, experimental studies allow to draw strong causal inferences regarding the relationships among variables hence overcoming common limitations of survey-based research (Campbell and Stanley 1963; Cook, Campbell, and Day 1979; Kerlinger and Lee 2000; Sprinkle 2003): Cross-sectional studies are unable to show direct causation since there is no clear sequence of cause and effect. In cross-sectional studies, the cause-effect relationship between variables is largely determined by theory, but may still remain ambiguous. For instance, Judge et al. (2001) consider the relationship between employee satisfaction and job performance to be the "Holy Grail" of industrial-organizational psychology as several theories take different positions toward causality. While longitudinal studies can demonstrate that an effect follows a cause, these studies lack an appropriate counterfactual, i.e., researchers have to mitigate concerns such as omitted correlated variables and selection bias (Sprinkle 2003).

A valid experimental design provides reasonable assurance that no other variable (referred to as a "confounding variable") affects the relationship between the independent and dependent variables (Shadish, Cook, and Campbell 2002). Manipulating the independent variable as the only variation in the experimental design and randomization "allow the investigator to control the research setting and isolate the effects of variables that are confounded" (Sprinkle 2003, p. 289). The variables "to be controlled" depend on the research setting. For instance, participants suffering from color vision deficiency should be evenly distributed across the treatments when the experimental task requires them to identify colors, such as in the Stroop Color-Word Test (1935), whereas this even distribution is less problematic if the goal of the study is to examine the effects of feedback frequency on creativity.[98] In this dissertation, I therefore complement the two survey-based studies with an experiment (Study 2) (also) to control for the influence of other dimensions of feedback (e.g., feedback timing) and to provide assurance about the cause-effect relationship between feedback frequency and employee creativity.

[98] If participants are randomly assigned to experimental treatments, even seemingly less relevant attributes of participants should be evenly distributed across treatments.

It has to be noted that the *direct* influence of a controlled confounding variable on the dependent variable does not need to be considered (Döring and Bortz 2016). With respect to the previous example, incentives for participants (feedback recipients) to generate creative ideas has been shown to influence creative performance (e.g., Kachelmeier, Wang, and Williamson 2019), such that quantity-based compensation leads participants to generate more ideas than with fixed compensation. However, as long as incentives are identical across the treatment groups, i.e., the incentives are controlled for, the direct effect of incentive system scheme is similar for any condition, so it does not affect relative differences across treatments due only to the manipulation of the independent variable (level of feedback frequency). If interaction effects between incentive system design and feedback frequency on recipients' creativity are to be investigated, a separate and novel experiment – ideally with active manipulation of the incentive system design – must be devised.

2.2.2 Threats to Experimental Validity

The internal and external validity of experiments is of particular importance (Döring and Bortz 2016). Internal validity is the truth of "interference about whether observed covariation between A and B reflects a causal relationship from A to B in the form in which the variables were manipulated or measured" (Shadish, Cook, and Campbell 2002, p. 53). In other words, the effects on the dependent variables have to be undoubtedly attributable to the effect of the independent variable. Therefore, plausible alternative explanations must be ruled out to ensure the internal validity of an experimental setting. Internal validity, in turn, is a prerequisite for external validity, which is defined as "the validity of inferences about whether the causal relationship holds over variation in persons, settings, treatment variables, and measurement variables" (Shadish, Cook, and Campbell 2002, p. 38). The external validity of an experiment indicates the extent to which the observed cause-effect relationship (respectively its strength) is transferable, generalizable, to other conditions. External validity typically decreases with increasing artificiality of the experimental setting compared to the natural environment of the target population (Döring and Bortz 2016). Nonetheless, low external validity does not argue against the existence of the observed effect, but can be considered as an extension of the research question (Döring and Bortz 2016), such as why the effect is (not) present in this particular setting.

Campbell and Stanley (1963) established a framework for confounding variables ("sources of invalidity") that threaten the internal and external validity of an experiment. In other words, these variables either display alternative explanations for differences in the dependent variable other than the manipulation of the independent variable (1-9, threats to the internal validity) or cast doubt whether the observed causal relationship between independent and dependent variables is generalizable (10-14, threats to the external validity). Considering the revision of Shadish, Cook, and Campbell (2002), whose work adds the ambiguity regarding the temporal procedure (1) and restructures the threats to external validity while retaining the Campbell and Stanley's (1963) original line of reasoning, these "threats" (variables) are presented and described[99] below:

Threats to the *internal validity* of the experiment:

(1) *Ambiguous temporal procedure*: When there is no clear chronological sequence of cause (precedeng) and effect (following), there is uncertainty as to which variable is the cause and which is the effect.

[99] The attached numerical values do not constitute a ranking. The variable description is also based on elaborations by Schnell et al. (2018), and Döring and Bortz (2016).

(2) *History*: External influences other than the manipulation(s) could affect the dependent variable. The greater the time interval between manipulation and variable measurement, the higher the risk of unrelated events.

(3) *Maturation*: Participants might change over the course of the experimental study, for instance, they get older, change their mood, or become more experienced in dealing with experimental tasks.

(4) *Testing*: An initial measurement could influence the outcome of a preceding measurement due to participants' practice, familiarity with the measurement or awareness of the purpose of the study.

(5) *Instrumentation*: Various measures used during the experimental study can yield various outcomes. These include measurement instruments, experimental material, such as instructions, but also experimenters who differ in gestures, voice pitch, or degree of concentration – among each other and, over time, also among themselves.

(6) *Regression*: Participants who show extreme traits (e.g., performance or mood) on an initial measurement could be statistically prone to show less extreme traits on a subsequent measurement ("regression to the mean"), regardless of the manipulation. This effect is particularly vulnerable when pretests are used for participant selection.

(7) *Selection*: It is not the manipulation(s) but the heterogeneity of participants between treatment groups that could drive differences in the dependent variable (e.g., age, gender or profession).

(8) *Mortality / Attrition*: Systematic differences in the "survival" of participants, e.g., only participants with certain attributes (e.g., highly motivated, high performing, healthy) fully complete the experiment or the experimental procedure influences terminations, could drive differences in the dependent variable. In the latter

case, these differences could then also be attributed to the heterogeneity of surviving participants between treatment groups.

(9) *Interaction*: The mentioned threats (to the internal validity of the experimental procedure) could occur not only individually, but could also interact in combination.

Threats to the *external validity* of the experiment:

(10) *Interaction of the causal relationship with units*: The observed causation could not be transferred to other types of participants. In pharmaceutical research, for example, experimental studies are conducted predominantly with young male participants. However, the observed effects of a drug (or the required dose) may not apply to female or elderly patients.[100]

(11) *Interaction of causal relationship over treatment variations*: The observed causation could not be transferred to other treatment designs. For instance, in an experimental setting on effort allocation, Hannan et al. (2013) found that when public relative performance information (RPI) is given, individuals perform worse when they can choose the proportion of time they spend on each task than when the time allocation is determined externally. However, these performance differences between these treatments are not present when a private or no RPI is given. If the authors had omitted the public RPI treatment, their results could have been (falsely) interpreted to mean that there are no performance differences between choosing or not choosing the time spent on each task, regardless of whether an RPI is provided (then only private RPI) or not.

[100] See https://www.spiegel.de/panorama/gendermedizin-medizin-forscht-fast-nur-an-maennern-mit-folgen-fuer-frauen-a-7d5bf557-f8dd-4cbc-ab2a-c02be3e777bd [retrieved 05-21].

(12) *Interaction of causal relationship with outcomes*: The observed causal relationship might not be maintained if other outcome observations are used. For instance, while the causation between a chosen feedback frequency and performance has been shown to be curvilinear (Holderness, Olsen, and Thornock 2020), the effects of feedback frequency on job satisfaction could be insignificant despite it is a known driver of employee performance (Judge et al. 2001).

(13) *Interactions of causal relationship with settings*: The observed causation could not be transferred to other settings. For instance, a field experiment conducted in a bank showed that a gamified anti-corruption training (i.e., elements of digital games were involved in the training) was preferred over traditional, non-gamified anti-corruption training and lead to better learning results (Baxter, Holderness, and Wood 2016). However, it is possible that this causation (respectively its strength) can only be replicated in work environments that are also otherwise poor in gamification.

(14) *Context-dependent mediation*: Unobserved variables that mediate the cause-effect relationship between independent and dependent variables could not convey the effect in different settings. For instance, a field experiment might show that if a supervisor reduces the frequency of his or her feedback, this does not cause a drop in employee performance because the observed employees respond by (actively) increasing the frequency of feedback they receive from their colleagues. However, this mechanism is not available to employees who work in departments with few or no colleagues.

More specifically, participants (e.g., age, gender, mood, prior experience), situational factors (e.g., daytime, illumination, background noise), or experimenter (e.g., behavior, gender, pitch of voice) practically constitute potential sources of variation among different experimental treatments – beyond the manipulation of the independent variable – that may influence the relationship between independent and dependent variables (Döring and Bortz 2016). Thus, the validity of an experimental setting increases with the rigor of controlling for these variables. Figure 2.17 provides an overview about systematic techniques that can be used to control for confounding variables.

Considering the implications of Shadish, Cook, and Campbell's (2002) framework for my experimental study (chapter 4), these techniques (where applicable) are routinely used in this study, e.g., all participants received standardized feedback, no additional information about the supervisors (i.e., feedback providers) other than their first names is divulged to eliminate the potential influence of other characteristics, such as age or physical appearance, participants are randomly assigned to treatments, or feedback valence is measured and included it as control variable in the further analysis (statistically controlling for its influence).

Technique	Description
Keep confounding variable constant	- Maximal standardization of the experimental procedure - Confounding variable has the same mean and no variance in all treatments - Applicable to sources of variation: situational factors and experimenter - Examples: The experiment is conducted at the same time in only one laboratory room (i.e., same background noise for all participants); only one experimenter is used for all conditions
Eliminate confounding variable	- Any influence of the confounding variable is prevented - Confounding variable is “zero” in all treatments - Applicable to sources of variation: situational factors and experimenter - Examples: Participants’ cell phones are collected and put away before the start of the experiment; the experimental procedure is introduced to the participants through a manual instead of an experimenter
Parallelize (match) confounding variable	- Prior to assignment to treatments, one (or multiple) participant-derived variable(s) are measured. Participants are then assigned to treatment groups in a way that ensures that this (these) previously measured variable(s) are similar in mean and variance across treatment groups - Confounding variable has a similar mean and variance in all treatments - Applicable to sources of variation: participants, but only for limited number of variables (otherwise the treatment assignment becomes too complex) - Example: In a pre-experimental survey, the age of the participants is measured. According to the age ranking, participants are then alternately assigned to the first or the second treatment, ensuring a similar age distribution in both treatments

Randomize	- Participants are randomly assigned to treatments - If treatments cells are sufficiently large (n > 30), systematic differences are unlikely for *any* confounding variables that originate from participants - *All* confounding variables originating from participants have similar mean and variance across all treatments - In contrast to parallelization/matching or statistical control, *no ex ante selection* of relevant confounding variables is required since *all* personal variables are homogenized - Applicable to sources of variation: participants - Example: If 200 students are randomly assigned to four experimental treatment groups, the groups should have similar mean and variance scores in terms of age, motivation, gender, income, sociodemographic background, etc.
Control confounding variable statistically	- Measurement of confounding variable that is later included as additional independent variable (i.e., control variable) in statistical assessment of causal relationship between independent and dependent variables - Confounding variable may differ between treatments regarding its mean or variance - Applicable to sources of variation: participants, situational factors, and experimenter - Example: Participants have to indicate their motivation for participating in the post-experimental questionnaire; the duration of the experimental sessions is measured (situational factor); demographic data (e.g., age, gender) of the experimenter are collected
Keep experimenter blind	- (Un)conscious influence of the experimenter is limited if he or she is not aware of the experimental condition he or she is supervising - Applicable to sources of variation: experimenter - Example: Experimenter welcomes participants to experimental session in the computer lab and explains (general) procedure, while the experimental manipulation is conveyed through information provided via the IT-application

Figure 2.17: Overview of Control Techniques for Confounding Variables[101]

[101] Adapted from Döring and Bortz (2016) and Schnell, Hill, and Esser (2018).

2.2.3 Elemental Experimental Design Choices

In addition to the control techniques described above, the general choice of environment in which the experimental study takes place has grave implications for its internal and external validity (also hereafter: Harrison and List 2004; Levitt and List 2009; Shadish, Cook, and Campbell 2002). While a laboratory experiment offers potential to control multiple confounding variables by creating an artificial environment, "external validity often is thought to be the Achilles heel of [laboratory] experimentation" (Sprinkle 2003, p. 289). Nonetheless, "conceding the lack of representativeness (external validity) the well-done laboratory experiment still has the fundamental prerequisite of any research: internal validity" (Kerlinger and Lee 2000, p. 581). Laboratory experiments are therefore frequently used in management accounting research (e.g., Holderness, Olsen, and Thornock 2020; Hecht, Newman, and Tafkov 2019).

Hence, I conduct a field experiment (Study 2) to take advantage of the (laboratory) experimental cause-effect assurance (internal validity) and the external validity of a field study. Since field studies take place in the natural environment of the participants, the absence of "artificiality" in the study context increases the generalizability of the study: "In this way, field experiments provide a useful bridge between laboratory and naturally occurring data in that they represent a mixture of control and realism usually not achieved in the laboratory or with uncontrolled data" (Floyd and List 2016, p. 438). It is therefore not surprising that field experiments have also recently become increasingly popular in management accounting research (e.g., Casas-Arce, Lourenco, and Martinez-Jerez 2017; Eyring and Narayanan 2018; Li and Sandino 2018).

Other hybrid experimental forms that also attempt to combine the advantages of experimental and field studies, such as artefactual field experiment (laboratory experiment with non-standard pool of individuals, e.g., managers or accountants), framed field experiment (laboratory experiment that builds on participants' prior knowledge and experience, e.g., participants whose (regular) job involves supervising employees must write work appraisals for (fictitious) employees), natural field experiment (experiment that takes place in natural environment of individuals who are normally unaware that they are part of an experiment and in which no active manipulation takes place), or natural experiment (ex post analysis of an exogenous "shock" that then demonstrates the experimental manipulation, e.g., reactions to the Corona crisis (shock) by firms whose business model or work processes allow for much or little home-office work[102]), are not considered an empirical methodology to investigate the research question for reasons such as lacking of active manipulation of the independent variable (e.g., feedback frequency) or limitations on the external validity of the experiment (e.g., restriction to non-standard participants).

Quasi-experiments constitute one special form of experiments (Döring and Bortz 2016; Shadish, Cook, and Campbell 2002). In contrast to real experiments, no (active) randomization of participants is possible or has been performed.[103] Instead, the treatments are undergone by predetermined groups. As previously discussed, this reduces the internal validity of the study, since other confounding variables originating from participants might drive the causation between independent and dependent variables. However, there are reasons why a quasi-experimental design may be

[102] While little attention has been paid to the effects of explicit external "shocks" in experimental management accounting research, this research method is more common in financial accounting research (e.g., Gippel, Smith, and Zhu 2015; Gow, Larcker, and Reiss 2016) and is related to event studies (e.g., Klein, Zwergel, and Heiden 2009).

[103] Therefore, quasi-experiments are also referred to as "non-randomized experimental studies". Some researchers even do not distinguish between quasi-experiments and natural experiments, as both do not randomize participants (e.g., Harrison and List 2004). However, quasi-experiments still involve active manipulation of an (possibly second) independent variable (Döring and Bortz 2016), so this work does not follow this line of reasoning.

unavoidable: First, the research question requires observing a variable that cannot be manipulated. For instance, when Wang (2017) examined the effects of personality traits (i.e., the dark triad) on counterproductive work behaviors, she could not actively manipulate levels of narcissism, Machiavellianism, or psychopathy because these traits are inherent to the participants. Similarly, when testing some hypotheses in Study 2, the recipient gender cannot be manipulated but serves as a variable to determine experimental groups. Second, the existing structure in the pool of participants must be adopted for pragmatic reasons. To give an example, Presslee, Vance, and Webb (2013) manipulate incentives for employees (cash-reward vs. points equal to the retail value of cash rewards), but participants' assignment to either treatment was determined by their departmental affiliation. Third, the experimental procedure can generate (intermediate) results, which in turn can be used as variables for the analysis of the experiment. Specifically, in my later experimental field study (chapter 4), feedback valence is used as (control) variable for the influence of feedback frequency on recipients' creativity. Since valence depends only on individuals' (prior) performance, it cannot be actively manipulated.

3. Study 1: Multiple Feedback Sources and Mediation of Feedback Frequency Effects[104]

3.1 Introduction

The first study of this dissertation aims to examine the individual and joint effects of feedback frequency from multiple sources on employee creativity and to uncover the intrapersonal processes that convert feedback frequency into employee creativity. In doing so, the study was designed to integrate into and contribute to service research literature, as service research has devoted much attention to fostering the creativity of (service) employees (for an overview see Schepers, Nijssen, and van der Heijden 2016) and involving customers in the service development process (Vargo and Lusch 2008a). The boundary-spanning position of service employees in particular, i.e., bridging perspectives and being confronted with potentially conflicting demands from inside (e.g., supervisor) and outside (e.g., customer) the organizational unit (Agnihotri et al. 2014a; Coelho, Augusto, and Lages 2011; Yoo and Arnold 2016), has been impetus for service research to incorporate customer knowledge, perception, and perspective into the generation of novel (service) ideas (Storey and Larbig 2018), especially as creative service ideas are considered critical to satisfying continuously changing customer needs (Agnihotri et al. 2014a). This approach stands apart from feedback research, that has largely focused only on the supervisor as the primary source of feedback. In addition, service research has provided suggestions for the interplay of feedback from multiple sources, namely supervisor and customers (Agnihotri et al. 2014a; Challagalla and Shervani 1996; Wilder, Collier, and Barnes 2014), which has been surprisingly neglected by feedback research. In conclusion, the research objectives of this study integrate well with the context and current discussions of service research.

[104] This chapter builds on a joint paper written by the examinee and his supervisor, Prof. Dr. Nevries.

Nonetheless, this study contributes to the broad strand of feedback literature as well, not only because it is the first to examine the effects of feedback frequency on employee creativity, but also because it uncovers the intrapersonal processes that (fully) mediate these effects and because it identifies multiple functions of feedback frequency (e.g., Ilgen, Fisher, and Taylor 1979).

Faster-spinning market changes and increasingly fine-grained customization conspire to demand firms to come up with creative responses more often. More than ever, creative employees are key to the survival and effectiveness of an organization (Oldham and Cummings 1996), and service employees[105] (hereafter SE) are at the epicenter of these challenges, “consistently exposed to ever-changing environments” (Agnihotri et al. 2014b, p. 57) through their unique boundary-spanning position. In a move to avoid information becoming increasingly outdated, trigger thought processes, and create more opportunities to develop new ideas, many firms are speeding up their feedback cycles. This is argued to be a suitable and effective solution, as SEs with a higher number of feedback contacts are better enabled and motivated to come up with new ideas (Adler and Borys 1996; Coelho and Augusto 2010), avoiding periods of unwanted creative dormancy. Somewhat logically, studies argue that “the more frequent the feedback, the better” (Ilgen, Fisher, and Taylor 1979, p. 354; Hackman and Oldham 1975; Lam et al. 2011; Johlke and Duhan 2000).

This however might turn out as an ill-advised decision for service managers, as a number of recent publications point to the harmful effects of frequent feedback such as information overload, myopia, perceptions of being controlled and monitored, loss

[105] Consistent with the study’s integration with service research and observation of service employee behavior, we refer to service employees as a representative group of employees in this chapter.

of autonomy, and additional ambiguities and conflicts (Casas-Arce, Lourenco, and Martínez-Jerez 2017; Holderness, Olsen, and Thornock 2020), which curtail behavioral deviations serving as breeding ground for creativity (Amabile, Goldfarb, and Brackfield 1990). As every service manager has to decide about how frequently to engage in the time-consuming task of providing feedback – even no feedback has its frequency – in a faster-paced environment, it is surprising that extant service literature has yet to substantiate informed feedback strategy choices. To fill this gap, one of the overarching goals of this study is to explore the implications of feedback frequency for service employee creativity (hereafter SEC[106]).

Furthermore, we need to recognize the SEs' unique boundary-spanning position and go beyond previous studies that are mostly restricted to a single source of feedback (i.e., supervisor, Alvero, Bucklin, and Austin 2001). Literature has repeatedly stressed the importance of distinguishing different sources of feedback (e.g., Ilgen, Fisher, and Taylor 1979). While supervisors are proclaimed as the prime providers of feedback working proximally to and commanding formal authority over the employee as a feedback receiver (e.g., Greller and Herold 1975), SE's boundary-spanning position also inherently provides privileged access to first-hand customer feedback. Supervisors and customers occupy juxtaposed work positions which means they provide disjunct non-redundant and thus potentially highly effective additive feedback (Dokko, Kane, and Tortoriello 2013). Importantly, they likely provide feedback at different points in time, yielding a more dynamic perspective on how feedback frequencies interact to drive SEC. As Sijbom et al. (2018, p. 355) note, by "gaining diverse viewpoints from others […] and cognitively processing and integrating these viewpoints with their own, individuals should be able to increase their creative

[106] Consistent with the study's integration with service research and observation of service employee behavior, we synonymously refer to employee creativity as service employee creativity in this chapter.

performance". Still, in practice it can be observed that some supervisors intentionally prevent customers from providing feedback to SEs. Such behavior may be rooted in an attempt to safeguard their power and control in the "three-cornered fight" among SE, customer, and supervisor (Bateson 1985), but also to avoid confusion and role stress among SEs (Zeithaml, Berry, and Parasuraman 1988). Despite the plethora of previous publications from diverse strands of literature, research has yet to clarify the joint effects of feedback frequencies originating from supervisors and customers.

To address these research gaps and support more informed managerial choices on how frequently service managers and customers should provide feedback to help SEs unlock their creative potential, we develop a conceptual framework building on dynamic feedback-specific Stimulus-Organism-Response theory and role theory. The resulting moderated mediation analysis model (see Figure 3.1) was empirically tested by using cross-sectional questionnaire-based responses of SEs from 385 firms operating across all major industries.

Our study complements previous output in several ways. First, we contribute to a large strand of literature that centers around SEC (for an overview see Schepers, Nijssen, and van der Heijden 2016). We introduce the frequency of feedback as an important complementary driver of SEC, building on recent insights from publications in a variety of disciplines that have pointed to the potential of the frequency of feedback in effectively stimulating employee responses (Casas-Arce, Lourenco, and Martínez-Jerez 2017; Holderness, Olsen, and Thornock 2020; Thornock 2016). Indeed, our results demonstrate how powerful – and potentially harmful – this choice is for managing SEC.

Second, service literature has made much of involving the customer in the service generation process. From a service-dominant logic perspective in particular it seems mandatory and inherently appropriate for the SE to receive feedback from their customers (Vargo and Lusch 2008a). Indeed, value co-creation in the service generation process is argued to rest upon integrating customer information, as SEs need to understand customers to motivate creativity guided towards effective customer solutions. Recent publications acknowledge that “the understanding of how a firm absorbs new customer knowledge during service development is limited” (Storey and Larbig 2018 p. 104) and call for a more nuanced analysis of customer involvement (Cabiddu, Frau, and Lombardo 2019; Ostrom et al. 2015). We contend that our choice of introducing the customer as a complementary feedback source enriches this discussion, as it goes beyond previous studies which have only recognized a single feedback source (Agnihotri et al. 2014a; Coelho and Augusto 2010; Johlke and Duhan 2000; for an exception, see Siahtiri 2018). Indeed, our results indicate the need of a selective involvement of customers.

Our study design also allows feedback interplay and thereby delivers a new layer of understanding feedback functions. This approach is noteworthy as more frequent supervisor feedback may alter SEs’ perceptions of each customer feedback, and vice versa, through reducing delay, short-term memory losses between respective feedbacks and strengthening the ability of sensemaking. Indeed, several authors suggest that an explanation function of supervisor feedback enables SEs to better understand customer feedback (Agnihotri et al. 2014a; Challagalla and Shervani 1996; Wilder, Collier, and Barnes 2014), yet there is a lack of empirical proof. We uncover explanation, information, motivation and controlling functions that specify the dynamics of supervisor and customer feedback. In a grander scheme, our results add to the knowledge regarding the boundary spanning role of SEs (Agnihotri et al. 2014a; Yoo and Arnold 2016).

Third, research has yet to look behind the façade of how SEs perceive feedback frequencies to understand how they change their creative response. Role theory guides us in recognizing the importance of the role stressors role conflict and role ambiguity as pivotal mediators of the relationship between feedback frequency and creativity (Coelho, Augusto, and Lages 2011; Solomon et al. 1985). While there seems to be consensus in service literatures on the substantial implications of role conflict and role ambiguity for SEC (Bettencourt and Brown 2003; Coelho, Augusto, and Lages 2011; Schepers, Nijssen, and van der Heijden 2016), our results reveal that these effects change with the level of the frequency of feedback. Contrary to expectations that more frequent feedback originating from two sources following their own agenda (Yoo and Arnold 2016) provokes creativity-stimulating inter-sender conflicts, our results do not show a consistently negative relation.

3.2 Conceptual Framework

We utilize the dynamic Stimulus-Organism-Response (S-O-R) framework (Jacoby 2002) and role theory (Katz and Kahn 1978) to structure our conceptual model that aims to explore the relationship between feedback frequency and SEC (see Figure 3.1). The dynamic S-O-R framework explains how an external stimulus elicits behavioral responses by recognizing affective and cognitive processes within the individual (organism), and intertemporal linkages between multiple stimuli. Since the genesis of creative ideas is an intrapersonal process that is responsive but imperceptible from the outside, the S-O-R framework is particularly well-suited to recognize the complex translation mechanisms that link feedback (stimulus), role stressors (organism), and SEC (response). Our approach thus allows for much-needed further glimpses behind the façade of SE perceptions of feedback frequencies as well as the responses to it.

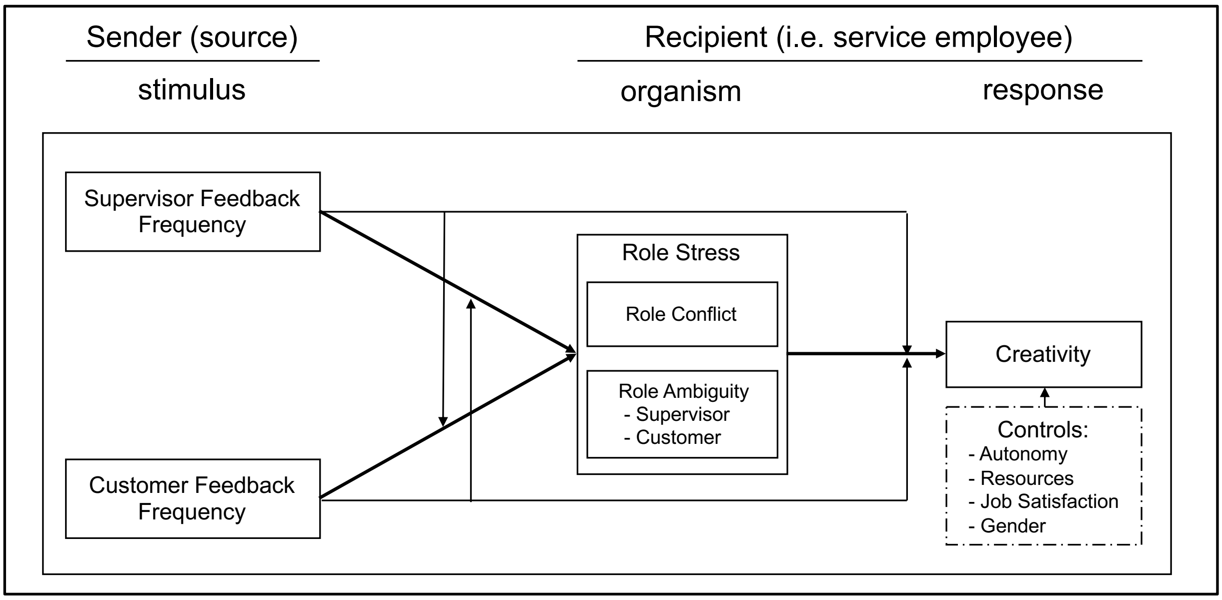


Figure 3.1: Conceptual Model building on the S-O-R Framework and Role Theory

Feedback is an important stimulus to affect service personnel behavior (Ilgen, Fisher, and Taylor 1979) “as feedback provides information […], it stimulates the employee to explore different courses of action in the pursuit of a better result” (Coelho and Augusto 2010, p. 429). Ilgen, Fisher, and Taylor (1979) describe that the perception of a feedback stimulus depends on the choice of a source (i.e., sender) and frequency of the feedback stimulus. Feedback frequency displays how frequently this stimulus-organism-response process occurs. As each feedback stimulus is associated with biunique cognitive, emotive and behavioral consequences, the higher the frequency of the communicative stimuli, the more often the recipient (i.e., SE) runs through the stimulus-organism-response process. Therefore, it matters how often feedback is provided.

The dynamic S-O-R framework also explains that more frequent feedback stimuli develop a dynamic relationship that transcends additive outcomes (Jacoby 2002). Since each feedback process changes the recipients’ set of “prior experiences, knowledge, beliefs, […] cognitive networks” or expectations (Jacoby 2002, p. 54), it also forms a different sounding board for the next feedback stimulus. Indeed, the recipient has by then morphed into a different state that changes the processing of a subsequent stimulus. For instance, supervisor feedback that mandates the SE to increase the efficiency of responding to customer emails allows the SE to translate customer feedback complaining about excessive reaction times into a creative service solution such as programming an automatic acknowledgement of receipt. Without the previous feedback stimulus from the supervisor the SE may have resorted to a scripted behavior (e.g., manually confirming incoming orders).

Higher feedback frequencies, therefore, do not only translate into higher numbers of stimulus-organism-response processes, but ceteris paribus also reduce the time interval to the previous feedback stimulus (Thornock 2016), such that there is a higher likelihood for the feedback to be more timely and more relevant (Ilgen, Fisher, and Taylor 1979). This prevents the provided information from becoming obsolete or superseded by disruptive events, and being lost in short-term memory (Crowder 2014). Furthermore, a higher feedback frequency implies reduced average delays of receiving a feedback stimulus after an event, and thus yields more time for creative ideas to emerge, or to translate the stimulus into a high-creativity idea (Kachelmeier, Wang, and Williamson 2019). Nonetheless, studies have also uncovered that frequent feedback may appear as a relentless flux of interruptions that exacerbates myopic behavior (Casas-Arce, Lourenco, and Martínez-Jerez 2017), which jars with time necessary for creative ideas to build – referred to as creative incubation (Kachelmeier, Wang, and Williamson 2019).

As feedback is a person-to-person interaction, there is a general recognition in the feedback literature regarding the importance of specifying the source (Greller and Herold 1975). In service research, much has been made of recognizing customers and supervisors as key counterparts of the SE (e.g., Bateson 1985), however, this did not translate into jointly studying customers and supervisors as sources of feedback that impact on SEC. Yet it is important to recognize that both feedback sources provide complementary starting points for initiating creative ideas. On one hand, the supervisor has to assess compliance with the specifications of the service process and can provide information regarding global transparency (Adler and Borys 1996). On the other hand, the customer is unique in providing the SE with insights into customer judgments and the customer process (Storey and Larbig 2018) so that SEs "receive evaluations of their performance from the people best able to render an appraisal – the customers they just served" (Markey, Reichheld, and Dullweber 2009, p. 44). Similar to SEs, supervisors cannot fully put themselves in the customer's shoes. Thus, neither supervisors nor customers can substitute each other's unique relation to the SE, which is defining for the SE's boundary-spanning role.

Our conceptualization of mediators (organism) is guided by role theory given that role theory is suited to "explain employee reactions to feedback" (Bell and Luddington 2006, p. 222). According to role theory, role senders communicate their perceived compliance of a recipient's behavior with a role senders' expectations to recipients through feedback (Katz and Kahn 1978). Specifically, role theory literature highlights that feedback stimuli influence role stress (e.g., Malhotra and Ackfeldt 2016; Singh 1993; Solomon et al. 1985). We set the focus of our study on inter-sender role conflicts and role ambiguity as they are a likely outcome of receiving feedback from both supervisor and customer and have been shown to provoke and inhibit creative ideas in service literature (e.g., Coelho, Augusto, and Lages 2011; Schepers, Nijssen, and van der Heijden 2016). Utilizing the structure of the dynamic S-O-R model and specifying the perception of the feedback receiver (organism) is critical for exploring the feedback frequency-SEC link.

3.3 Hypotheses Development

3.3.1 Feedback Frequency and Role Stressors

Particularly in the context of boundary-spanning SEs, the emergence of inter-sender role conflicts appears inherent, as incompatible expectations “make it difficult, if impossible, for the worker to meet concurrently” (Coelho, Augusto, and Lages 2011, p. 34; Bell and Luddington 2006; Zeithaml, Berry, and Parasuraman 1988). Since feedback transmits the extent of accordance of the recipient’s behaviors with the role sender’s expectations (Katz and Kahn 1978), this information is more often associated with higher feedback frequency. For instance, repeated customer feedback that demands extended availability of a firm’s service hotline makes call center SEs – who are aware that their supervisor disapproves overtime – more conscious of the incompatibility of these demands.

A dynamic perspective of the feedback provision process, however, also allows SE and the feedback source to align their role expectations during repeated interactions (Malhotra and Ackfeldt 2016; Ng, Plewa, and Sweeney 2016). This leads us to hypothesize:

Hypothesis 1a: Supervisor feedback frequency is positively related to SE role conflict

Hypothesis 1b: Customer feedback frequency is positively related to SE role conflict

We enter established research territory with respect to the relationship between feedback and role ambiguity**.** Role ambiguity reflects “the extent to which a person is uncertain about the expectations of relevant role partners” (Challagalla and Shervani 1996, p. 91) and “how to satisfy those expectations” (Zeithaml, Berry, and

Parasuraman 1988, p. 43). Therefore, role ambiguity is caused by a lack of information about role partners (Singh 1993).

Consequently, high levels of feedback frequency reduce role ambiguity since “regular manager feedback is the mechanism by which employees continually recalibrate their understanding of their organizational roles” (Agnihotri et al. 2014a, p. 175; Coelho and Augusto 2010). Indeed, Zeithaml, Berry, and Parasuraman (1988, p. 43) highlight that “the more frequently managers provide clear and unambiguous communication about these topics, the lower employees' role ambiguity will be”. We argue that SEs benefit from frequent feedback in contemporary organizations where role expectations are likely to be constantly changing. Thus, we suggest the following hypotheses:

Hypothesis 2a: Supervisor feedback frequency is negatively related to SE’s supervisor role ambiguity

Hypothesis 2b: Customer feedback frequency is negatively related to SE’s customer role ambiguity

3.3.2 Interaction Effects: Feedback Frequencies and Role Stressors

Service literature has made much of the juxtaposed requirements of supervisors and customers that confront boundary-spanning SEs. The more often they give feedback, respectively, the more often are they able to comment on feedback from other sources, elaborate on additional, incompatible facets between their role expectations, or stress the dominance of their feedback information, which in turn increases the perception of the SE of an enlarged gulf between each role expectation. For example, frequent feedback from the supervisor outlining efficiency requirements reduces the SE’s flexibility during interactions with the customer (Coelho, Augusto, and

Lages 2011; Singh 1993). Following this logic, frequent feedback from multiple sources conspire to cause inter-sender role conflict. We argue that the more often customers and supervisors provide feedback, the more likely is a smaller time gap between feedback from both sources leading to a more pronounced perception of inter-sender role conflict. On the other hand, the supervisor may be able to help SEs to make sense of and relate to customer communication (Agnihotri et al. 2014a; Challagalla and Shervani 1996; Wilder, Collier, and Barnes 2014), thereby reducing inter-sender conflicts. We argue as follows:

Hypothesis 3: Supervisor feedback frequency moderates the relationship between customer feedback frequency and role conflicts such that more supervisor feedback frequency increases the positive relationship between customer feedback frequency and role conflicts

The conceptual considerations of several papers underscore the importance of supervisor feedback to an SE's understanding of customer communication (e.g., Challagalla and Shervani 1996; Wilder, Collier, and Barnes 2014), while empirical insights into the interactional effects of supervisor feedback are still limited (e.g., Eva et al. 2019; Siahtiri 2018). The underlying rationale is that the more often supervisors provide SEs with feedback, the higher is the probability that it is received shortly after customer communication, so that the SE is able to reproduce the original, almost verbatim, content. Under these circumstances, supervisor feedback is of particular value as it explains and clarifies the customer's statements to the SE. Indeed, supervisor feedback enables SEs "to respond more quickly to customer concerns, have more meaningful customer interactions, and most importantly, offer more innovative solutions" (Agnihotri et al. 2014a, p. 171; Lages and Piercy 2012). Along the same lines, Johlke and Duhan (2000, p. 162) suggest that „the extent of customer contact […] may moderate the effects of managerial communication". We subscribe to this line of thinking and hypothesize:

Hypothesis 4a: More customer feedback frequency increases the negative relationship between supervisor feedback frequency and supervisor role ambiguity

Hypothesis 4b: More supervisor feedback frequency increases the negative relationship between customer feedback frequency and customer role ambiguity

3.3.3 Role Stressors on Creativity

Drawing on role theory, we posit that role conflict stimulates SEs to generate novel and creative ideas. When "no standard response […] is able to satisfy the contradictory demands encountered" (Schepers, Nijssen, and van der Heijden 2016, p. 801), searching for novel ideas displays the only solution to "address perceived incompatibility among the expectations of their role partners" (Coelho, Augusto, and Lages 2011, p. 34). Thus, role conflict "activates them to develop ideas […] to rethink current practices and come up with ideas to update service scripts" (Schepers, Nijssen, and van der Heijden 2016, p. 801).

Hypothesis 5: Role conflict is positively related to SEC

Role ambiguity hinders SEs "from fully using their expertise and creative-thinking skills in executing their jobs, thus negatively impacting on creativity" (Coelho, Augusto, and Lages 2011, p. 34). More specifically, a clear understanding of customer preferences represents a prerequisite to generate creative ideas for customizing the service in a better manner (Lages and Piercy 2012). This is of critical importance as role ambiguity has been associated with reduced job satisfaction, psychological withdrawal and decreased motivation to engage in generating novel ideas (Bettencourt and Brown 2003).

As previous studies observing the relationship between role ambiguity and SEC do not distinguish between customer and supervisor role ambiguity, we are particularly interested whether previous results indicating a negative relationship are confirmed for both SE role partner ambiguities, and whether one is dominant regarding its implications for SEC.

Hypothesis 6a: Supervisor role ambiguity is negatively related to SEC

Hypothesis 6b: Customer role ambiguity is negatively related to SEC

3.3.4 Interaction: Role Stressors, Creativity, and Feedback Frequencies

Role theory suggests that the frequency of feedback affects SE's behavioral reaction (i.e., creative activities) to inter-sender role conflicts. A basic premise of role theory is that recipients strive to avoid negative feedback, so that role conflict displays a situation in which SEs generally have to decide which role partner's needs are given preferential consideration, and frequent feedback raises the tension within the SE to fulfill the feedback source's needs and to omit the conflicting needs of the other role partner.

Given that the supervisor commands hierarchical power over the SE and can control how the SE utilizes resources, it has been argued that "feedback is all too frequently controlling" and "emphasizes how people should behave and implies that the manager is in control" (Deci, Connell, and Ryan 1989, p. 585). More frequent supervisor feedback may force SEs to follow the supervisor's directions instead of searching for creative solutions to satisfy the needs of both role partners.

Similarly, we expect that this rationale applies to the frequency of customer feedback, such that a higher frequency of customer feedback curtails the positive implications of SE’s role conflict for creativity. However, it is unclear whether the customer has sufficient power over the SE given that “the more the source is seen to control valued outcomes, the more likely is the recipient to try to respond to feedback from that source. Increased power should increase compliance from the recipient” (Ilgen, Fisher, and Taylor 1979, p. 359).

Hypothesis 7: Feedback frequencies moderate the relationship between role conflict and creativity such that (a) higher supervisor feedback frequency and (b) higher customer feedback frequency reduces the positive relationship between role conflict and SEC

We follow insights from the creativity and feedback literatures to argue that the relationship between role ambiguity and creativity depends on how often feedback is provided. Receiving feedback with less delay (i.e., more frequent feedback) allows SEs to better exploit less role ambiguity to work on creative ideas both faster as well as more extensively.

SEs that are capable of generating novel ideas because of low levels of role ambiguity will engage in creative activities more often when they receive more frequent feedback, as “the frequency of feedback should enhance motivation on the task” (Ilgen, Fisher, and Taylor 1979, p. 363), inspire confidence (Agnihotri et al. 2014a) and is “perceived as a limited resource that has value” (Fedor and Buckley 1987, p. 173) that strengthens a SE’s desire to recompense. We therefore conclude:

Hypothesis 8: Higher supervisor feedback frequency strengthens the negative relation between (a) supervisor role ambiguity and (b) customer role ambiguity and SEC

Hypothesis 9: Higher customer feedback frequency strengthens the negative relation between (a) supervisor role ambiguity and (b) customer role ambiguity and SEC

3.4 Methodology

3.4.1 Sample and Data Collection

To test the hypothesized relationships, we collected data from management accountants. We selected this internal service provider (Hogreve et al. 2017; Schneider and Bowen 2019) as management accountants constitute one of the most fundamental service departments for firms and their *raison d'être* is to provide services to "customers [that] are internal to the company" (Heskett et al. 2008, p. 128). Management accountants collect, process, and supply information as the basis for managerial decision-making (e.g., providing regular and ad hoc analyses, cost and efficiency information, and offering specialized tools) and customer interactions are, as such, a fundamental part of their day-to-day role. Managers reflect ideal internal customers, e.g., concerning the restriction of changing service partners (Hsieh et al. 2013).

The finalized questionnaire was sent to German management accountants by personalized e-mails. In total, 473 responses were obtained[107] (response rate of 63.9%). 88 responses were deemed unusable, owing to incomplete questionnaires, straight-lining, and inadequate participants, such as those not currently working as

[107] The data was collected through the WHU Controller Panel. We gratefully acknowledge the substantial help of Christian Pfennig in gathering empirical data.

management accountants, resulting in a sample size of 385 (adjusted response rate of 52.1%).

Moreover, the descriptive data showed that the respondents work for a variety of companies across all major industries. Whereas some prior studies restrict their observations to a limited number of industries, our aim was to pool individuals across multiple company sizes and industries to contribute to the generalizability of findings. More specifically, respondents came from diverse industries such as industrial companies (24.6%), media and IT (23.8%), or consumer goods retailing (13.5%), had an average job experience of 9.4 years[108], and were employed by firms of various sizes regarding number of employees and total sales (cf. Figure 3.2).

SE's Firm Representation	Percentage
Industry	
Industrial company	24.6%
Media and IT	23.8%
Chemical and Pharmaceutical	13.8%
Consumer Goods Retailing	13.5%
Logistics	13.0%
Banking and Insurance	11.3%
Sales (Mio. EUR)	
< 100	38.3%
100-1,000	30.3%
> 1,000	31.4%
Employees	
< 500	40.5%
500-2,000	31.9%
> 2,000	27.6%

Figure 3.2: Sample Characteristics

[108] To assess experience, respondents were asked how long they have been employed in their current firm (in years).

A comparison of early and late respondents (each 25% of the sample) was completed to detect a possible non-response-bias (Armstrong and Overton 1977). We did not find any significant differences in the relevant variables. Furthermore, a χ^2-test showed that the drawn sample is not significantly different from the original mailing list regarding age, industry, and firm size.

3.4.2 Assessment of Common Method Bias

Because the data was collected from the same individuals at the same point in time, there is an eligible concern that common method variance may be present (Podsakoff et al. 2003). As such, we followed several procedural remedies as suggested by the literature (Podsakoff et al. 2003). First, we ordered the questions for dependent and independent variables in such a manner that the respondents could not draw conclusions regarding the underlying conceptual model of the research. Second, respondents were informed that there are no right or wrong answers, and they were encouraged to respond honestly and comprehensively. Third, we ensured and communicated anonymous responses by utilizing self-administered questionnaires as opposed to personal data collection. Fourth, we largely used prevalidated measurement scales and extensively pretested adaptions made to ensure comprehension and to avoid social desirability biases.

Furthermore, we tested for common-method-bias using Harman’s single-factor test (Lages and Piercy 2012; Podsakoff et al. 2003) and employed a full collinearity approach (Kock 2015) with results suggesting that common method variance did not affect our results adversely. We also controlled for potential common method variance with the marker variable test (Rönkkö and Ylitalo 2011) by including career orientation as a marker variable in our model because it is theoretically unrelated with any of our variables. The results show that including the marker variable in our model does not alter path significance levels, further alleviating concerns that common method bias accounted for the results.

3.4.3 Measures

All scales used in this survey were drawn from existing scales and were measured with multiple items on a 7-point Likert-type scale ranging from "does not apply at all" to "fully applies" (cf. Figure 3.3).

Latent Variable	Measures (Standardized Loadings)
Supervisor Feedback Frequency (α = 0.91, CR = 0.94, AVE = 0.79)	My supervisor often lets me know how well I am doing my job (.91)
	I always receive information from my supervisor about the results of my work (.91)
	Anytime, I have the opportunity to get feedback from my supervisor about my job performance (.79)
	My supervisor often lets me know about my job performance (.94)
Customer Feedback Frequency (α = 0.89, CR = 0.93, AVE = 0.76)	My customers often let me know how well I am doing my job (.89)
	I always receive information from my customers about the results of my work (.90)
	Anytime, I have the opportunity to get feedback from my customers about my job performance (.76)
	My customers often let me know about my job performance (.93)
Supervisor Role Ambiguity (α = 0.89, CR = 0.90, AVE = 0.76)	I know exactly what my supervisor expects of me (rc) (.92)
	The expectations of my supervisor often are not clear to me (.83)
	I receive clear instructions from my supervisor about my job duties (rc) (.84)
	I know the demands of my supervisor very well (rc) (.90)
Customer Role Ambiguity (α = 0.84, CR = 0.89, AVE = 0.68)	I know exactly what my customers expect of me (rc) (.87)
	The expectations of my customers often are not clear to me (.72)
	I receive clear instructions from my customers about my job duties (rc) (.82)
	I know the demands of my customers very well (rc) (.87)

Role Conflict (α = 0.85, CR = 0.91, AVE = 0.77)	I try to meet conflicting demands of various departments and people (.88)
	I have to deal with or satisfy too many different people (.88)
	I receive incompatible requests from two or more people (.87)
Creativity (α = 0.85, CR = 0.90, AVE = 0.69)	I make constructive suggestions for service improvements regarding new accounting tools and services (.75)
	I make constructive suggestions for service improvements regarding service processes (.83)
	I share creative solutions to customer problems with other unit members regarding service processes (.87)
	I encourage coworkers to contribute ideas and suggestions for service improvement regarding service processes (.85)
Resources (formative)	The technical resources in my department allow me to perform my duties without any problems
	The personnel resources in my department allow me to perform my duties without any problems
Autonomy (α = 0.90, CR = 0.93, AVE = 0.77)	My job frequently permits me to decide for myself how to go about doing work (.90)
	My job provides me with a very high degree of freedom in executing the tasks assigned to me (.90)
	In my job, I can very often use my own judgement to carry out my work activities (.91)
	In my job I often have the opportunity to independently define my own area of responsibility (.79)
Job Satisfaction (α = 0.93, CR = 0.95, AVE = 0.82)	I am very pleased with my job overall (.88)
	Altogether, my job is very close to my ideal of a perfect workplace (.93)
	Compared to other jobs where I could use my qualifications, I like my job much better overall (.90)
	All in all, I am very happy with my job (.91)
Note: α = Cronbach's alpha; CR = composite reliability; AVE = average variance extracted; rc = reverse coded.	

Figure 3.3: Measurement Models

Service employee creativity was measured with items from Lages and Piercy (2012) that emphasize the SE's communication of ideas and contribution to service improvements for customer problems. Due to comments in our pre-study, we added one item addressing creativity regarding the service process. Similar to previous research (e.g., Agnihotri et al. 2014a; Schepers, Nijssen, and van der Heijden 2016), creativity has been assessed with a self-report measure since "an individual's creativity is unlikely to be accurately assessed by any observer" (Coelho, Augusto, and Lages 2011, p. 41).

Building on the work of Eisenberger, Rhoades, and Cameron (1999) we derived items for feedback frequency from established feedback measures contained in the job diagnostic survey (Hackman and Oldham 1975) and in the job characteristics inventory (Dubinsky and Skinner 1984). The two reflective measurements of supervisor feedback frequency and customer feedback frequency, respectively, each contain five items that only differ regarding the feedback source's denotation in the respective items.[109]

For role ambiguity, we adopted five items for our measurement from Hartline and Ferrell (1996), Rizzo, House, and Lirtzman (1970) and Singh (2000), before reverse-coding them. In alignment with Singh's work (1993, p. 14) which highlights that "feedback is likely to have *localized* effects", we account for the multifacetedness of role ambiguity by specifying supervisor role ambiguity and customer role ambiguity separately. The measurement of role conflict encompasses four items from Rizzo, House, and Lirtzman (1970) and Singh (2000).

In order to rule out alternative explanations we gather further data about SEs' resource endowment, job autonomy, gender, and job satisfaction based on established measures.

---

[109] Specifically, in the questionnaire we ask for feedback from the "direct supervisor" such that we control for feedback source's power distance to the recipient (Ilgen, Fisher, and Taylor 1979), and we explicitly refer to feedback from "internal customers" since management accountants occasionally might have external customers as well, e.g., public institutions.

3.4.4 Data Analysis

To assess the consistency and validity of our measurement, the items were analyzed using partial least squares structural equation modeling (PLS-SEM; SmartPLS 3.0, Ringle, Wende, and Becker 2015). Following the guidelines of Hair et al. (2012) and Hair et al. (2019) for reflective outer model evaluation (as outlined in chapter 2.1.4.2), we established indicator reliability, internal consistency and convergent validity. Regarding the evidence for discriminant validity, we assessed the conservative heterotrait-monotrait (HTMT) criterion (Henseler, Ringle, and Sarstedt 2015). All disattenuated correlations are below the threshold of 0.85 and each indicator loads highest on the intended construct.[110] Therefore, discriminant validity is established.

3.5 Results

Structural equation modeling (SEM) was employed to test the hypotheses. Variance-based partial least squares (SmartPLS 3.0, Ringle, Wende, and Becker 2015) was used due to the leniency regarding distributional assumptions, the ability to handle complex structural models, as well as the availability of complementary analytical techniques such as FIMIX-PLS and PLS-POS.

3.5.1 Unobserved Heterogeneity

Prior to the analysis of the structural equation model, we applied the FIMIX-PLS approach to explore unobserved heterogeneity within the inner structural model (Becker et al. 2013; Hair et al. 2016; Janka and Guenther 2018). FIMIX-PLS creates transparency regarding unobserved heterogeneity, which – though likely to occur –

[110] Tested with bootstrapping (5,000 samples) and 95% confidence intervals.

otherwise remains hidden. To do so, FIMIX-PLS aims to disentangle different populations behind the aggregated sample, which might lead to severe misinterpretations if opposing effects of the underlying subpopulations remain undetected (see chapter 2.1.2.4).

First, following the guidelines of Hair et al. (2016) we utilized FIMIX-PLS to assess an adequate number of segments. Accordingly, we referred to Akaike's information criterion (AIC_3; AIC_4), the Bayesian information criterion (BIC), the consistent Akaike's information criterion (CAIC), and the normed entropy (EN) as criteria for determining an appropriate number of segments. The respective results of the FIMIX-PLS procedure for different numbers of segments are displayed in Figure 3.4, indicating that a two-segment solution should be considered. The results are robust to the exclusion of the feedback frequency moderator terms in the structural model.

No. of Segments	**2**	**3**	**4**
AIC_3 (Modified AIC with Factor 3)	4038.91	**4018.34**	4019.12
AIC_4 (Modified AIC with Factor 4)	**4087.91**	4092.34	4118.12
BIC (Bayesian Information Criteria)	**4183.62**	4236.88	4311.50
CAIC (Consistent AIC)	**4232.62**	4310.88	4410.50
EN (Entropy Statistic Normed)	**0.90**	0.76	0.66

Note: Following Hair et al. (2016) bold values mark the optimal value for the respective evaluation criteria.

Figure 3.4: Evaluation Criteria for FIMIX-PLS

Second, PLS-POS was employed to identify a potential explanatory moderator variable, which can determine relevant subpopulations (Becker et al. 2013; Matthews et al. 2016). PLS-POS gradually reallocates observations between the segments in order to maximize the explained variance of all endogenous latent variables in the structural model and is better able to identify the underlying segment structure than FIMIX-PLS (Ringle et al. 2013). In an exploratory analysis, we find a significant correlation ($p < 5\%$) between the suggested PLS-POS segmentation and the mean-split distribution of SE's experience (Becker et al. 2013, p. 668). This indicates that the experience of a SE has a decisive influence on the effectiveness of feedback frequency onto creativity. Again, this result is robust to the exclusion of the feedback frequency moderator terms in the structural model. FIMIX-PLS discrete partition supports the decisive character of the SE's experience as the overlap with its mean-split classification is above the cut-off value of 60% (Matthews et al. 2016). Subsequent data analysis reveals that reliability and validity are established in accordance with the previously mentioned criteria for the measurement instruments in both subsamples.

3.5.2 Evaluation of Structural Model and Testing of Hypotheses

Results from our structural equation analysis provide support for the conceptual model. Figure 3.5 provides an overview of the path coefficients significance levels with respect to the level of SE experience.[111] More specifically, in H1 we argue that higher feedback frequencies increase role conflict. Surprisingly, we find that feedback frequency reduces role conflict (H1a, β_{LE} (low experience) = -.15, $p < .05$; H1b, β_{HE} (high experience) = -.23, $p < .01$). Next, in alignment with our second hypothesis, supervisor (customer) feedback frequency influences supervisor (customer) role ambiguity (H2a, β_{LE} = -.56, $p < .01$, and β_{HE} = -.48, $p < .01$; H2b, β_{LE} = -.46, $p < .01$, and β_{HE} = -.52, $p < .01$).

With respect to the moderating effects of supervisor and customer feedback frequency, the results reveal that supervisor feedback frequency reduces the relation

[111] The moderating effect of SE experience holds both for multigroup-analyses and when treated as a continuous moderator. Furthermore, we tested and found no evidence for nonlinear direct effects.

of customer feedback frequency and customer role ambiguity (H4b, $\beta_{LE} = -.13$, $p < .05$). However, neither role conflict nor supervisor role ambiguity are influenced by the interaction of both feedback frequencies (H3 and H4a).

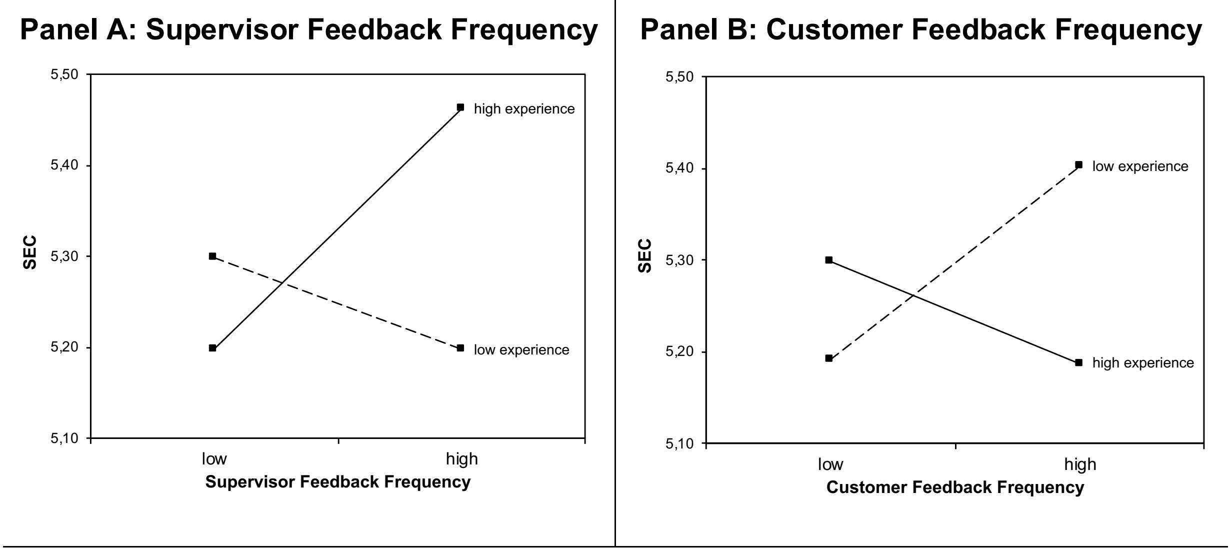
Furthermore, when examining the relation between role stressors and SEC, role conflict shows a consistently positive relation with SEC (H5, $\beta_{LE} = .32$, $p < .01$, and $\beta_{HE} = .24$, $p < .01$), whereas the impact of supervisor role ambiguity (H6a, $\beta_{HE} = -.27$, $p < .05$) and customer role ambiguity (H6b, $\beta_{LE} = -.21$, $p < .01$) is conditional on the SE's level of experience.

	Hypothesis	All sample		Low Experience		High Experience	
		pc	**t-value**	**pc**	**t-value**	**pc**	**t-value**
Estimated Paths							
SupFF → RC	H1a	-0.03	0.46	-0.15 *	2.02	0.19	1.83
CusFF → RC	H1b	-0.09	1.32	0.00	0.03	-0.23 **	2.69
SupFF → SupRA	H2a	-0.53 **	10.96	-0.56 **	10.05	-0.48 **	6.82
CusFF → CusRA	H2b	-0.48 **	10.75	-0.46 **	7.70	-0.52 **	7.34
CusFF X SupFF → RC	H3	0.01	0.18	0.03	0.52	-0.04	0.42
CusFF X SupFF → SupRA	H4a	0.00	0.06	-0.01	0.13	0.00	0.06
CusFF X SupFF → CusRA	H4b	-0.10 *	2.37	-0.13 *	2.47	-0.04	0.57
RC → SEC	H5	0.29 **	6.23	0.32 **	4.95	0.24 **	2.87
SupRA → SEC	H6a	-0.02	0.29	0.10	1.13	-0.27 *	2.18
CusRA → SEC	H6b	-0.11	1.51	-0.21 *	2.57	0.17	1.24
RC X SupFF → SEC	H7a	-0.16 *	2.54	-0.18 *	2.10	-0.18 *	2.28
RC X CusFF → SEC	H7b	0.00	0.07	0.07	0.96	-0.13	1.45
SupRA X SupFF → SEC	H8a	-0.12 *	2.26	-0.09	1.28	-0.18 *	2.40
CusRA X SupFF → SEC	H8b	0.05	0.84	0.00	0.03	0.14	1.55
SupRA X CusFF → SEC	H9a	0.13 *	2.30	0.14	1.88	0.10	0.85
CusRA X CusFF → SEC	H9b	-0.14 *	2.48	-0.15 *	2.07	-0.04	0.43
n		385		241		144	

Note: pc = path coefficient; SupFF = supervisor feedback frequency; CusFF = customer feedback frequency; SupRA = supervisor role ambiguity; CusRA = customer role ambiguity; RC = role conflict; SEC = service employee creativity; *p < .05; **p < .01 (two-tailed test), results of 5,000 bootstrapping samples.

Figure 3.5: Standardized Estimates of Inner Model

Finally, we examined the moderating effects of supervisor and customer feedback frequency, respectively, on the effects of role stressors for SEC. We found that supervisor feedback frequency reduces the impact of role conflict on SEC (H7a, β_{LE} = -.18, $p < .05$, and β_{HE} = -.18, $p < .05$), and that supervisor (customer) feedback frequency reduces the relation between supervisor (customer) role ambiguity and creativity, contingent upon SE experience (H8a, β_{HE} = -.18, $p < .05$; H9b, β_{LE} = -.15, $p < .05$). The inclusion of control variables does not alter the results. Following Aiken and West (1991) and Kenny (2013) we evaluated the effect sizes (f^2) of the hypothesized moderations, finding large effect sizes that indicate substantial moderation effects of the respective feedback frequency (see chapter 2.1.3.6).



Note: SE = service employee; SEC = service employee creativity; low (high) values of feedback frequency refer to 1SD below (above) their respective mean.

Figure 3.6: Moderating Effect of SE Experience on the Relationship between Feedback Frequency and SEC

Total effects show that both customer and supervisor feedback frequencies have a substantial effect on SEC, contingent upon the moderator SE experience (total effects supervisor feedback frequency: β_{LE} = -.10, $p < .05$ and β_{HE} = .18, $p < .01$; total effects customer feedback frequency: $\beta_{LE} = .10$, $p < .05$ and $\beta_{HE} = -.14$, $p < .05$; see Figure 3.6 Panel A and Panel B), and all effects are fully mediated through role stressors.

3.6 Discussion

Many firms respond to increasingly crisis-prone and rapidly changing markets by speeding up their feedback cycles to ensure that employees receive timely information to enable them to adjust their services. Services continuously satisfying changing customer demands is fundamental to firm success, and creativity is key to this (e.g., Agnihotri et al. 2014a). Yet the literature is silent on whether the frequency of feedback is an effective tool to manage service employee creativity (SEC). Therefore, our cross-sectional cross-industry study set out to explore how the frequency of feedback promotes SEC utilizing a dynamic S-O-R framework (Ilgen, Fisher, and Taylor 1979; Jacoby 2002) and role theory (Katz and Kahn 1978). We find that the frequency of feedback represents a powerful driver of SEC. Importantly, our results provide a more nuanced understanding of the long-held assumption that “the more frequent the feedback, the better” (Ilgen, Fisher, and Taylor 1979) by showing how feedback frequencies from customers and supervisors interact to influence the creativity of SEs in their boundary-spanning position. These results help us to make important contributions to service literature in several ways and provide strategies for decision makers responsible for managing service employee creativity.

3.6.1 Feedback Frequency as crucial Choice to manage Creativity

To start with, we contribute to service studies that assign “utmost importance” (Coelho and Augusto 2010, p. 427) to understanding SEC and its antecedents. Indeed, prior service research has contributed valuable insights on distinct drivers of service employee creativity, including the relevance of service employee's affective state (e.g., Bettencourt and Brown 2003; Lages and Piercy 2012), skills and abilities (e.g., Agnihotri et al. 2014a; Lages and Piercy 2012), relationships at work (Coelho, Augusto, and Lages 2011), and organizational factors such as organizational support (Lages and Piercy 2012), job complexity (e.g., Coelho, Augusto, and Lages 2011), and supervisor feedback (e.g., Agnihotri et al. 2014a; Coelho and Augusto 2010). Our study infuses this strand of literature with a temporal dimension by demonstrating the relevance of the frequency of feedback as an effective lever to drive SEC. This is particularly notable given market environments defined by ever-faster change and literature that increasingly focuses on better understanding how firms can cope with higher volatilities through agility, responsiveness, and speed of change (Singh et al. 2013).

Furthermore, to reflect the boundary-spanning role of SEs, we are the first to combine both supervisor and customer as feedback sources that impact on SE behavior. Our results support this choice as we find substantial effects of customer feedback frequency for SEC, indicating that previous findings restricted to supervisor feedback might be distorted. However, we also uncover the potential of reducing SEC through providing more frequent customer feedback. This adds more color to unambiguously positive arguments suggesting that "codesign fosters creative thinking" (Trischler et al. 2018, p. 89), such as that present in the literature on customer involvement, customer participation, and co-creation (Chang and Taylor 2016; Storey and Larbig 2018) and echoes recent indications of more nuanced and cautious customer engagements in the service process (e.g., Brodie et al. 2011; Cabiddu, Frau, and Lombardo 2019). Our study uncovers the risks involved when translating general conceptual considerations which argue that "the customer is always a co-creator of value" (FP6, Vargo and Lusch 2008a, p. 7) into allowing customers to provide frequent feedback to every SE.[112] This provides an interesting insight into the times at which customers need to be mindful of not overtaxing their relevance for service employees. It could be argued that other ways exist for the customer to be involved in co-creating services over and above providing feedback, yet it is difficult to imagine a customer as a co-creator of value (FP6) who "identifies the logic of interactivity" (Vargo and Lusch 2008b, p. 32) without providing feedback. It is more reasonable to assume that feedback represents a dominant means to realize customer co-creation. Therefore, we would liken the negative effect of more frequent customer feedback for some SEs to what Cabiddu, Frau, and Lombardo (2019) have termed value co-destruction (VCD). Because our conceptual model singles out creativity as one element of the co-creation process, our results allow a clearly defined addition to previously documented VCD-social interactions (Cabiddu, Frau, and Lombardo 2019).

3.6.2 Interplay of Supervisor and Customer Feedback

Furthermore, pertinent questions with respect to how customer feedback interacts with supervisor feedback have been left unanswered. Extending conceptual considerations (Agnihotri et al. 2014a; Challagalla and Shervani 1996; Wilder, Collier, and Barnes 2014), our analysis provides first empirical indications for the ability of supervisors to explain customer communication to SEs. Leveraging their strategic overview, expertise, and market knowledge (Adler and Borys 1996), supervisors can contextualize customer feedback. Our results, therefore, support recent publications to position leader influence "as a facilitator of the customer–employee interactions" (Dong et al. 2015, p. 1373).

[112] While within the conceptual framework of service-dominant logic it could be argued that service employee creativity does not necessarily equate to the final service in use and that therefore our model escapes any value co-creation/co-destruction considerations, it seems undisputed that creativity is an indispensable and positive driving element of the service generation process.

Interestingly, we do not find any influence of customer feedback frequency on the effectiveness of supervisor feedback frequency. This exposes any concerns of supervisors who perceive customer feedback as an interference with their relation to the SE in the “three-cornered fight” among SEs, customers, and supervisors (Bateson 1985) as unjustified.

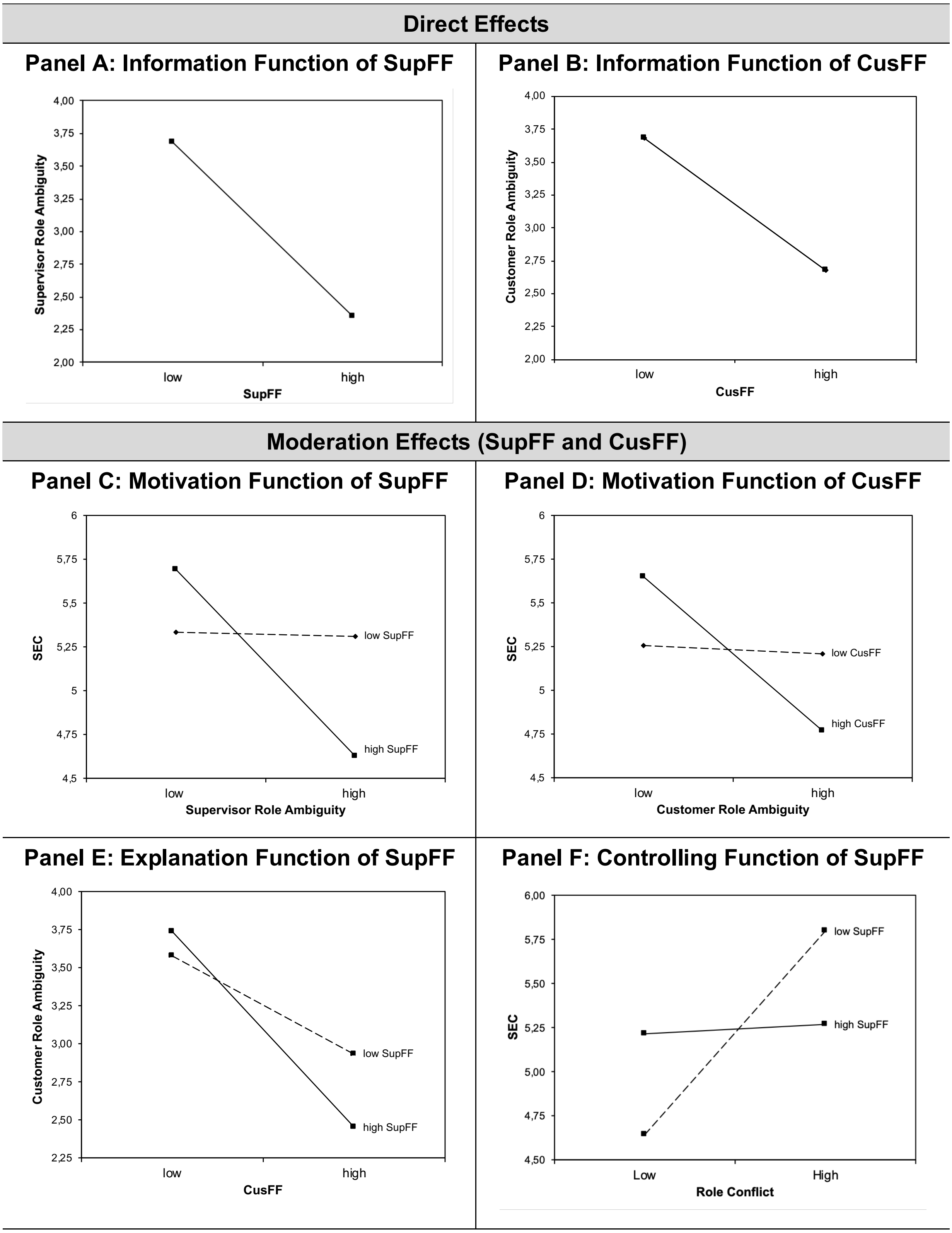
Our findings surprisingly also contest the commonly held assumption that multiple feedback sources interfere to cause inter-sender role conflicts. A detailed analysis revealed that SEs perceive neither the same levels of respective customer and supervisor feedback frequency nor feedback frequency dominance (van der Borgh, de Jong, and Nijssen 2019) as a cause of role conflict (untabulated). One suggestion to explain these findings comes from boundary-spanner research, which shows that boundary spanners develop specific skills to accept diverse expectations from different sources as a normality inherent to their job (Malhotra and Ackfeldt 2016). In a grander scheme, our results contribute to a more comprehensive understanding of SEs’ boundary-spanning role (Agnihotri et al. 2014a; Yoo and Arnold 2016).

3.6.3 Opening the Black Box: Role Theory-derived Mediators

Next, we unlock the black box that links frequent feedback stimuli with behavioral responses. Our focus on the internal state of SEs enables a more nuanced understanding of previous findings regarding the implications of role conflict on SEC (e.g., Bettencourt and Brown 2003; Coelho, Augusto, and Lages 2011; Schepers, Nijssen, and van der Heijden 2016), as we show that this positive relation is conditional on the level of supervisor feedback frequency and might turn into a negative relation at high levels of supervisor feedback frequency (see Figure 3.5). Our evidence thereby substantiates theoretical considerations that higher supervisor feedback frequency puts employees in a behavioral corset and thereby hinders creativity as “feedback is all too frequently controlling; it emphasizes how people should behave and implies that the manager is in control" (Deci, Connell, and Ryan 1989, p. 585). With these results, we complement research in leader behaviors and employee creativity (cf. Amabile et al. 2004).

By our selection of SE's internal reactions to feedback frequency, we can also provide a more differentiated picture regarding the effectiveness of role ambiguity on SEC. Whereas previous research (e.g., Bettencourt and Brown 2003; Coelho, Augusto, and Lages 2011) consistently acknowledges the negative influence of role ambiguity on the SE's search for novel ideas, our results show that this negative relationship is magnified by increasing levels of feedback frequency. Drawing on notions of the ability of feedback to foster intrinsic motivation (e.g., Coelho and Augusto 2010), we conclude that if SEs are not motivated to generate novel ideas by sufficiently frequent feedback, they will not engage in creative behaviors regardless of the level of any of their role ambiguity facets (cf. Figure 3.7, Panels C and D).

Our moderated mediation research design enables us to single out a number of distinct functions of feedback (cf. Ilgen, Fisher, and Taylor 1979, p. 351). Consistent with the multiple ways in which feedback impacts as described by the broader feedback literature (cf. Agnihotri et al. 2014a; Challagalla and Shervani 1996; Coelho and Augusto 2010; Deci, Connell, and Ryan 1989; Ilgen, Fisher, and Taylor 1979; Mullins, Agnihotri, and Hall, 2020), we find indications of the information (H2a, H2b, Figure 3.5), motivation (H8a, H8b, Figure 3.5), controlling (H7a, Figure 3.5), and explanation (H4b, Figure 3.5) functions of feedback for SEs (cf. Figure 3.7). Interestingly, while information and motivation functions apply to both customer and supervisor feedback frequency, controlling and explanation functions are only present if the supervisor is the feedback source.



Note: SupFF = supervisor feedback frequency; CusFF = customer feedback frequency; SEC = service employee creativity; low (high) values of the moderator and the predictor variable refer to 1SD below (above) their respective mean; Panels C and D base on subsamples (n = 144 and n = 241 respectively), Panel A, B, E, and F base on all sample (n = 385).

Figure 3.7: Functions of Customer and Supervisor Feedback Frequencies in the Context of SEC

3.6.4 Importance of Recipient Characteristics

Finally, we find that SE experience moderates feedback frequency effectiveness. Previous studies have tested multi-factor models and outlined that employee creativity depends on individual domain- and creativity-relevant skills and abilities (e.g., Agnihotri et al. 2014a; Amabile et al. 2004), and human capital theory (Becker 1964) has taught us that these change with growing experience so that more experienced employees command a different set of skills and abilities. In congruence with these studies, our results show that employee experience is indeed an important factor that changes the reaction of SEs to feedback stimuli. This is also interesting for the general feedback literature as it paves the way to fuse two opposing theoretical camps that speak for both a positive (e.g., Ilgen, Fisher, and Taylor 1979; Holderness, Olsen, and Thornock 2020) and a negative (e.g., Casas-Arce, Lourenco, and Martínez-Jerez 2017; Deci, Connell, and Ryan 1989) relationship between feedback frequency and employee performance, respectively. In other words, the relevance of the frequency of feedback as a driver of SEC does not diminish with more experience as key conceptual feedback studies indicate (Anseel et al. 2015; Ilgen, Fisher, and Taylor 1979); rather, more experienced employees approach creative activities in different ways.

3.7 Managerial Implications

From a managerial perspective, this study provides a plethora of practical propositions that advocate the positive potential inherent to the frequency of feedback as an effective management lever for elevating SEC. This is important to recognize for any manager, as it is an inevitable decision regarding how often feedback should be provided. In practice, realizing the time-pressured work contexts that many supervisors live in, supervisors may hesitate to provide feedback more often because of resource restrictions (e.g., time, cognitive capability), particularly when responsible for multiple employees. Quite attractively, our results draw an escape strategy for such a management dilemma, as they guide supervisors to concentrate feedback capacities towards selected SEs, thereby optimizing the aggregate levels of employee creativity. This is a distinguishing feature of our study, as some academic studies inherently imply the availability of free resources to implement their managerial implications.

More specifically, firstly, supervisors can directly control SEC. To do so effectively, supervisors need to be aware of their SE's experience level. For less experienced SEs, it is vital to recognize that supervisor feedback frequency can reduce creativity (cf. Figure 3.6, Panel A). For more experienced SEs, however, supervisors command a powerful design element of their control package, as more frequent feedback leads to more creative ideas. It should be noted that supervisors need to brace for a shift in SE reactions, however, as few feedback contacts primarily provoke employee role conflict to stimulate SEC, while more frequent feedback contacts stimulate SEC through conveying a clearer picture of SE's job expectations. Our results also motivate supervisors to go the extra mile, as every additional feedback contact is two times more effective compared to less frequent feedback.

Secondly, our suggestions extend towards customers, as we find that customer feedback activities only pay for less experienced SEs (cf. Figure 3.6, Panel B). Supervisors should therefore motivate these SEs to get into contact with their customers, and vice versa, through breaking down organizational and technological barriers that restrict customer exchange (Giebelhausen et al. 2014), by cultivating a feedback-seeking climate (Auh et al. 2019) towards customers, and by ensuring sufficient resources, education, and information about how to enquire for feedback in the most appropriate manner (Auh et al. 2019). However, our results also reveal a potentially fatal trap: If customers start with infrequent feedback contacts to establish a relationship, the SE’s response indicates that the SE is not receptive to customer feedback. Only those customers that persist and counter-intuitively provide higher levels of feedback frequency are rewarded with a substantial creative response (cf. Figure 3.6, Panel B).

Thirdly, our implications for how to best exploit SEs’ creative potential also extend to firm strategies as firms could adapt their hiring, training, and career strategies utilizing our findings. When hiring employees, firms can single out employees who are sensitive to different levels of feedback frequency (e.g., narcissistic employees that “respond better to more frequent feedback”, Young et al. 2016, p. 50), to strategically elevate the firm’s creative potential. Moreover, training and career planning could (temporarily) allocate communicative and open-minded personnel to the organizational units which are in most need of creative potential.

Thus far, we had the implicit assumption that feedback frequency can be changed. In some situations, however, this might not be possible – for instance due to fixed feedback routines, inherent job characteristics or a predetermined dominance of either customers or supervisors as a source of feedback. Our results nevertheless provide managers with valid implications for these situations as they suggest to strategically select candidates with accordingly low or high levels of experience.

On a meta-level, firms should establish structures that develop supervisors’ communicativeness, openness to requests, and cybernetic leadership style, but which also curfew supervisors’ misuse of hierarchical power in the “three-cornered fight” (Bateson 1985).

3.8 Limitations and Future Research

Our study has limitations which in turn may open up avenues for further research. To begin, some methodological issues that apply to all survey-based field studies are present. First, our data is sampled across multiple industries and organizations. Although this contributes to the generalizability of our findings, we were limited in gathering information per respondent to disentangle customer relationship characteristics. Future studies are encouraged to specify more customer characteristics that impact the relationship between customers and SEs.

Second, in accordance with previous studies, we measured creativity using a self-report measure (Agnihotri et al. 2014a; Coelho, Augusto, and Lages 2011; Lages and Piercy 2012). However, not all ideas are necessarily novel or feasible to the organization or customer, and they thus fail to display starting points for service innovations, which could be picked up by future research attached to research focusing on high-creativity ideas (e.g., Kachelmeier, Wang, and Williamson 2019). Still, with respect to better understanding how to stimulate creativity, it can be argued to be less important whether feedback frequency induces an objectively novel idea and more important whether the SE invests cognitive and emotional resources and is enabled to think outside his or her box. Future studies are thus encouraged to take a more comprehensive perspective on the interaction of feedback frequency and the innovation process.

Furthermore, we focused on two important feedback sources close to the SE, which implies that we do not know how the feedback dynamics would develop in the presence of additional feedback sources, for example by recognizing different types of customers or internal counterparts. Along these lines, a promising research impulse hails from recent studies in information technology journals (e.g., Krancher, Luther, and Jost 2018) which focus on automated feedback sources to reflect the ongoing digitization revolution. Similarly, Ostrom et al. (2015, p. 131) state that “[m]any of the roundtables perceived technology, particularly digital technology, as one of the key opportunities and challenges related to service innovation”. Automated feedback sources might rival our focus on human feedback sources as their speed and frequency, ubiquitous availability, but also inherent inability to build inter-human trust differentiate human and machine feedback sources. Therefore, juxtaposing machine feedback and human feedback regarding the “last human stronghold” (Beniaminy 2020) of creativity could be a promising addition to our understanding of the important relationship between feedback frequency and SEC.

4. Study 2: Gender of Feedback Source and Quality of Creativity

4.1 Introduction

In Study 1, the S-O-R framework and role theory have been drawn upon to show that feedback frequency affects employee creativity and that this effect is mediated by role stress (i.e., role conflict and role ambiguity). Also, the creative response to feedback frequency depends on the feedback source (i.e., customer or supervisor) and on recipient characteristics (i.e., employee experience).

Study 2 aims at observing the influence of gender on the creative effectiveness of feedback frequency and distinguishing the implications of feedback frequency on the quantity of creativity and on the quality of creativity by applying feedback theory (Ilgen, Fisher, and Taylor 1979) and gender role theory (Eagly and Karau 1991; Lanaj and Hollenbeck 2015). Thereby, Study 2 complements the findings of Study 1 by, first, providing evidence for the relationship between feedback frequency and employee creativity using a different empirical methodology (i.e., a field experiment). The (field) experimental approach overcomes the weaknesses inherent in the survey-based research approach in Study 1 by demonstrating the presumed direction of the cause-effect relationship. Second, utilizing feedback theory (Ilgen, Fisher, and Taylor 1979), this research approach allows me to disentangle the effects of feedback frequency from other feedback dimensions that are often related to it (e.g., feedback valence and feedback timing). Thus, Study 2 is able to rule out alternative explanations and to establish an interference-free relationship between feedback frequency and employee creativity. Third, the joint application of feedback theory (Ilgen, Fisher, and Taylor 1979) and gender role theory (Eagly and Karau 1991; Lanaj and Hollenbeck 2015) in Study 2 reveals that gender of the feedback source moderates the creative effectiveness of feedback frequency, while it also shows that this effectiveness is not influenced by gender of the feedback recipient. These results add to the moderating effects of recipient characteristics (i.e., employee experience) and feedback source (i.e., customer and supervisor) identified in Study 1. Fourth, while a particular limitation of Study 1 is that it does not differentiate between the potential of creative ideas to represent starting points for innovation (see chapter 3.8), Study 2 builds on the work of Kachelmeier and colleagues (2008, 2019) and finds that feedback frequency causes employees to generate ideas that have a high quality and hence offer the greatest potential for innovations and organizational development. Finally, considering arguments from feedback theory (Ilgen, Fisher, and Taylor 1979), Study 2 provides further insights into the mechanisms behind the creative benefits associated with higher levels of feedback frequency. Thus, Study 2 addresses gaps in the feedback and management accounting literature, but also contributes to gender and leadership research.

The frequency of feedback constitutes a crucial component of the feedback stimulus (Ilgen, Fisher, and Taylor 1979). While feedback theory (Ilgen, Fisher, and Taylor 1979; Taylor, Fisher, and Ilgen 1984) clearly pursues a distinct conceptualization of feedback dimensions (e.g., feedback frequency, feedback quantity, or feedback timing), empirical studies not only show that certain feedback dimensions are highly correlated (Kinicki et al. 2004; Larson 1986), but some empirical operationalization of feedback frequency even overlap with other dimensions. For instance, Hackman and Oldham's (1975) commonly used measure of feedback (quantity) derived from their job characteristics model, is biased toward feedback frequency (Fedor and Buckley 1987), whereas in experimental studies not only feedback frequency but simultaneously feedback quantity has been manipulated (e.g., Holderness, Olsen, and Thornock 2020 or Lam et al. 2011). Similarly, investigations of feedback timing have been accused of actually observing feedback frequency effects (e.g., Goomas, Smith, and Ludwig 2011 in Bechtel et al. 2015). Furthermore, the most frequently studied source of feedback, the supervisor (Alvero, Bucklin, and Austin 2001), perceives providing negative feedback as an unpleasant and eluded task (Larson 1986), possibly causing a confounding effect of feedback frequency with feedback valence (Ilgen, Fisher, and Taylor 1979).[113]

While these feedback dimensions related to feedback frequency have been shown to impact on employee creativity (e.g., Coelho and Augusto 2010; Kim and Kim 2020; Zhou 1998) or there is a clear logic behind their effects, for example, feedback timing determines the time in which creative ideas can emerge, referred to as creative incubation period (Dodds, Ward, and Smith 2012; Gilhooly 2016), little is known whether feedback frequency (i.e., how often feedback is provided) alone will affect employee creativity. Given that prior research has not neatly delineated the effects of feedback frequency and its related feedback dimensions, a fresh reexamination is warranted. Thus, one goal of this field experimental study is to decouple the effects of feedback frequency on recipient (i.e., employee) creativity from feedback quantity, feedback valence, and feedback timing, and to demonstrate an interference-free cause-effect relationship.

As the generation of creative ideas is a non-standardized individual process, personal characteristics and "the interaction between personal characteristics and the work environment" (Shalley and Gilson 2004, p. 34) play a major role in this intrapersonal process (Shalley, Zhou, and Oldham 2004; Shalley and Gilson 2004). Women and men exhibit systematically different characteristics in work contexts, particularly when it comes to non-standard activities such as creativity or interpersonal

[113] Following this line of argumentation, feedback frequency effects may also be attributed to recipient capability effects, since feedback valence arises from prior performance.

communication. In organizational psychology and gender research, it is well established that women and men differ in their perceptions of the behaviors of others, due to the prescriptive nature of gender stereotypes projected onto female or male counterparts, as well as inherent traits specific to men and women (Heilmann and Chen 2005; Carli 2010). This particularly applies to feedback contexts where gender influences both the perception of the (female or male) source's behavior and the (female or male) recipient's response due to the perceptual bias and the prescriptive nature of stereotypical gender roles (Eagly and Karau 1991; Eagly and Wood 2012; Lanaj and Hollenbeck 2015). The importance of gender in work context is reinforced when one considers that gender roles are more salient than organizational roles, such as supervisor or subordinate (Eagly and Karau 1991). Moreover, feedback theory (Ilgen, Fisher, and Taylor 1979; Taylor, Fisher and Ilgen 1984) emphasizes the potential of interaction effects between the feedback stimulus with the characteristics of the feedback source and the feedback recipient, such as gender, that ultimately determine the recipient's response. Nevertheless, the effects of gender are not trivial, as three systematic influences of gender on the relationship between feedback frequency and employee creativity emerge from the interpersonal dyad of source and recipient:

First, being stereotypically associated with communal behavior, women's actions are perceived, ceteris paribus, as more helpful and prosocial (Eagly and Karau 1991). Therefore, feedback from a female feedback source should increase the intrinsic motivation of recipients and create an environment where "trial and error" is perceived to be less critically judged, both of which should enhance employee engagement in the search for novel ideas (Shalley and Gilson 2004). In contrast, men tend to be seen as natural leaders, competent and objective (Carli 2010; Eagly and Karau 1991). According to feedback theory (Ilgen, Fisher, and Taylor 1979; Taylor, Fisher, and Ilgen 1984), these attributions should elicit the (creative) responses of the recipients, and the effect of this reinforcement should correspond to the frequency with which the feedback is provided. Consequently, there is ambiguity about the influence of gender of the feedback source on the effectiveness of feedback frequency.

Second, extant gender research has emphasized that “feedback has different effects for females and males” (Wozniak 2012, p. 170). Gender stereotypes are prescriptive of female and male behavior (Heilman and Chen 2005; Eagly and Wood 2012). While the male stereotype requires men to be self-reliant and dominant, women are expected to be modest and build their self-concept more on relational aspects (Burton and Hoobler 2006; Carli 2010; Miller, Perlman, and Brehm 2007). Thus, women desire more frequent communication and are more receptive to feedback information (Burton and Hoobler 2006; Sturm et al. 2014; Röhner and Schütz 2020). In other words, women have a more positive attitude toward additional feedback stimuli and render provided feedback more meaningful and stimulating, potentially influencing the impact of feedback frequency on the creative responses of (female or male) feedback recipients.

Third, gender (in)congruence reflects a degree of congruence between feedback source and feedback recipient on belonging to the same social groups (Biernat and Danaher 2012; Sturm et al. 2014). Since the psychological and feedback literature has stressed the importance of source-recipient similarity for feedback effectiveness (e.g., Biernat and Danaher 2012; Carli 2010; Sturm et al. 2014), I examine whether gender (in)congruence moderates the relationship between feedback frequency and employee creativity, and thereby respond to recent calls that “research assessing the translation of feedback from both ingroup and outgroup members [(i.e., gender congruent and gender incongruent dyads, respectively,)] is needed” (Biernat and Danaher 2012, p. 276).

Given the large body of gender literature on feedback (e.g., Baier et al. 2018; Biernat et al. 2020; Biernat and Danaher 2012; Burton and Hoobler 2006; Ertac and Szentes 2011; Fletcher 1999; Gill and Prowse 2014; Park et al. 2018; Sturm et al. 2014; Taylor and Hood 2011; Vescio et al. 2005; Wozniak, Harbaugh, and Mayr 2016), it is surprising that extant literature on feedback as driver of creativity has not previously picked up on the tension inherent in gender, namely the feedback source's gender, feedback recipient's gender, and gender (in)congruence.

Thus, with this study, I address these research gaps in the feedback and management accounting literature and contribute also to gender and leadership research. First, contributions are made to the extensive feedback literature by examining the implications of feedback frequency for employee creativity. The results reveal that the frequency of feedback yields a higher quality of creative ideas from feedback recipients, while it does not affect the number of ideas generated. The field experimental research design allows me not only to substantiate the cause-effect relationship between feedback frequency and creativity, but also to rule out the effects of several other feedback dimensions that might interfere with this relationship, such as feedback valence, feedback timing, and feedback quantity (Ilgen, Fisher, and Taylor 1979; Lechermeier and Fassnacht 2018; Thornock 2016; Zhou 1998). Since these dimensions are often confounded with feedback frequency in practice (and in some of previous research), the results provide more certainty about the true impact of feedback frequency on employee responses.

Furthermore, I complement recent management accounting research on antecedents of employees' high-creativity idea generation, such as the type of compensation (Kachelmeier, Reichert, and Williamson 2008; Kachelmeier, Wang, and Williamson 2019), by adding feedback frequency to the short list of known determinants. Due to the specific feedback design, recent calls that "while technological advancements have made it easier for companies to provide frequent relative performance information (RPI) feedback, research on the effectiveness of such feedback is limited" (Holderness, Olsen, and Thornock 2020, p. 156) are also addressed.

Second, the results show that gender is a critical characteristic of the feedback source for the effectiveness of feedback frequency. I find feedback from a female feedback source to be more powerful, and surprisingly so regardless of the recipient's gender, which contradicts the feedback theory suggestions but is consistent with the arguments of gender role theory (Eagly and Karau 1991; Eagly and Wood 2012; Lanaj and Hollenbeck 2015). While feedback research has repeatedly emphasized the importance of the feedback source characteristics (e.g., Fedor et al. 2001; Holderness, Olsen, and Thornock 2017; Steelman and Rutkowski 2004), gender has received scant attention so far (see the reviews by Alvero, Bucklin, and Austin 2001 and Lechermeier and Fassnacht 2018, which systematically analyze feedback source and source characteristics). Contrary to gender role theory, I unexpectedly find that recipient gender and gender (in)congruence are irrelevant to the relationship between feedback frequency and recipient creativity.

Third, the supplemental analysis provides insights into the mechanisms by which feedback frequency operates. The results show that recipients perceive more frequent feedback as additional effort from their supervisors and as a greater disruption to their daily routine, suggesting that frequent feedback induces motivation through reciprocity (Fedor and Buckley 1987) and, consistent with theoretical considerations (Ilgen, Fisher, and Taylor 1979; Kluger and DeNisi 1996), that a frequent feedback stimulus is more salient and hence less likely to be overshadowed by other stimuli in the recipient's environment. Surprisingly, I also find that high levels of feedback frequency help employees process the feedback in units they can digest better. This result contrasts with previous lines of reasoning that claim frequent feedback causes information overload, e.g., Lam et al. (2011) and Holderness, Olsen, and Thornock (2020), who, however, do not control for the quantity of feedback.

4.2 Theory and Hypotheses Development

Feedback theory is utilized to substantiate the conceptual model (Ilgen, Fisher, and Taylor 1979). Feedback is an important stimulus to affect recipient's behavior since "as feedback provides information […], it stimulates the employee to explore different courses of action in the pursuit of a better result" (Coelho and Augusto 2010, p. 429). The recipient's processing of the feedback stimulus, that ultimately determines his or her response, such as the generation of novel ideas, is an imperceptible intrapersonal process that is not only affected by the nature of the feedback stimulus (e.g., its frequency), but also depends on the characteristics of the feedback source and recipient (e.g., gender) (Ilgen, Fisher, and Taylor 1979). These characteristics influence the likelihood that individuals will accept and favorably respond to feedback (Ilgen, Fisher, and Taylor 1979; Kluger and DeNisi 1996; Taylor, Fisher and Ilgen 1984).

In their seminal work Ilgen, Fisher, and Taylor (1979) posit that feedback frequency influences the feedback perception and recipient's desire to respond to the feedback, which ultimately determine the recipient's (creative) response. With regard to feedback perception being the initial step in the recipient's feedback processing, more frequent feedback allows recipients to better refer the feedback to specific (since more recent) behaviors, which increases the perceived accuracy of the feedback (Ilgen, Fisher, and Taylor 1979). Similarly, more frequent feedback enables recipients to make better comparisons with past performance, so recipients gain a better understanding of how their behavior relates to specific outcomes, potentially increasing creative self-efficacy and intrinsic motivation (Tierney and Farmer 2002).

The frequency of the feedback also influences the perception of the work environment in which recipients decide whether to engage in creative activities. According to feedback theory (Ilgen, Fisher, and Taylor 1979; Taylor, Fisher, and Ilgen 1984), which originated in cybernetics, feedback recipients strive to avoid negative feedback. Engaging in creative activities, however, increases employees' risk of not performing as required because the pursuit of creative avenues regularly involves trying new behaviors that may (initially) fall short of the performance of established and well-known behaviors or require resources (e.g., time, budget) that could be invested for (non-creative) performance purposes. As a result, some contemporary management approaches (e.g., Sagiv et al. 2010; Whitney 2018) rely on early recommendations (Guilford 1950; Wallas 1926) urging that "when a creative idea is generated, total freedom must be ensured by eliminating directional guidance, constraints, criticism, and thinking within bounded scopes" (Sagiv et al. 2010, p. 1088). In particular, frequent feedback not only underscores the perceived importance of performance evaluations by requiring costly organizational resources (Uhl-Bien and Arena 2018), but simultaneously shortens the time between two evaluations, making employees more likely to be driven to meet short-term goals that allow less to compensate when a creative trial has yielded inferior performance. This rationale aligns with Nohria and Gulati's (1996) line of reasoning about the negative consequences of too little budgetary slack for experimentations impeding innovations.

On the contrary, considering that feedback is a vehicle for employees to obtain important information that has the potential to be leveraged to improve their performance and hence experience self-efficacy, receiving feedback can also be perceived as access to valuable organizational resources that benefit the feedback recipient (Fedor and Buckley 1987). Thus, frequent feedback can be understood as a signal of caring and support from supervisors. More specifically, supervisors are typically more experienced than their subordinates, have access to other information sources within the organization, and, because of their strategic overview, can point out how the subordinate’s behavior relates to other organizational processes (Adler and Borys 1996) – all of which can benefit subordinates and provide a complementary perspective that enhances their understanding of their work duties. In addition, frequent interactions between supervisor and subordinate strengthen their relationship and foster interpersonal trust (Dagger, Danaher, and Gibbs 2009; Doney and Cannon 1997), which provides fertile ground for creativity (Shalley and Gilson 2004; Hughes et al. 2018).

Feedback recipients’ desire to respond to the feedback provided subsumes the long-recognized motivational effects of feedback (Ilgen, Fisher, and Taylor 1979): “Performance feedback performs a motivational function by providing the proper environment for the recipient to meet higher order needs through task accomplishment” (Ilgen, Fisher, and Taylor 1979, p. 350; Deci 1975; Hackman and Oldham 1975). Given the described benefits in feedback perception that result from feedback frequency, such as a higher accuracy and a higher self-efficacy, feedback frequency should have a positive impact on creativity, as “proper feedback […] motivates employees to be more creative” (Agnihotri et al. 2014a, p. 171). Moreover, since adequate feedback enables recipients to “obtain a better understanding of their jobs, and this fosters their domain-relevant skills” (Coelho and Augusto 2010), “the frequency of feedback should enhance motivation on the task to the extent that it increases the recipient's perceptions of competence.” (Ilgen, Fisher, and Taylor 1979, p. 363).

A further motivational facet of feedback frequency arises from the conceptualization of feedback as organizational resource (Ashford and Cummings 1985). Since feedback is “perceived as a limited resource that has value to potential organizational recipients” (Fedor and Buckley 1987, p. 173) and is provided by a member of the organization, e.g., supervisor, frequent feedback addresses reciprocity mechanisms (Gouldner 1960) that motivate the recipients to engage in activities that benefit the organization, such as raising their (creative) performance.

4.2.1 Feedback Frequency and the Quantity and Quality of Creativity

In this study, I draw on the conceptualization of creativity by Kachelmeier and colleagues (2008, 2019), who distinguish between the quantity of creativity (the number of ideas) and the quality of creativity (a subset of all ideas that are classified as high-creativity ideas). The quantity of creativity determines a measure of creative productivity, which is particularly important for firms that rely on the generation of creative ideas to succeed, as these firms “must generate creative innovations while also maintaining high ongoing productivity [and] want their employees to be both productive (i.e., high volume) and creative” (Kachelmeier, Reichert, and Williamson 2008, p. 342). Yet, the generation of arbitrary creativity can become ineffective because only a fraction of creative ideas (i.e., the ‘best’ ideas) are converted into innovations due to capability constraints. Consequently, creativity (i.e., the production of ideas) is “a necessary, but insufficient condition for innovation” (i.e., the implementation of ideas) (Rietzschel, Nijstad, and Stroebe 2010, p. 48; Amabile 1988; West 2002). Not surprisingly, a strand of creativity literature is devoted to the study of the process of selecting the ideas of highest quality from the sea of ideas generated (e.g., Faure 2004; Putman and Paulus 2009; Rietzschel, Nijstad, and Stroebe 2006, 2010). Therefore, the quality of creative ideas, the assessment of whether the ideas generated are genuinely more than just bizarre and erratic, but are truly original, implementable, and have the potential to lead to real advancements for the business, is a distinct and critical dimension in the evaluation of creativity.

Importantly, the nature of feedback frequency offers potential to affect both creativity dimensions. In general, more frequent feedback offers more starting points for novel ideas (quantity) as feedback can reveal challenges that require creative solutions and make the recipient apparent and conscious about them. For instance, negative feedback about long on-hold times when customers try to reach the company's call center may prompt a call center employee to suggest that customers schedule appointments, allowing the call center to optimize their capacity utilization, reduce typical peak times, and minimize wait times for calling customers. Without the feedback, the employee would not be aware of the necessity to take action and start looking for new ideas to solve this issue, but instead would continue to act according to established and scripted behaviors. In other words, "negative feedback alerts recipients to a creativity-standard gap, and thus may offer an opportunity to improve creativity" (Kim and Kim 2020, p. 584). Hence, the more frequent the feedback, the more opportunities exist for challenges to come at light, such that the more often information are presented should correspond to the quantity of creative ideas, or as Ilgen, Fisher, and Taylor (1979, p. 360) put it, "the frequency of feedback should be positively correlated with the frequency of correct responses". This reasoning about the effectiveness of feedback frequency goes beyond feedback quantity, because in the example above, if feedback was too infrequent, the challenge associated with customer wait times would have become obsolete in the interim to the next feedback, as customers became dissatisfied with the services and abandoned the company, automatically reducing wait times – to the company's chagrin.

Complementarily, feedback frequency may also affect the quality of recipients' creativity. Feedback divides the idea generation process into smaller time segments[114] characterized by different experiences and knowledge of the recipient, as feedback provides the recipient with an evaluation of current efforts and updated information (Coelho and Augusto 2010). In this way, feedback recipients are better able to focus their efforts on improving promising ideas, abandoning former ideas of inferior quality and searching more effectively for qualitatively superior ideas during the ideation process. However, it has been shown that frequent feedback can also cause recipients to lose a holistic picture and overemphasize single (recent) outliers in the evaluation (Lurie and Swaminathan 2009), which is especially more likely with smaller periods evaluated, that eradicate the potential of detecting gradual changes in performance and hence particularly harm the quality of creativity.

Moreover, previous research has emphasized the consequences of different incentives (i.e., different motivators) on the creation of high-quality ideas (Kachelmeier, Reichert, and Williamson 2008; Kachelmeier, Wang, and Williamson 2019). Since "a plausible role for incentives is to motivate people to think differently" (Kachelmeier, Wang, and Williamson 2019, p. 252), I also assume that due to the motivational effects of feedback frequency, highly frequent feedback particularly motivates recipients to go the extra mile and generate highly creative ideas.

Nonetheless, processing feedback is more effortful when it is frequent and therefore diverts cognitive resources away from other work activities (Lam et al. 2011;

[114] I generally assume that one or more feedbacks are given while employees are working on generating creative ideas, which corresponds to a dynamic view of the feedback provision process. Nevertheless, my arguments hold even when this assumption is violated, as more frequent feedback increases the likelihood of receiving feedback during the idea generation process.

Holderness, Olsen, and Thornock 2020). Creative activities in particular should be prone to receive fewer resources considering that they are cognitively intense (Lechermeier and Fassnacht 2018) and cutting them off releases plenty of cognitive resources. While in principle this affects both the quantity and the quality of creativity, I assume that this crowding out of cognitive resources is particularly relevant for the generation of higher quality creative ideas, since refraining from seeking such ideas should free up relative more resources. Besides, feedback frequency, ceteris paribus, interferes with the creative incubation period – a period in which the employee is left alone – and thus potentially harms the generation of ideas with high quality. This leads me to hypothesize:

Hypothesis 1: More frequent feedback causes a higher quantity (a) and higher quality (b) of creative ideas

4.2.2 Interaction: Feedback Frequency, Gender, and Creativity

To better understand how feedback frequency is processed and transformed into a behavioral response (Ilgen, Fisher, and Taylor 1979) such as creativity, I recognize gender as distinctive characteristic of the feedback source and the recipient and hence as a potential moderator of the effectiveness of feedback frequency since "social stereotypes are indeed powerful environmental stimuli that do not depend on conscious, personal endorsement for their effects to be palpable" (Jost and Kay 2005, p. 498). My conceptual considerations are infused by gender role theory (Eagly and Karau 1991; Eagly and Wood 2012; Lanaj and Hollenbeck 2015) to explain why the gender of the feedback recipient, the gender of the feedback source, and gender (in)congruence influence the effects of feedback frequency on employee creativity. This study refers to "gender" as the "psychosocial implications of being male or female"

(Powell et al. 2019, p. 62), which is distinct to “sex” describing a physiological distinction of female and male (Archer and Lloyd 2002; Borna and White 2003; Powell et al. 2019; Unger 1979), and therefore follow “accepted definitions of these terms in the social science literature” (Borna and White 2003, p. 90). Hence, “the “sex” of a person is biologically determined, the “gender” of a person is culturally and socially constructed” (Borna and White 2003, p. 90).

Gender role theory postulates the existence of “shared social role expectations of how women and men ought to behave that affects people’s perceptions in the workplace” (Lanaj and Hollenbeck 2015, p. 1478). These expectations “arise from the distribution of women and men into different specific social roles in natural settings” (Eagly and Karau 1991, p. 686). Thus, the behavior of other women (men) shapes the expectation each individual faces and the way others perceive her (or his) behaviors. For instance, since most leadership positions are held by men, “men are viewed as more appropriate for leadership positions” (Lanaj and Hollenbeck 2015, p. 1477).

Gender literature has devoted extensive attention to the portrayal of the two gender roles (e.g., Biernat and Danaher 2012; Eagly and Karau 1991; Frieze and Li 2010; Gardner and Gabriel 2004; Heilman and Chen 2005; Lanaj and Hollenbeck 2015; Miller, Perlman, and Brehm 2007; Wood and Eagly 2002). Generally, men are expected “to possess high levels of agentic qualities, including being independent, masterful, assertive, and competent” (Eagly and Karau 1991, p. 686) and considered “more leader-like, intellectual, analytical, able to think abstractly, and able to solve problems” (Carli 2010, p. 343). Women, on the contrary, are expected to be more modest, to exhibit prosocial behaviors, and “to possess high levels of communal attributes, including being friendly, unselfish, concerned with others, and emotionally expressive” (Eagly and Karau 1991, p. 686). Also, gender differences in interpersonal communication were identified, such that men show more dominance and speak more assertively, while women convey warmth, supportiveness, and collaboration (Carli 2010; Hancock and Rubin 2015).

4.2.2.1 The moderating Effect of the Gender of the Feedback Source

With regard to the feedback source, it is important to note that gender role theory describes a perceptual bias that is "largely driven by recognition-based priming that operate[s] at relatively low levels of awareness" (Lanaj and Hollenbeck 2015, p. 1478), with the result that "gender automatically triggers stereotypes and attributions that influence perceptions of leadership" (Lanaj and Hollenbeck 2015, p. 1478), such as the provision of feedback. In turn, "perceivers usually assume that people's behaviors reflect their intrinsic characteristics" (Eagly and Wood 2012, p. 462), implying that the same behaviors of a woman and a man are valued and perceived differently. Similarly, stereotype fit theory suggests that individuals and their behaviors are "perceived in a certain way because of the societal group to which they belong (e.g., women)" (Sturm et al. 2014, p. 662; Dipboye 1985; Heilman 1983; Lee et al. 2015).

Thus, gender role theory suggests that gender stereotypes infect perceptions of feedback. Accordingly, since "being a helper is central to female gender stereotype" (Heilman and Chen 2005, p. 431) and women "tend to be seen by others as more helpful, kind, and devoted to other people" (Frieze and Li 2010, p. 325), feedback from female supervisors should be perceived as more supportive. Consistent with a large body of creativity research that has emphasized the motivational and reciprocal mechanisms of supervisor support in subordinates' creative idea generation (e.g., Shalley and Gilson 2004; Tierney, Farmer, and Graen 1999), feedback from a female source should enhance employee's creativity more than feedback from a male feedback source. Moreover, as feedback from male supervisors "focus more on their subordinates' mistakes and failures" (Carli 2010, p. 338; Eagly, Johannesen-Schmidt, and Van Engen 2003), feedback from female supervisors who “engage more than men do in the socially oriented aspects of interaction and [are] concerned about others’ feelings and group harmony” (Eagly and Karau 1991, p. 686) fosters a work environment in which employees feel psychologically more secure and less fearful of blame for mistakes. This feedback-driven work environment motivates employees to be creative (Blake and Mounton 1985; Edmondson 1999) since in developing “new and useful products or processes, individuals have to be willing to try and to possibly fail” (Shalley and Gilson 2004, p. 36).

In contrast, traditional feedback theory (e.g., Ilgen, Fisher, and Taylor 1979; Kluger and DeNisi 1996; Taylor, Fisher, and Ilgen 1984) predicts that male gender stereotypes favor the effectiveness of feedback. Men are viewed as more leader-like and competent (Biernat and Danaher 2012; Carli 2010), and thus should possess relatively more expertise and power as feedback source. In addition, when giving feedback, male supervisors also “confine themselves more strictly to task concerns and [..] forgo [..] greater involvement in interpersonal matters” (Eagly and Karau 1991, p. 705), which should promote the perception of feedback objectivity and hence the credibility of a male feedback source. However, gender role theory suggests that the male stereotype of being dominant and assertive (Carli 2010) is reflected in the provision of feedback in that men “focus more than women do on controlling their environment and obtaining tangible outcomes” (Eagly and Karau 1991, p. 686). Feedback from a male supervisor may therefore be less likely to enhance employee creativity, as controlling supervisor behaviors have frequently been shown in creativity research to have adverse effects on the recipient’s creativity by reducing self-determination and intrinsic motivation (e.g., Deci, Connell, and Ryan 1989; Shalley and Gilson 2004; Zhou 1998).

So far, gender stereotypes have largely been viewed as a perceptual bias that affects how female or male behaviors are perceived by others. However, gender role theory also states that behaviors of other individuals of the same gender are prescriptive for men and women. Given that meta-analyses support the notion that men communicate more and interrupt their counterparts more often (Anderson and Leaper 1998; Hancock and Rubin 2015; Leaper and Ayres 2007), evidence suggests that women and men communicate differently. Following the prescriptive tenet of gender role theory, this divergence in behavior shapes expectations of gender-specific communication behavior. Meeting the stereotypical gender role expectations, such as men to communicate frequently, is “likely to be disregarded or ignored” (Heilman and Chen 2005, p. 431). Accordingly, frequent feedback from a female supervisor should more likely be perceived as a special effortful act of kindness toward the recipient, while infrequent feedback is merely socially expected. In contrast, frequent feedback from a male supervisor should be more taken for granted, while infrequent feedback from a male supervisor should be perceived as deviation from gender role expectations and thus as particularly negative. I therefore hypothesize:

Hypothesis 2: Feedback source gender moderates the relationship between feedback frequency and recipient creativity, such that women who provide high levels of feedback frequency are associated the highest quantity (a) and quality (b) of creative ideas

4.2.2.2 The moderating Effect of the Gender of the Feedback Recipient

I draw arguments for why women and men respond differently to feedback and thus why feedback recipient gender influences the effect of feedback frequency on recipient creativity by taking up the tenet of gender role theory suggesting that gender stereotypes are also prescriptive and “dictate norms about how women [and men] should behave” (Heilman and Chen 2005, p. 431; Eagly and Wood 2012). While men are socialized to be more self-reliant and dominant, women are expected to be more modest and base their self-concept more on relational aspects (Burton and Hoobler 2006; Carli 2010; Miller, Perlman, and Brehm 2007). Women therefore rely more on the “interpersonal domain and reflected appraisals (reactions of others to them) [and consequently] exhibit greater sensitivity to and a higher need for social approval” (Sturm et al. 2014, p. 671; Burton and Hoobler 2006). Thus, it is not surprising that empirical studies have shown that women respond more sensitively to feedback, for instance, regarding their subsequent effort (Biernat et al. 2020), productivity (Gill and Prowse 2014), emotions (De Castro et al. 2013; Motro and Ellis 2017; Vescio et al. 2005), or self-evaluation (Roberts and Nolen-Hoeksema 1989).

Consequently, gender changes how feedback is perceived and processed. Biernat et al. (2020, p. 142) echo Deci’s (1975) thoughts, arguing that “women may be more likely than men to perceive praise as “controlling” rather than informative”, because women are socialized to be more concerned with external evaluation, whereas men are socialized to be more independent of others and to be more self-oriented. In other words, women perceive feedback as forcing them in a behavioral corset of behaving in order to be liked and positively evaluated, which reduces their willingness to attempt creative ventures that may fail. Besides, it lowers their sense of autonomy and undermines their intrinsic motivation, which in turn is detrimental to employee creativity (Shalley, Zhou, and Oldham 2004; Shalley and Gilson 2004).

Furthermore, by being more reliant on external evaluation, women are more prone to internalize feedback and view “feedback as a more accurate assessment of their abilities than [do] men” (Biernat et al. 2020, p. 142). Particular in case of receiving negative feedback, women “are more vulnerable to attributing these types of setbacks to their own failure” (De Castro et al. 2013, p. 3), while men rather externalize negative feedback, e.g., by giving less credit to the feedback source or finding excuses for inferior performance. So, women face an ambiguity when receiving feedback. On the one hand, women “rely more on feedback and reassurance and when it doesn’t happen, they interpret it as not being valued” (Sturm et al. 2014, p. 671) and therefore yearn for feedback from others (Fletcher 1999). On the other hand, feedback poses a potential threat to their self-image. In alignment to this ambiguity, Wozniak, Harbaugh, and Mayr (2016) found that women may even avoid feedback in competitive environments due to their fear of negative feedback, while men pursue feedback for ego-based utility reasons regardless of their confidence. Avoiding feedback that falls short of one’s expectations may also explain why women under-predict how other organizational members, such as subordinates, peers, and supervisor, rate their performance (Taylor and Hood 2011). Similarly, Sturm et al. (2014) conclude that differences in feedback needs, learned gender-roles, and lack of self-confidence are responsible for the fact that women anticipate to receive less positive feedback. Lastly, drawing on the notion of gender role theory that gender stereotypes impact on the individuals’ perception, female recipients should perceive feedback as more supportive than men since, in general, “higher levels of positive social behaviors are directed at women than at men” (Carli 2010, p. 338; Carli 1989; Johnson, Clay-Warner, and Funk 1996). I therefore assume that women are relatively more receptive to feedback stimuli and consequently respond with higher creative quantity and quality when feedback is provided more frequently.

In addition, women have been shown to exhibit higher levels of reciprocity (Heinz, Juranek, and Rau 2012), and meeting the gender role expectations also requires them to engage more in prosocial behaviors (Gardner and Gabriel 2004;

Wood and Eagly 2002). Since frequent feedback is a valuable organizational resource provided to feedback recipients that helps employees to improve their performance and to experience higher levels of self-efficacy (Fedor and Buckely 1987), women are more likely than men to feel obliged to respond to feedback with greater effort and dedication to compensate for the support provided, such as by generating more and better creative ideas. I argue as follows:

Hypothesis 3: Feedback recipient gender moderates the relationship between feedback frequency and recipient creativity, such that women receiving high levels of feedback frequency display the highest quantity (a) and quality (b) of creative ideas

4.2.2.3 The moderating Effect of Gender (In)congruence

Gender role theory, feedback theory, and the psychological literature offer several arguments that the effects of feedback frequency on recipient creativity depend not only on the gender of the feedback source or recipient, but as well on their combination. Gender role theory suggests that perceptions of gender-specific stereotypical norms of leadership behaviors, such as feedback provision, are contingent on the gender of the observer. For instance, men in particular attribute leadership skills to men (Carli 2010). Consequently, men discredit their female leaders and resist their influence, while they are more likely to be influenced by male leaders (Carli 2010). Moreover, “men are especially likely to help women, whereas women are equally likely to help women and men” (Frieze and Lie 2010, p. 326; Eagly and Crowley 1986). This behavior fits the stereotype of the “gentle man” – helping a woman in need. Similarly, the perception of female feedback recipients is also biased towards the gender of the feedback source. Women expect male feedback providers to have reservations about them as women, so they are predominantly seen as women rather than as members of the organization (Sturm et al. 2014), which is consistent with Eagly and Karau's (1991) concept of gender roles being more salient than work roles. As a result, women believe that "no matter what they do or how they behave, their male bosses will still rate them lower because of perceived gender stereotypes and roles" (Sturm et al. 2014, p. 663). However, in case of a female feedback source, this perceived stereotypical prejudice against female recipients will be less pronounced because female feedback providers belong to the same social group (Sturm et al. 2014). Nevertheless, empirical results also show reverse effects, as feedback from a male feedback source in a male domain has been shown to increase women's confidence, self-efficacy, task interest, and task identification more than that from a female feedback source, while male responses remained unaffected by the gender of the feedback source (Park et al. 2018).

More generally, Carli (2010) concludes that "gender differences in communal behavior generally are more pronounced in same-gender interactions, [while] gender differences in agentic behavior generally are most pronounced in mixed-gender interactions" (Carli 2010, p. 338 f.), such that "women are less agentic when interacting with men, and men are more inclined to take charge in interactions with women" (Carli 2010, p. 339) and "the greatest warmth shown is in all-female interactions and the least in all-male interactions" (Carli 2010, p. 338). Thus, the previously described divergence in perceptions of leadership behavior that results from gender-stereotypical attribution of agentic and communal characteristics to the feedback source is dependent on the gender of the feedback recipient. Consequently, the effectiveness of a female or male feedback source also depends on the gender of the recipient.

Considerations of the similarity of feedback recipient and feedback source allow a further perspective on the combination of their genders. It is beyond doubt that gender plays an important role in the individual’s assessment of social group membership. Hence, women “see themselves as being dissimilar to their male bosses in particular because they belong to a different social (gender) group” (Sturm et al. 2014, p. 664). Psychological research has revealed that individuals distrust evaluators who do not originate from their social group and question their objectivity, which is referred to as “attributional ambiguity account” (Biernat and Danaher 2012; Major and Sawyer 2009; Mendes et al. 2008). Similarly[115], the feedback literature has emphasized the impact of psychological closeness of feedback source and feedback recipient on the effectiveness of feedback (Greller and Herold 1975; Ilgen et al. 1979; Luckett and Eggleton 1991), as “individuals rely most upon sources close to themselves” (Ilgen et al. 1979, p. 353) and sources “closer to the individual, are seen as providing more feedback information than those sources identified as […] psychologically distant” (Greller and Herold 1975, p. 255). Therefore, gender congruence should increase the effect of feedback frequency on employee creativity. I state my fourth hypothesis as follows:

Hypothesis 4: Gender congruence between feedback source and feedback recipient moderates the relationship between feedback frequency and recipient creativity, such that high levels of feedback frequency and gender congruence results in the highest quantity (a) and quality (b) of creative ideas

4.2.3 Decoupling of Feedback Frequency from Feedback Timing Effects

The implications of several dimensions of feedback on recipient creativity have been observed, e.g., valence (Hoever, Zhou, and van Knippenberg 2018; Kim and Kim

[115] Brookins et al. (1996, p. 245) state that “psychological closeness is the most basic form of group identification”.

2020; Zhou 1998), quantity (Coelho and Augusto 2010; Hildebrand et al. 2013), and source (Agnihotri et al. 2014a; Hon, Chan, and Lu 2013). In this study, a major target is to examine the effects of feedback frequency on recipients' creativity. While other feedback dimensions are related to frequency, such as feedback valence, since supervisors try to avoid negative feedback "because of the unpleasant nature of relaying such information to others" (Ilgen, Fisher, and Taylor 1979, p. 367; Larson 1986), or feedback quantity[116], particularly feedback timing should highly correlate[117] to feedback frequency in practical settings.

According to Bilodeau (1966), feedback timing consists of two sub-dimensions: *feedback delay*, which refers to the time between employee's response and the corresponding subsequent feedback, and *postfeedback interval*, which is "the interval between the feedback and the next response" (Ilgen, Fisher, and Taylor 1979, p. 354). It is widely recognized that feedback delay has negative effects on feedback effectiveness (Ammons 1956; Berger and Ludwig 2007; Goomas, Smith, and Ludwig 2011; Houde et al. 2013; Kirby 2009; Luke and Alavosius 2011; Ludwig and Goomas 2009; Papa, Aldrich, and Schumacker 1999; Shimada et al. 2014). The rationale behind this notion is that interfering activities and events that occur between the employee's response and the corresponding feedback dilute the employee's memory of the previous response and make the comparison with the provided feedback less effective (Ilgen, Fisher, and Taylor 1979) – not to mention that this is equally true for the feedback provider, who may remember the employee's response less accurately.

[116] It is to be expected that the sum of 12 monthly feedback sessions will last longer than an annual feedback meeting.
[117] While feedback frequency determines the number of feedbacks within a given period of time, feedback timing determines the time span between two employee responses to which feedback is given (Andiola 2014; Bildeau 1966). Assuming a series of employee responses separated by feedbacks on the latest response (Andiola 2014), feedback timing mathematically represents the inverse of feedback frequency.

However, when cognitively intensive tasks are examined, such as learning (Metcalfe Kornell, and Finn 2009; Nakata 2015; Thornock 2016), delayed feedback has been shown to have positive effects. Similarly, it has been proposed that immediate feedback (i.e., there is no feedback delay) impairs employee creativity due to cognitive overload (Lechermeier and Fassnacht 2018). Furthermore, a lower feedback delay, ceteris paribus, increases the postfeedback interval. Metcalfe, Kornell, and Finn (2009) argue that previous studies on feedback delay do not (sufficiently) control for the postfeedback interval and show that the benefits of immediate feedback are less pronounced when controlling for the postfeedback interval (which the authors refer to as “lag to test”). Particularly in the context of creativity, a longer postfeedback interval allows more time for creative ideas to emerge, referred to as creative incubation period, during which the creative task is set “aside and though about consciously” (Weisberg 1999, p. 45; Dodds, Ward, and Smith 2012; Gilhooly 2016), because “effective incubation necessitates time away from the task rather than just additional time devoted to the task” (Kachelmeier, Wang, and Williamson 2019, p. 252). Thus, it remains ambiguous whether feedback timing impacts employee creativity. I therefore pose the following research question:

Research Question: Does feedback timing affect the (a) quantity and (b) quality of recipient’s creativity?

4.3 Methodology

To test the hypotheses, I conduct a 2 × 2 between-subject field experiment in which I vary two factors. First, I manipulate the frequency of the provided feedback at two levels: low and high. Second, I provide feedback to participants from a female or a male supervisor. As participants indicate their gender, this simultaneously influences gender congruence between feedback provider and feedback recipient.

4.3.1 Overview of Experimental Procedure

The experimental field study is implemented in a university course. 199 business students participate in a two-stage online tutorial (see Figure 4.1). In the first stage, the students receive a worksheet with five tasks that reflect assignments in the final exam (the students are informed about this). During a work period of 11 days, the participants submit their individual solutions to one of the two supervisors to whom they have been assigned. After a short period of time to evaluate the submissions, the students receive individual feedback on their performance from their supervisor via email. Subsequently, the second stage begins, which is designed similarly except for the tasks to be completed. Importantly, the fifth task in the second stage (i.e., task 10) allows students to submit creative ideas for multiple-choice tasks that relate to the content of the university course. A total of 528 creative ideas were submitted.

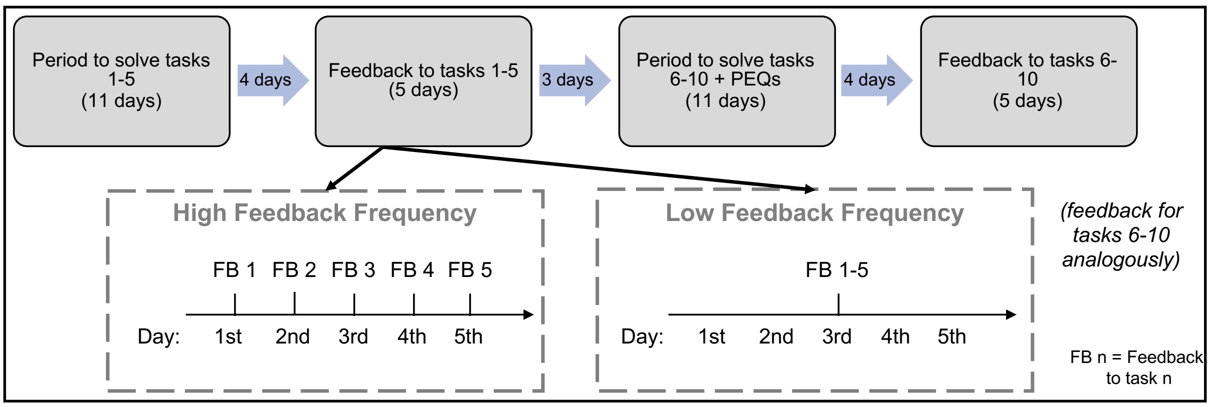


Figure 4.1: Sketch of the Experimental Procedure and the Feedback Frequency Manipulation

4.3.2 Experimental Tasks

The tasks reflect common assignments that are conducted in the course. More specifically, each task refers to a specific course topic (e.g., cost accounting). I do this with the aim that feedback to any task is meaningful and provides non-redundant information to the participants. To reinforce the importance of every single feedback, the participants are explicitly told that the tasks are equally important.[118] Prior to the experiment, the course procedure offered no external performance evaluation, and performance on the experimental tasks has no relevance for the students' course grade, which is based solely on the final exam. Although participants did not have the opportunity to interact or communicate with other participants owing to the formal course organization, it is emphasized that the tasks were useful preparation for the final exam if the tasks are worked on alone and that group work is not accepted[119] to prevent individual participants from collaborating. The submitted data do not indicate such behavior (e.g., redundancy of submitted solutions and timing of submission is examined). No financial incentives are offered.

4.3.3 General Feedback Design

Relative performance information (RPI) is employed to provide feedback to the participants about their performance in each task. The RPI indicates the proportion of outperformed participants in a continuous range between 0 and 1. Despite participants receive private RPI (i.e., they know their own performance but not the performance of any particular other participant) (Hannan et al. 2013; Tafkov 2013), they are aware of the number of participants. I choose RPI specifically for conveying feedback, as it is

[118] Pretests with 6 participants (who were later excluded from the analysis) substantiate my intention that the tasks were perceived as equally important.

[119] Since performance in the experimental tasks has no relevance to the final course grade, part of the motivation for participants to follow the rules is to gain access to the task solutions, which are only available to participants who fully complete the experimental procedure while following the rules.

particularly prone to distract recipients from other activities (Holderness, Olsen, and Thornock 2020). Due to the field design of the experiment, a powerful feedback is required that is not drowned out by the day-to-day activities of the participants. In particular, it is to be expected that participants have several other university courses to complete and exert effort into while the field experiment takes place. The experimental design therefore follows the advice of Tafkov (2013) to enhance the effect of the RPI by using the same tasks between participants, by having participants who share similar attributes (all participants are business students taking the same university course), and by designing the field setting of the experiment to ensure that the domain is important to participants and that they desire to do well in their job (e.g., the tasks represent common course activities). I also refrain from giving absolute feedback (e.g., “in task one, you scored 55% of all possible points”), first to better cover the range of possible feedback valence, second to account for the phase of the university course in which the experiment takes place, which does not expect participants to solve all tasks perfectly correctly, third to avoid confounding feedback valence with task difficulty, and fourth to avoid interference with the RPI provided (i.e., there is a unambiguously unique feedback valence).

The assessment of participants’ performance is highly standardized (see chapter 2.2.2). Nonetheless, both supervisors evaluate all participants independently to ensure homogeneity. Apart from negligible minor differences that can be easily resolved, a uniform and consistent assessment is thus ensured. Pretests and experimental results show that the participants’ performance is well distributed across this range. In order to increase trust in the experimental procedure and to allow for comparisons between treatments, the RPI is based on the performance of all sample. Both supervisors send standardized emails to transmit the feedback to their respective students. Specific solutions to the tasks are not made available until after the experimental procedure. Figure 4.2 shows a feedback email in the low feedback frequency condition, in which each email provides feedback on only one task.

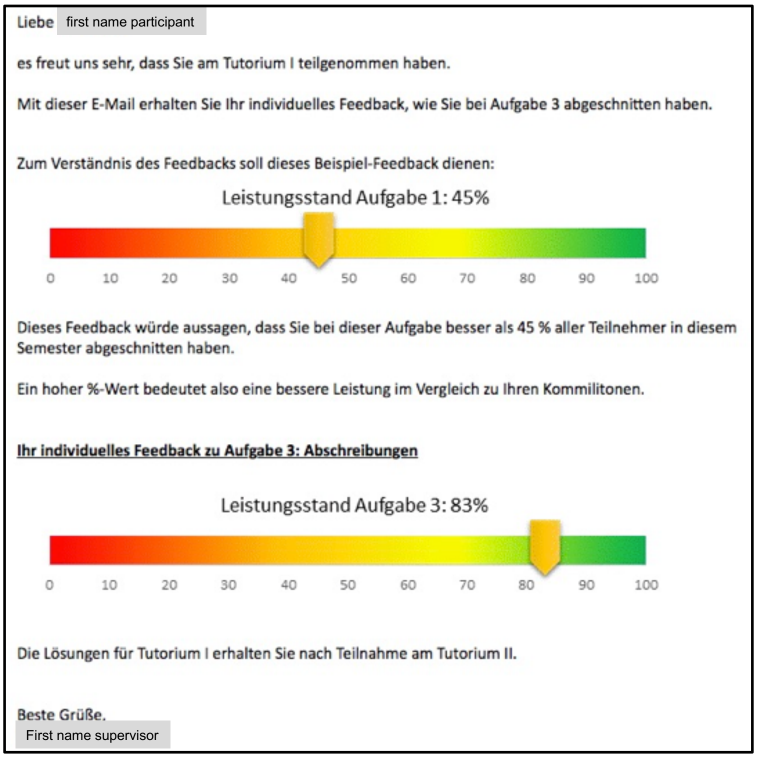


Figure 4.2: Feedback Email in the low Feedback Frequency Condition

4.3.4 Manipulations

To test the predictions, I manipulate the feedback provided to the participants in two ways and randomly assign participants to the treatment groups. The *level of feedback frequency* determines whether participants receive feedback on five consecutive days (high feedback frequency) or on a single day on all tasks simultaneously (low feedback frequency). In other words, participants in the high feedback frequency treatment receive five emails – each containing feedback on one specific task, while participants in the low feedback frequency treatment receive only one email that contains feedback, which is still reported separately for each task (see Figure 4.3). Importantly, this procedure ensures that the overall amount of feedback and the level of feedback detail remain constant across the different treatments (Casas-Arce, Lourenco, and Martinez-Jerez 2017). A manipulation check provides support for the effectiveness of the manipulation, since the differences in participants' perception of the feedback frequency ("my supervisor provides feedback frequently") are consistent with the manipulation (means 4.22 vs. 3.39, p = 0.0%).

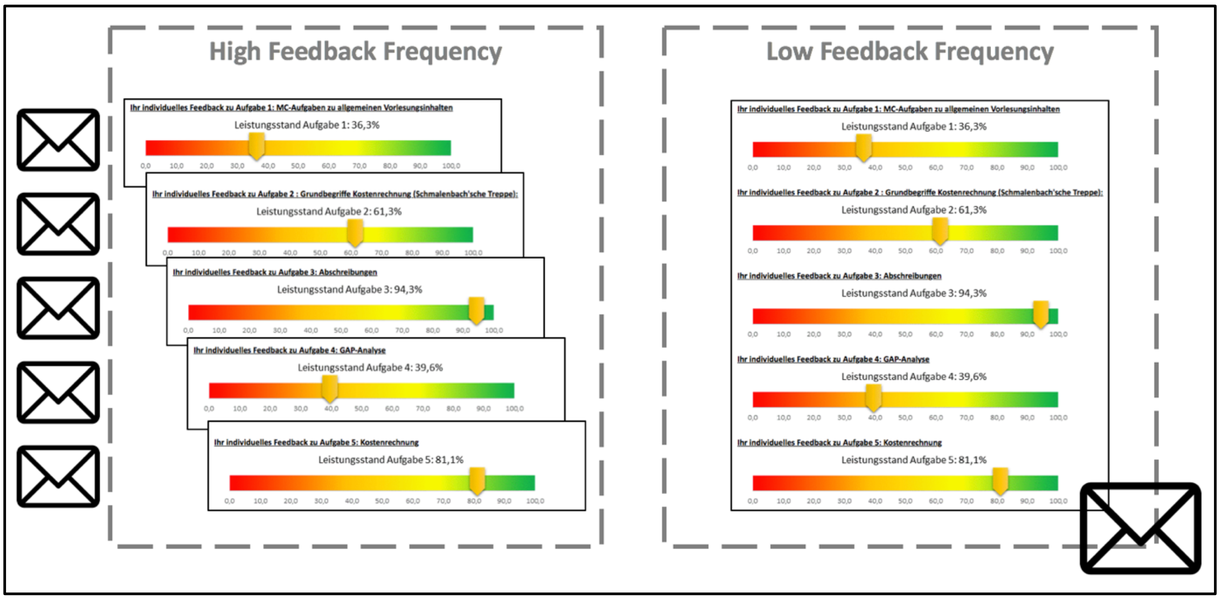


Figure 4.3: Visualization of the Manipulation of Feedback Frequency

To manipulate the *gender of the feedback source,* we use the first names ("Julia" and "David") of the students' supervisors to emphasize their gender (non-deception) when they communicate with the students (Kaas and Manger 2011; Hardacre and Subasic 2018). Their surnames are among the most common names in Germany in the participants' generation.[120] I refrain from providing further details about the supervisors (see chapter 2.2.2), such as pictures or entries on the department's homepage, in order to control for secondary attributes and to prevent private

[120] I refer to the site www.beliebte-vornamen.de [retrieved 05-21] and the "Gesellschaft für deutsche Sprache e.V." (https://gfds.de/vornamen/beliebteste-vornamen/#topten [retrieved 05-21]). Unlike, for instance, Great Britain (UK Office for National Statistics, ONS), no official statistics for first names are available in Germany.

investigations of the participants (such as social media or asking student friends). The supervisors are neither visible nor audible to the students before and during the experimental procedure. They communicate with the students only via standardized emails. More specifically, in addition to the feedback emails (see Figure 4.2), we send out a welcome email as priming. Again, a manipulation check supports the effectiveness of the manipulation.

4.3.5 Dependent Variables

After receiving feedback on the first stage tasks, participants are asked to submit creative ideas on multiple-choice questions. This opportunity is integrated into the tasks of the second stage (task 10) to allow a natural embedding into the flow of the online course. In doing so, participants explicitly face the trade-off of investing more (less) effort in generating creative ideas that could otherwise be invested in solving the other (non-creative) tasks. Participants are informed that their submissions to task 10 are evaluated based on quantity (i.e., number of ideas) and quality (i.e., originality, innovativeness, and cleverness). As an incentive for submission, it is announced that a selection of these creative ideas will be included in the final exam.

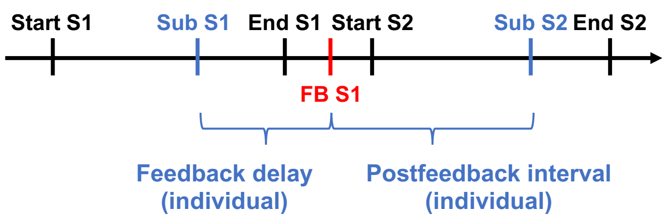
Similar opportunities for students to propose self-created assignments for the final exam exist in several other courses at the participants' university; in addition, creating new assignments is also a convenient approach to preparing for the exam that was addressed in the course, so students latently search for creative assignment ideas. Thus, the overall experimental setting (inclusion in ordinary procedure of a university course) and the tasks (natural course assignments) correspond to the distinction between laboratory and field experiments according to Bloomfield, Nelson, and Soltes (2016): "Researchers elicit dependent variables when they pose questions and present tasks to people and observe responses that otherwise would not occur, as is the case in surveys, lab studies, and lab experiments" (p. 356), whereas in field experiments, dependent variables are not elicited, but researchers "observe dependent variables as data occurs naturally in practice settings" (p. 360). The number of ideas that can be submitted is not limited.

To measure creativity, I take the mean number of all ideas submitted by participants to evaluate the *quantity of creative ideas*, while I take the mean number of high-creativity ideas only to evaluate the *quality of creative ideas*. In order to identify high-creativity ideas, I adhere to the procedure of Kachelmeier, Wang, and Williamson (2019):[121] 6 creativity raters independently and privately evaluate all submitted ideas on a scale from 1 (lowest) to 5 (highest). These raters are volunteering graduated business students who are well-versed in the university course's content and are not experimental participants. The order of the ideas to be evaluated is randomly determined per rater. The evaluation is based on the instructions provided to the experimental participants. In addition, raters are blind to submitting participant and experimental treatment condition. Second, to reduce noise in the dependent variable, the rater with the lowest correlation to the average evaluation of the five other raters is eliminated. Furthermore, the highest and lowest individual scores for each idea are omitted to reduce the impact of outliers. Third, each raters' evaluations are z-standardized. If an idea's average z-value of the three remaining evaluations exceeds 1 (i.e., +1 standard deviation), the idea is considered to be a high-creativity idea.

[121] The only difference to the procedure of Kachelmeier, Wang, and Williamson (2019) lies in the scale used for evaluation (1 to 10) and the associated predefined threshold for classifying ideas as highly creative. Regarding the scale, pretests have shown that the raters feel more comfortable with a 5-point scale and consider it to be sufficiently differentiated. In turn, the rigid threshold (of 6) used by Kachelmeier, Wang, and Williamson (2019) to identify high-creativity ideas is no longer applicable. Nonetheless, the share of ideas identified as highly creative in this study (15.9%) is in the same range (14.1% and 16.1% in Kachelmeier, Wang, and Williamson 2019).

4.3.6 Further Experimental Variables

In addition to feedback frequency and the gender of the feedback source, the experimental design allows the measurement of three other dimensions of feedback that are critical to feedback effectiveness (cf. Ilgen, Fisher, and Taylor 1979) and are related to feedback frequency. These variables are not actively manipulated, but are measured to statistically control for alternative explanations (see chapter 2.2.2). First, participants receive a certain *feedback valence* on the RPI-continuum of outperformed peer participants (values range from 0 to 1), so higher values indicate a (more) positive feedback. Importantly, the feedback provision is designed independently of feedback valence, as “interpersonal sources of feedback may delay giving negative feedback” (Ilgen, Fisher, and Taylor 1979, p. 367; Larson 1986) and previous field studies of feedback frequency may be biased toward feedback valence.



Legend	Controlled?	Description
Start S1	yes, fixed date	Start of first stage (i.e., distribution of tasks 1-5)
Sub S1	*no, individually*	Submission of solutions (tasks 1-5) of a particular participant
End S1	yes, fixed date	General deadline for submissions of solutions to tasks 1-5
FB S1	yes, fixed date	Feedback on tasks 1-5
Start S2	yes, fixed date	Start of second stage (i.e., distribution of tasks 6-10)
Sub S2	*no, individually*	Submission of solutions (tasks 6-10) of a particular participant
End S2	yes, fixed date	General deadline for submissions of solutions to tasks 6-10

Figure 4.4: Definition of Feedback Delay and Postfeedback Interval based on the Experimental Procedure

Second, participants are allowed to submit their solutions at any time after the tasks are made available, but before the announced deadline. Therefore, I can assess how long each participant has to wait individually for feedback (on tasks 1-5) after completing: The earlier a participant submits in the first stage, the longer he or she has to wait for feedback on the first stage tasks, which is provided to all participants at the same fixed date (see Figure 4.4). I refer to this time period as *feedback delay* (Thornock 2016). Third, similar differences at participant level emerge in the amount of time participants spend processing the provided feedback (on tasks 1-5) when working on the second stage (tasks 6-10): The earlier a participant submits in the second stage, the shorter the time he or she can work with the provided feedback from the first stage (see Figure 4.4). I refer to this time span as *postfeedback interval* (Ilgen, Fisher, and Taylor 1979). In particular, this time period is measured since creative incubation (recall: Creative ideas are submitted by means of task 10) has been shown to boost the generation of creative ideas (Dodds, Ward, and Smith 2012; Gilhooly 2016). Since participants in the high feedback frequency condition receive feedback over a longer period of time (i.e., five days), I consistently refer to the median day[122] (i.e., day three) of the feedback provision phase when measuring the latter two variables. Thus, the research design allows me to separate the effects on recipient creativity of the various factors that are often intertwined in practice (and in some publications). A definition of all experimental variables and gender variables can be found in Figure 4.5.

---

[122] Using the first or the last (i.e., the fifth) day as the reference point for either variable does not substantially alter the results.

Variable Name	Description
Feedback frequency	Manipulated variable. Dummy variable indicating whether the participant receives feedback on tasks 1-5 (and on tasks 6-10) once for all five tasks (i.e., low feedback frequency = 1) or separately for each task on five consecutive days (i.e., high feedback frequency = 2)
Feedback valence	Mean of participant's share of outperformed peers (RPI-score) for tasks 1-5 as indicated in the feedback
Submitting participants	Share of participants that submit at least one creative idea
Quantity of creativity	Mean number of ideas submitted per participant
Quality of creativity	Mean number of high-creativity ideas submitted per submitting participant
Feedback delay	Participant's individual time span (in hours) between his or her submission of tasks 1-5 and feedback on tasks 1-5
Postfeedback interval	Participant's individual time span (in hours) between feedback on tasks 1-5 and his or her submission of tasks 6-10
Gender source	Manipulated variable. Dummy variable indicating whether the participant receives feedback from a female (= 1) or a male (= 2) feedback source
Gender recipient	Dummy variable indicating the participant's (i.e., feedback recipient's) gender as indicated by the participant in both sessions (1 = female; 2 = male)[123]
Gender congruence	Dummy variable indicating whether the participant (i.e., the feedback recipient) and his or her assigned supervisor are of the same gender (= 1) or not (=2)[124]

Figure 4.5: Definition of Experimental Variables and Gender Variables

[123] We allow and control for participant gender fluidity and nonbinarity by measuring participant gender in each session and offering "divers" as an selectable option (Butler 2011; Fotaki, Metcalfe, and Harding 2014; https://www.health.harvard.edu/blog/gender-fluidity-what-it-means-and-why-support-matters-2020120321544 [retrieved 06-21]). However, none of the participants used this option or altered their indication during the experimental procedure.

[124] We follow Romero et al. (2021) in their definition of "gender congruence" as the interpersonal alignment of sender and recipient gender. Graham, Dust, and Ziegert (2018) speak similarly of "supervisor-employee gender (dis)similarity". Nonetheless, we are aware of the term "gender congruence" – but explicitly do not refer to it – in the way it has been used in other literature streams as an intrapersonal measure of harmony in one's gender (Jones et al. 2019; https://genderspectrum.org/articles/understanding-gender [retrieved 06-21]) or as a match of consumers' gender with the perceived gender-specificity of a product (Fugate and Phillips 2010).

4.3.7 Post-experimental Procedure

Participants complete the post-experimental questionnaire after they submitted the solutions to tasks 6-10. In alignment to their experimental condition (low/high feedback frequency; female/male feedback source), participants also receive feedback on tasks 6-10. Subsequently, all participants who have fully completed the experimental procedure receive access to the detailed solutions to tasks 1-9. Afterwards, all participants are offered to take part in a presentation of the research questions and results (debriefing). Although all participants granted permission to analyze the data scientifically, follow-up interviews with some participants confirm that, as intended, the experimental setting was not noticed. Thus, there do not appear to be any reactive effects (Campbell and Stanley 1963) that could impair the external validity of the experiment. Figure 4.6 provides a more detailed overview of the experimental procedure.

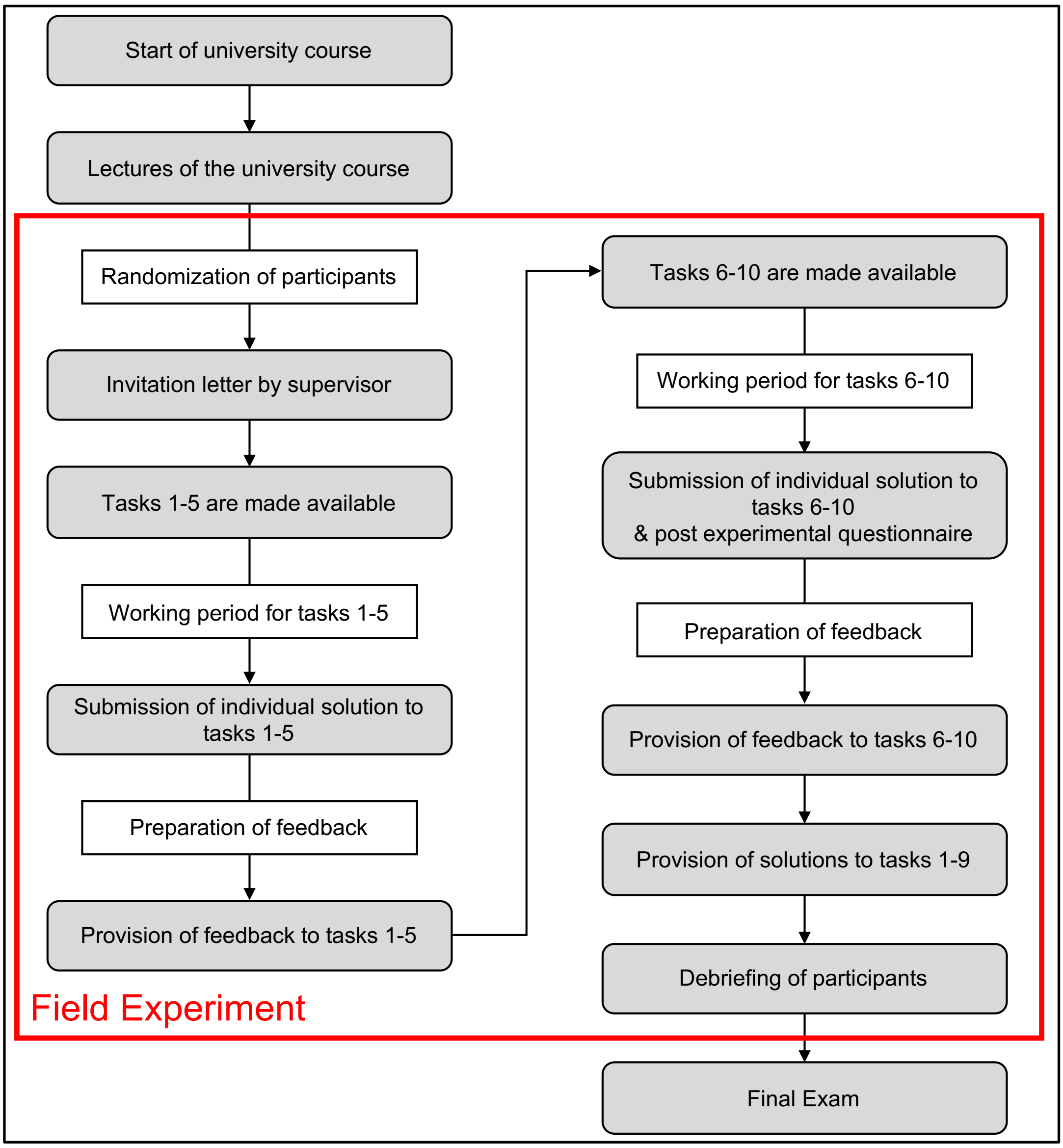


Figure 4.6: Overview of the Experimental Procedure and its Integration into the University Course

4.3.8 Sample Description

The final sample includes 105 female (55.6%) and 84 male (44.4%) business students[125] with an average age of 23.8 years. Descriptive cross-sectional data show that the randomization appears to be successful (see Figure 4.7), as I am not able to find significant differences (Kruskal-Wallis test, all p > 10%, two-tailed) across the

[125] 10 participants are excluded from the study for dropping out after the first stage (5), submitting too late (3), or indicating in the post-experimental questionnaire that they received no feedback (2). Their inclusion does not substantially alter results.

treatments that may interfere with the results (Kruskal and Wallis 1952). A correlation matrix of experimental and cross-sectional variables is provided in Figure 4.8.

Cross-Sectional Data Summary Statistics – by manipulations (Feedback Frequency and Gender Source)

Descriptive Statistics
(Means [Standard Deviations])

	Treatment [a]			
Variable [b]	Low FF + FeSou	Low FF + MSou	High FF + FeSou	High FF + MSou
Age	23.7 [4.6]	23.1 [3.3]	24.2 [3.7]	24.1 [3.9]
Gender recipient	1.36 [0.5]	1.43 [0.5]	1.53 [0.5]	1.46 [0.5]
German as native language	1.09 [0.3]	1.13 [0.3]	1.19 [0.4]	1.08 [0.3]
Undergraduate study	1.02 [0.1]	1.02 [0.1]	1.00 [0.0]	1.02 [0.1]
Employment	0.81 [0.4]	0.79 [0.4]	0.70 [0.5]	0.69 [0.5]
Self-assessment of tasks 1-5	56.7% [15.8%]	56.8% [14.1%]	56.2% [18.4%]	54.6% [19.9%]
Expected valence on tasks 1-5	35.9% [21.4%]	29.4% [16.1%]	32.3% [17.6%]	32.2% [20.7%]
Perceived difficulty of tasks 1-5	3.5 [1.0]	3.5 [1.1]	3.3 [1.0]	3.5 [1.0]
Interest in course	5.6 [0.9]	5.7 [1.2]	5.7 [1.1]	5.4 [1.3]
Achievement striving	5.4 [1.0]	5.0 [1.2]	4.8 [1.2]	4.9 [1.1]
n	47	47	47	48

a. *Low FF* = low feedback frequency; *High FF* = high feedback frequency; *FeSou* = female feedback source; *MSou* = male feedback source.

b. *Age* = age of participants in years; *Gender recipient* = dummy variable for participants' gender as indicated by each participant in both sessions (1 = female; 2 = male); *German as native language* = dummy variable for participants' self-assessed ability in the German language (1 = maximal value on the 5-point scale; 2 = any other value); *Undergraduate study* = dummy variable for participants' current study program (1 = Bachelor; 2 = Master); *Employment* = dummy variable for whether participants are employed besides their studentry (0 = no; 1 = yes); *Self-assessment of tasks 1-5* = participants were asked to estimate which proportion of tasks 1-5 they answered correctly; *Expected valence on tasks 1-5* = participants were asked to estimate the share of participants they expect to outperform in tasks 1-5 (prior to the feedback); *Perceived difficulty of tasks 1-5* = participants were asked to indicate how difficult they perceived tasks 1-5 (on a 7-point scale, with higher scores expressing higher perceived difficulty); *Interest in course* = participants were asked to indicate their interest in the university course the experiment was integrated into (on a 7-point scale, with higher scores expressing higher perceived interest); *Achievement striving* = participants' personality trait assessed with four items from the respective subscale of the revised NEO Personality Inventory (Costa and McCrae 1992; McCrae and John 1992).

Figure 4.7: Cross-Sectional Data Summary Statistics – by Manipulations (Feedback Frequency and Source Gender)

Correlation Matrix

Variable [a]	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1 Gender source	1																		
2 Feedback frequency	.005	1																	
3 Feedback valence	.026	-.102	1																
4 Submitting participants	.015	-.048	-.004	1															
5 Feedback delay	-.011	-.043	**.242**	.101	1														
6 Postfeedback interval	.032	-.073	**-.150**	**-.205**	**-.473**	1													
7 Quantity of creativity	.018	-.046	.090	**.759**	.136	**-.207**	1												
8 Quality of creativity	-.062	**.335**	.094	n/a	.152	**-.207**	**.368**	1											
9 Age	-.042	.097	-.015	.029	**-.176**	.037	-.067	-.083	1										
10 Gender recipient	-.005	.102	.022	.017	.061	-.025	-.034	.079	.091	1									
11 Gender congruence	.111	.069	-.045	-.006	.132	-.047	.031	.033	-.066	-.005	1								
12 German as native language	-.051	.047	**-.214**	.011	**-.197**	**.143**	.067	-.123	**.150**	-.138	.014	1							
13 Undergraduate study	.042	-.043	-.114	.030	-.121	.099	.008	-.032	.105	-.114	.042	**.212**	1						
14 Employment	-.021	-.118	.041	.010	-.038	.048	.008	.132	-.014	.008	.003	-.006	-.023	1					
15 Self-assessment of tasks 1-5	-.023	-.042	**.457**	-.054	.045	-.009	-.066	.120	-.075	**.229**	-.03	-.112	.082	-.065	1				
16 Expected valence on tasks 1-5	-.087	-.009	**.213**	-.024	**.155**	-.005	.009	.116	-.015	**.341**	.065	-.100	.050	.050	**.485**	1			
17 Perceived difficulty of tasks 1-5	.050	-.034	**.202**	-.138	-.034	.041	-.098	-.014	-.089	.038	.029	-.016	.055	.026	.115	.137	1		
18 Interest in course	-.039	-.048	**.170**	.098	.062	**-.192**	**.190**	.058	-.048	-.038	-.03	.015	.025	-.052	.019	**-.149**	-.004	1	
19 Achievement striving	-.044	-.138	.122	.092	.086	.023	.086	-.110	-.119	**-.253**	-.116	-.083	.086	-.057	.066	-.081	-.031	**.176**	1

[a] p-values in bold are significant at $p < 0.05$ (two-sided); variables 1-9 are derived from the experimental procedure, variables 10-19 are cross-sectional data; please refer to Figure 4.5 for definitions of the experimental and gender variables; *Age* = age of participants in years; *German as native language* = dummy variable for participants' self-assessed ability in the German language (1 = maximal value on the 5-point scale; 2 = any other value); *Undergraduate study* = dummy variable for participants' current study program (1 = Bachelor; 2 = Master); *Employment* = dummy variable for whether participants are employed besides their studentry (0 = no; 1 = yes); *Self-assessment of tasks 1-5* = participants were asked to estimate which proportion of tasks 1-5 they answered correctly; perceived difficulty of tasks 1-5 = participants were asked to indicate how difficult they perceived tasks 1-5 (on a 7-point scale, with higher scores expressing higher perceived difficulty); *Expected valence on tasks 1-5* = participants were asked to estimate the share of participants they expect to outperform in tasks 1-5 (prior to the feedback); *Interest in course* = participants were asked to indicate their interest in the university course the experiment was integrated into (on a 7-point scale, with higher scores expressing higher perceived interest); *Achievement striving* = participants' personality trait assessed with four items from the respective subscale of the revised NEO Personality Inventory (Costa and McCrae 1992; McCrae and John 1992); n = 189.

Figure 4.8: Correlation Matrix of Experimental and Cross-Sectional Data

4.4 Results

4.4.1 Hypotheses Testing

The 2 x 2 factorial design to test hypotheses H1a/b and H2a/b, which predict that both feedback frequency (1 = low feedback frequency; 2 = high feedback frequency) and the interaction of feedback frequency and gender source (1 = female feedback source; 2 = male feedback source) influence recipient creativity, is consistent with the manipulations conducted. While an analysis of the descriptive cross-sectional data already indicated that randomization was successful (see sample description and Figure 4.7), the experimental control variables (i.e., feedback delay, postfeedback interval, and feedback valence) that could potentially confound the cause-effect relationship of feedback frequency on creativity also do not differ significantly between treatments (Kruskal-Wallis test, all $p > 10\%$, two-tailed). The cell sizes, means, and standard deviations are presented in Panel A of Figure 4.9.[126]

[126] Please refer to Figure 4.5 for definitions of the experimental and gender variables.

Panel A: Experimental Data Summary Statistics – by Feedback Frequency and Gender Source

Descriptive Statistics
(Means [Standard Deviations])

Variable [b]	Treatment [a] Low FF + FeSou	Low FF + MSou	High FF + FeSou	High FF + MSou
Feedback valence	45.0% [17.1%]	45.3% [16.9%]	41.0% [16.0%]	42.5% [18.1%]
Submitting participants	57.4% [50.0%]	57.4% [50.0%]	51.1% [50.5%]	54.2% [50.4%]
Feedback delay	208.4 [55.1]	214.1 [52.2]	210.7 [54.6]	202.8 [51.4]
Postfeedback interval	254.5 [69.1]	255.6 [57.6]	241.6 [67.5]	249.1 [70.4]
Quantity of creativity	2.87 [3.3]	3.02 [3.4]	2.60 [3.4]	2.69 [3.4]
Quality of creativity	0.22 [0.4]	0.56 [0.7]	1.54 [1.6]	0.85 [1.3]
n	47	47	47	48

a. *Low FF* = low feedback frequency; *High FF* = high feedback frequency; *FeSou* = female feedback source; *MSou* = male feedback source.

Panel B: Effect of Feedback Frequency and Gender Source on Quantity of Creativity

Factors [a]	pre-dicted	df	mean squares	F	p-value (two-tailed)
Constant		1	1475.521	130.743	0.000
Feedback frequency	+	1	4.400	0.390	0.533
Feedback frequency x Gender source	-	1	0.039	0.003	0.953
Gender source		1	0.684	0.061	0.806
Error		185	11.286		

a. DV = Quantity of creativity; n = 189.

Panel C: Effect of Feedback Frequency and Gender Source on Quality of Creativity

Factors [a]	pre-dicted	df	mean squares	F	p-value (two-tailed)
Constant		1	64.986	53.852	0.000
Feedback frequency	+	1	16.811	13.930	0.000
Feedback frequency x Gender source	-	1	6.865	5.688	0.019
Gender source		1	0.851	0.705	0.403
Error		100	1.207		

a. DV = Quality of creativity; n = 104.

Figure 4.9: Effect of Feedback Frequency and Gender Source on Creativity with descriptive Experimental Data

Panel A: Effect of Feedback Frequency and Gender Source on Quantity of Creativity – including Control Variables (ANCOVA)

Factors [a]	pre-dicted	df	mean squares	F	p-value (two-tailed)
Constant		1	58.462	5.333	0.022
Feedback frequency	+	1	4.938	0.450	0.503
Feedback frequency x Gender source	-	1	0.001	0.000	0.991
Gender source		1	1.109	0.101	0.751
Feedback valence		1	4.513	0.412	0.522
Feedback delay		1	2.036	0.186	0.667
Postfeedback interval		1	56.345	5.140	0.025
Gender recipient		1	2.802	0.256	0.614
Error		181	10.961		

a. DV = Quantity of creativity; n = 189.

Panel B: Effect of Feedback Frequency and Gender Source on Quality of Creativity – including Control Variables (ANCOVA)

Factors [a]	pre-dicted	df	mean squares	F	p-value (two-tailed)
Constant		1	0.405	0.335	0.564
Feedback frequency	+	1	14.261	11.784	0.001
Feedback frequency x Gender source	-	1	6.097	5.038	0.027
Gender source		1	0.815	0.674	0.414
Feedback valence		1	0.857	0.708	0.402
Feedback delay		1	0.635	0.524	0.471
Postfeedback interval		1	0.871	0.720	0.398
Gender recipient		1	0.062	0.051	0.822
Error		96	1.210		

a. DV = Quality of creativity; n = 104.

Figure 4.10: Effect of Feedback Frequency and Gender Source on Creativity with Controls

To test H1a and H2a, which predict that feedback frequency and the interaction of feedback frequency and *gender source* have a positive and negative effect on the quantity of recipients' creativity, respectively, an analysis of variance (ANOVA)[127] is conducted in a saturated model (see Panel B of Figure 4.9[128]). Surprisingly, the two-way analysis of variance yields that neither the main effect of feedback frequency ($F(3, 185) < 0.01$, $p = 0.81$) nor the interaction effect of feedback frequency and gender source ($F(3, 185) = 0.39$, $p = 0.53$) are significant, implying that there are no differences between the experimental groups regarding the quantity of creativity. A two-way analysis of covariance (ANCOVA) is subsequently performed to test the robustness of the previous results by controlling for feedback valence, feedback delay, postfeedback interval, and gender of the recipient (see Panel A of Figure 4.10[129]). As expected, the main and interaction effect sizes of feedback frequency and gender source remain both similar and insignificant. However, while the other controls show no significant influence, the postfeedback interval is found to have a significant effect on the quantity of creativity ($F(7, 181) = 5.14$, $p = 0.03$), such that early responses to the feedback

---

[127] ANOVA is based on F-tests. Therefore, it is implicitly assumed that (a) the dependent variable is normally distributed and (b) variance homogeneity exists between the analyzed groups when ANOVAs are performed (Blanca et al. 2018).
(a) However, as ANOVA results (respectively F-tests) have repeatedly been noted to be robust to violations of the non-normal data distribution assumption (e.g., Blanca et al. 2017; Glass, Peckham, and Sanders 1972; Harwell et al. 1992; Lix, Keselman and Keselman 1996), I do not consider this assumption any further.
(b) Moreover, recent research (Blanca et al. 2018) supports, through a Monte Carlo simulation, earlier considerations (Borneau 1960; Lindquist 1953) that F-tests are robust to variance heterogeneity "regardless of the total sample size and variance ratio" (i.e., ratio of the largest variance to the smallest variance across treatments) when the groups analyzed have equal sample sizes. Given the virtually equal distribution of the number of participants across treatments (including the following), the results should not be biased by possible variance heterogeneity.
Nonetheless, I also test for variance homogeneity. For the dependent variable *quantity of creativity*, variance homogeneity is established (Levene's test of homogeneity 1960, $p > 0.1$) across all factor combination, i.e., low and high feedback frequency, female and male source (and later also female or male feedback recipient, and gender congruence or incongruence); whereas for *quality of creativity*, due to the variable's nature (Blanca et al. 2018), I do not find variance homogeneity (Levene's test of homogeneity). Moreover, in addition to the ANOVA analyses (including the following), I perform Kruskal-Wallis tests, which do not assume variance homogeneity and have been shown to be robust to variance heterogeneity (Harwell et al. 1992). The results (untabulated) are consistently in line with and substantiate the ANOVA results.
[128] Please refer to Figure 4.5 for definitions of the experimental and gender variables.
[129] Please refer to Figure 4.5 for definitions of the experimental and gender variables.

stimulus involve more ideas on average than later responses. A separate single-factor linear regression (untabulated) supports this finding (β= -0.21, p < 0.01).[130]

Analogously, to test H1b and H2b, which predict that feedback frequency and the interaction of feedback frequency and gender source have a positive and negative effect, respectively, on the quality of recipients' creativity, a two-way ANOVA was conducted. The results of the saturated model (see Panel C of Figure 4.9[131]) reveal a significant main effect of feedback frequency on quality of creativity ($F(3, 100) = 13.93$, $p < 0.01$), such that a high feedback frequency (M = 1.18, SD = 0.21) leads to higher quality of creativity than a low feedback frequency (M = 0.39, SD = 0.08). Strikingly, I also find a significant interaction of feedback frequency and gender source ($F(3, 100) = 5.69$, $p = 0.02$), indicating that the effect of feedback frequency is greater in the treatment with female feedback source than in the male feedback source treatment. The main effect of gender source is insignificant ($F(3, 100) = 0.71$, $p = 40.3$). Thus, H1b and H2b are confirmed. Figure 4.11 presents an illustration of the results.

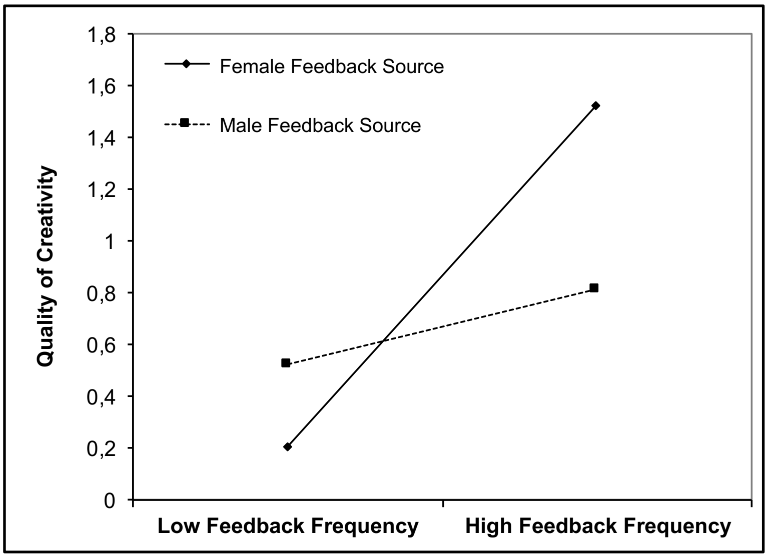


Figure 4.11: Visualization of the moderating Effect of Gender Source on the Relationship between Feedback Frequency and Quality of Creativity

[130] The magnitude of this relationship is similar for the low and high feedback frequency treatments.
[131] Please refer to Figure 4.5 for definitions of the experimental and gender variables.

To ensure robustness of the results, several variations of the analysis are conducted. First, non-submitters are included in the analysis of quality of creativity (i.e., not submitting any ideas is coded as generating zero high-creativity ideas). Nevertheless, the main effect of feedback frequency ($F(3, 185) = 8.65$, $p = 0.4\%$) and the interaction of feedback frequency and gender source ($F(3, 185) = 3.67$, $p = 5.7\%$) maintain to be significant consistently (untabulated). Second, the number of high-creativity ideas per participant is not considered (i.e., the dependent variable is coded as a dummy variable for participant submitting any high-creativity ideas (= “1”) or not (= “0”)). For this dependent variable, the results also agree with the previous results, as the main effect of feedback frequency ($F(3, 104) = 6.42$, $p = 1.3\%$) and the interaction effect of feedback frequency and gender source ($F(3, 104) = 10.01$, $p = 0.2\%$) remain significant (untabulated). Third, a two-way ANCOVA is performed to control for feedback valence, feedback delay, postfeedback interval, and gender recipient (see Panel B of Figure 4.10[132]). The significance and magnitude of both the main effect of feedback frequency ($F(7, 96) = 11.78$, $p = 0.01$) and the interaction effect of feedback frequency and gender source ($F(7, 96) = 5.04$, $p = 0.03$) persist when controlling for these variables, while none of the other variables have a significant effect on the quality of creativity.

Regarding the hypothesized moderating effects of *gender recipient* on the cause-effect relationship of feedback frequency and recipient creativity (H3a/b), the 2 x 2 factorial design represents a quasi-experimental setting, since the assignment of participants to the recipient gender conditions (1 = female feedback recipient, 2 = male feedback recipient) cannot be controlled. The respective cell sizes, means, and

[132] Please refer to Figure 4.5 for definitions of the experimental and gender variables.

standard deviations of the descriptive data are presented in Figure 4.12. In order to account for the quasi-experimental setting, a chi-square test of independence is performed to rule out a statistically unequal distribution of participants. Its results (χ^2(1, N = 189) = 1.96, p > 10%) confirm the visual analysis, rejecting an unequal distribution of participants among treatments. Moreover, Kruskal-Wallis test is conducted and significant differences across the treatments regarding the participants' self-assessment of tasks 1-5 (p = 0.7%), expected valence on tasks 1-5 (p < 0.1%), and achievement striving (p = 0.4%) are detected. These differences can be statistically traced to the gender of participants[133] and can be explained by inherent attributions to gender: While gender research has repeatedly observed a gender confidence gap (e.g., Balafoutas, Kerschbamer, and Sutter 2012; Baier et al. 2018; Dohmen and Falk 2011; Kamas and Preston 2012; van Veldhuizen 2017) that explains differences in self-assessment and expected valence prior to the feedback, the significantly higher level of achievement striving among women may seem surprising at first glance. However, Laher and Croxford (2013), who likewise use the NEO-PI measure, equally observe that women exhibit higher levels of achievement striving. This result is further supported by McCrae et al. (2005), who find that this difference holds for college-age individuals, while it reverses for older adults, and argue that this outcome is due to an increase in the career aspirations among young women. Thus, these differences in cross-sectional data are not surprising, nor should they bias the results as they are inherent in the different genders. Furthermore, the descriptive data of the experimental control variables (i.e., feedback valence, feedback delay, and postfeedback interval) (see Panel A of Figure 4.13[134]) show similar values for all treatments (Kruskal-Wallis test, all p > 10%, two-tailed), so these variables, which are often related to feedback

[133] The results of the Kruskal-Wallis test correspond to gender differences between recipients for the aforementioned cross-sectional variables (t-tests, all p < 1%, two-tailed).
[134] Please refer to Figure 4.5 for definitions of the experimental and gender variables.

frequency in practice and have been shown to impact recipient creativity, do not interfere with the cause-effect relationship between feedback frequency and recipient creativity.

Contrary to my expectations, the respective ANOVAs (see Panel B and Panel C of Figure 4.13[135]) testing H3a and H3b show that the interaction term of feedback frequency and recipient gender is not significant with respect to either quantity of creativity ($F(3, 185) < 0.01$, $p = 0.81$) or quality of creativity ($F(3, 100) = 0.18$, $p = 0.71$). These results hold if experimental control variables (i.e., feedback valence, feedback delay, and postfeedback interval) and gender source are added to the analysis (ANCOVA, untabulated). I find that the effects of feedback frequency on creativity do not differ with respect to the recipient gender and therefore reject hypotheses H3a and H3b.

[135] Please refer to Figure 4.5 for definitions of the experimental and gender variables.

Cross-Sectional Data Summary Statistics – by Feedback Frequency and Gender Recipient

Descriptive Statistics
(Means [Standard Deviations])

Variable [b]	Treatment [a] Low FF + FeRe	Low FF + MRe	High FF + FeRe	High FF + MRe
Age	23.0 [4.2]	24.1 [3.6]	24.1 [3.5]	24.3 [4.1]
Gender recipient	1.00 [0.0]	2.00 [0.0]	1.00 [0.0]	2.00 [0.0]
German as native language	1.12 [0.3]	1.08 [0.3]	1.21 [0.4]	1.06 [0.2]
Undergraduate study	1.04 [0.2]	1.00 [0.0]	1.02 [0.1]	1.00 [0.00]
Employment	0.81 [0.4]	0.78 [0.4]	0.67 [0.5]	0.72 [0.5]
Self-assessment of tasks 1-5	53.5% [14.0%]	61.8% [15.0%]	51.5% [19.5%]	59.3% [18.0%]
Expected valence on tasks 1-5	27.5% [17.5%]	40.5% [19.1%]	25.6% [16.4%]	39.1% [19.4%]
Perceived difficulty of tasks 1-5	3.3 [1.0]	3.8 [0.9]	3.5 [1.0]	3.3 [1.1]
Interest in course	5.7 [1.0]	5.5 [1.1]	5.5 [1.2]	5.6 [1.3]
Achievement striving	5.4 [1.0]	4.8 [1.2]	5.1 [1.1]	4.6 [1.1]
n	57	37	48	47

a. *Low FF* = low feedback frequency; *High FF* = high feedback frequency; *FeRe* = female feedback recipient; *MRe* = male feedback recipient.
b. *Age* = age of participants in years; *Gender recipient* = dummy variable for participants' gender as indicated by each participant in both sessions (1 = female; 2 = male); *German as native language* = dummy variable for participants' self-assessed ability in the German language (1 = maximal value on the 5-point scale; 2 = any other value); *Undergraduate study* = dummy variable for participants' current study program (1 = Bachelor; 2 = Master); *Employment* = dummy variable for whether participants are employed besides their studentry (0 = no; 1 = yes); *Self-assessment of tasks 1-5* = participants were asked to estimate which proportion of tasks 1-5 they answered correctly; *Expected valence on tasks 1-5* = participants were asked to estimate the share of participants they expect to outperform in tasks 1-5 (prior to the feedback); *Perceived difficulty of tasks 1-5* = participants were asked to indicate how difficult they perceived tasks 1-5 (on a 7-point scale, with higher scores expressing higher perceived difficulty); *Interest in course* = participants were asked to indicate their interest in the university course the experiment was integrated into (on a 7-point scale, with higher scores expressing higher perceived interest); *Achievement striving* = participants' personality trait assessed with four items from the respective subscale of the revised NEO Personality Inventory (Costa and McCrae 1992; McCrae and John 1992).

Figure 4.12: Cross-Sectional Data Summary Statistics – by Feedback Frequency and Gender Recipient

Panel A: Experimental Data Summary Statistics – by Feedback Frequency and Gender Recipient

Descriptive Statistics
(Means [Standard Deviations])

	Treatment [a]			
Variable [b]	Low FF + FeRe	Low FF + MRe	High FF + FeRe	High FF + MRe
Feedback valence	45.0% [17.1%]	45.5% [16.8%]	40.9% [15.0%]	42.6% [18.9%]
Submitting participants	57.9% [49.8%]	56.8% [50.2%]	50.0% [50.5%]	55.3% [50.3%]
Feedback delay	209.9 [54.5]	213.4 [52.4]	201.6 [53.6]	211.9 [52.0]
Postfeedback interval	257.3 [60.1]	251.6 [68.5]	244.9 [65.6]	245.8 [72.4]
Quantity of creativity	3.21 [3.5]	2.54 [3.0]	2.52 [3.3]	2.77 [3.4]
Quality of creativity	0.33 [0.5]	0.57 [0.5]	0.50 [0.5]	0.55 [0.5]
n	57	37	48	47

a. *Low FF* = low feedback frequency; *High FF* = high feedback frequency; *FeRe* = female feedback recipient; *MRe* = male feedback recipient.

Panel B: Effect of Feedback Frequency and Gender Recipient on Quantity of Creativity

Factors [a]	pre-dicted	df	mean squares	F	p-value (two-tailed)
Constant		1	1405.551	125.189	0.000
Feedback frequency	+	1	2.487	0.221	0.638
Feedback frequency x Gender recipient	-	1	2.082	0.185	0.667
Gender recipient		1	9.661	0.860	0.355
Error		185	11.227		

a. DV = Quantity of creativity; n = 189.

Panel C: Effect of Feedback Frequency and Gender Recipient on Quality of Creativity

Factors [a]	pre-dicted	df	mean squares	F	p-value (two-tailed)
Constant		1	63.520	49.647	0.000
Feedback frequency	+	1	15.190	11.873	0.001
Feedback frequency x Gender recipient	-	1	0.180	0.140	0.709
Gender recipient		1	0.087	0.068	0.795
Error		100	1.279		

a. DV = Quality of creativity; n = 104.

Figure 4.13: Effect of Feedback Frequency and Gender Recipient on Creativity with descriptive Experimental Data

Before testing hypotheses H4a/b, which predict that the interaction of feedback frequency and *gender congruence* influence recipient creativity, I test whether significant differences emerge in the descriptive cross-sectional data in the novel 2 x 2 grouping of participants for gender congruence (1 = gender congruence, 2 = gender incongruence) and feedback frequency. Since gender congruence cannot be actively manipulated, but is only indirectly influenced by the manipulation of the source's gender, this analysis bases on a quasi-experimental approach. However, significant differences (Kruskal-Wallis test, all p > 10%, two-tailed) between the gender congruence and feedback frequency treatments that may interfere with the results are not detected (see Figure 4.14). In addition, the distribution of participants (χ^2(1, N = 189) = 0.35, p > 10%) and the experimental control variables (Kruskal-Wallis test, all p > 10%, two-tailed) also indicate a suitable basis for further analyses (see Panel A of Figure 4.15[136]).

An ANOVA is conducted in a saturated model (see Panel B of Figure 4.15[137]) to test H4a, which predicts that gender congruence moderates the relationship between feedback frequency and quantity of creativity. Surprisingly, the results show that the interaction term of feedback frequency and gender congruence (F(3, 185) = 1.73, p = 0.19) is not significant, indicating that there are no differences across the (quasi)experimental groups regarding the quantity of creativity. Similarly, I test H4b (see Panel C of Figure 4.15[138]), which predicts that gender congruence moderates the relationship between feedback frequency and quality of creativity, and also find this interaction to be non-significant (F(3, 100) = 0.22, p = 0.64). The inclusion of experimental control variables (i.e., feedback delay, postfeedback interval, and feedback valence) and the remaining gender variables (i.e., gender source and gender

[136] Please refer to Figure 4.5 for definitions of the experimental and gender variables.
[137] Please refer to Figure 4.5 for definitions of the experimental and gender variables.
[138] Please refer to Figure 4.5 for definitions of the experimental and gender variables.

recipient) in an ANCOVA does not alter the results (untabulated). Since gender research does not homogeneously subsume the effects of any combination of source and recipient gender under gender congruence, I additionally run two 2 x 2 x 2 ANOVAs in a saturated model including feedback frequency, recipient gender, and source gender. However, the analyses confirm the previous results, as there is no significant main effect and interaction effect on creativity (quantity and quality), except for the interaction effect of feedback frequency and source gender and the main effect of feedback frequency on quality of creativity, which remain significant (untabulated). Thus, hypotheses H4a and H4b are rejected.

Cross-Sectional Data Summary Statistics – by Feedback Frequency and Gender Congruence

Descriptive Statistics
(Means [Standard Deviations])

	Treatment [a]			
Variable [b]	Low FF + Con	Low FF + InCon	High FF + Con	High FF + InCon
Age	23.9 [5.0]	22.9 [2.4]	24.3 [4.2]	24.1 [3.4]
Gender recipient	1.40 [0.5]	1.39 [0.5]	1.50 [0.5]	1.49 [0.5]
German as native language	1.10 [0.3]	1.11 [0.3]	1.14 [0.3]	1.14 [0.3]
Undergraduate study	1.02 [0.1]	1.02 [0.1]	1.00 [0.0]	1.02 [0.1]
Employment	0.82 [0.4]	0.77 [0.4]	0.66 [0.5]	0.73 [0.5]
Self-assessment of tasks 1-5	55.8% [13.1%]	57.9% [16.9%]	57.4% [20.1%]	53.6% [18.1%]
Expected valence on tasks 1-5	31.0% [17.4%]	34.4% [21.0%]	31.4% [19.8%]	33.0% [18.7%]
Perceived difficulty of tasks 1-5	3.4 [1.2]	3.4 [1.3]	3.8 [1.3]	3.7 [1.3]
Interest in course	5.6 [1.1]	5.6 [1.1]	5.6 [1.2]	5.5 [1.3]
Achievement striving	5.3 [1.2]	5.0 [1.0]	5.0 [1.0]	4.8 [1.3]
n	50	44	44	51

a. *Low FF* = low feedback frequency; *High FF* = high feedback frequency; *Con* = gender congruence; *InCon* = gender incongruence.
b. *Age* = age of participants in years; *Gender recipient* = dummy variable for participants' gender as indicated by each participant in both sessions (1 = female; 2 = male); *German as native language* = dummy variable for participants' self-assessed ability in the German language (1 = maximal value on the 5-point scale; 2 = any other value); *Undergraduate study* = dummy variable for participants' current study program (1 = Bachelor; 2 = Master); *Employment* = dummy variable for whether participants are employed besides their studentry (0 = no; 1 = yes); *Self-assessment of tasks 1-5* = participants were asked to estimate which proportion of tasks 1-5 they answered correctly; *Expected valence on tasks 1-5* = participants were asked to estimate the share of participants they expect to outperform in tasks 1-5 (prior to the feedback); *Perceived difficulty of tasks 1-5* = participants were asked to indicate how difficult they perceived tasks 1-5 (on a 7-point scale, with higher scores expressing higher perceived difficulty); *Interest in course* = participants were asked to indicate their interest in the university course the experiment was integrated into (on a 7-point scale, with higher scores expressing higher perceived interest); *Achievement striving* = participants' personality trait assessed with four items from the respective subscale of the revised NEO Personality Inventory (Costa and McCrae 1992; McCrae and John 1992).

Figure 4.14: Cross-Sectional Data Summary Statistics – by Feedback Frequency and Gender Congruence

Panel A: Experimental Data Summary Statistics – by Feedback Frequency and Gender Congruence

Descriptive Statistics
(Means [Standard Deviations])

Variable [b]	Treatment [a] Low FF + Con	Low FF + InCon	High FF + Con	High FF + InCon
Feedback valence	44.8% [16.4%]	45.7% [17.5%]	43.6% [18.4%]	40.1% [15.7%]
Submitting participants	52.0% [50.5%]	63.6% [48.7%]	59.1% [49.7%]	47.1% [50.4%]
Feedback delay	204.2 [51.8]	219.3 [54.8]	199.4 [53.0]	213.0 [52.4]
Postfeedback interval	262.1 [65.3]	247.1 [60.6]	243.3 [74.4]	247.1 [64.1]
Quantity of creativity	2.54 [3.2]	3.41 [3.5]	2.86 [3.4]	2.45 [3.4]
Quality of creativity	0.38 [0.7]	0.39 [0.5]	1.08 [1.4]	1.29 [1.7]
n	50	44	44	51

a. *Low FF* = low feedback frequency; *High FF* = high feedback frequency; *Con* = gender congruence; *InCon* = gender incongruence.

Panel B: Effect of Feedback Frequency and Gender Congruence on Quantity of Creativity

Factors [a]	pre-dicted	df	mean squares	F	p-value (two-tailed)
Constant		1	1491.507	133.501	0.000
Feedback frequency	+	1	4.732	0.424	0.516
Feedback frequency x Gender congruence	-	1	19.314	1.729	0.190
Gender congruence		1	2.449	0.219	0.640
Error		185	11.172		

a. DV = Quantity of creativity; n = 189.

Panel C: Effect of Feedback Frequency and Gender Congruence on Quality of Creativity

Factors [a]	pre-dicted	df	mean squares	F	p-value (two-tailed)
Constant		1	64.144	50.255	0.000
Feedback frequency	+	1	16.407	12.854	0.001
Feedback frequency x Gender congruence	-	1	0.276	0.217	0.643
Gender congruence		1	0.322	0.252	0.616
Error		100	1.276		

a. DV = Quality of creativity; n = 104.

Figure 4.15: Effect of Feedback Frequency and Gender Congruence on Creativity with descriptive Experimental Data

4.4.2 Supplemental Analysis

While the results confirm the hypothesized causality of feedback frequency on employee creativity in terms of the quality of creativity, the post-experimental questionnaire allows a deeper understanding of the underlying mechanisms that presumably drive individuals' internal creative processes. I build conceptually on the feedback stimulus and source characteristics as defined by Ilgen, Fisher, and Taylor (1979) and the corresponding operationalizations of Kinicki et al. (2004) and added feedback frequency and gender specific questions to the post-experimental questionnaire. Thus, I am able to determine the implications of feedback frequency and the gender of the feedback source, which were manipulated in the field experiment, on recipients' perception of feedback.

4.4.2.1 Insights into how Feedback Frequency works

Following Fedor and Buckley's (1987) line of reasoning that feedback recipients value frequent feedback as a costly organizational resource that is provided to them by the feedback source and helps them improve their performance, the results show that participants in the high feedback frequency treatment perceive the source's effort in the feedback provision as significantly higher (M = 2.44) than participants in the low feedback frequency condition (M = 2.07; difference p = 3.1%). According to the norm of reciprocity (Gouldner 1960) and social exchange theory (Bateman and Organ 1983), feedback recipients are morally obligated to recompense efforts of their feedback source that benefit the recipient, which is reflected in the difference in participants' indication of how much the feedback motivated them to invest additional time in task completion (M = 4.83 for high feedback frequency, M = 4.52 for low feedback frequency) that strives to significance (difference p = 7.5%). Thus, higher levels of feedback frequency should lead to a propensity to work harder, for instance, on generating novel ideas that have an exceptional level of creative quality.

Ilgen, Fisher, and Taylor (1979) describe that the recipient's perception of the feedback stimulus is determined by its frequency. Particularly in the setting of the experimental field study, where the feedback stimulus can easily be overlaid by impressions and stimuli in the participants' everyday life, the disruptive nature of the feedback, which removes the recipients from their daily routine and attracts their attention, influences how strongly (if at all) the feedback stimulus is perceived and processed. While it stands to reason that a single concentrated (i.e., less frequent) feedback containing information on all first-round tasks would be perceived as more salient disruptive event (M = 1.72), I find conversely that disseminating the same total amount of information over multiple days (i.e., more frequent feedback) is more likely to be perceived as an disruption of the daily routine (M = 2.18; difference p = 1.3%) and conclude that the feedback stimulus is more salient if it occurs more frequent, which is more likely to elicit a (more) creative response.

Furthermore, participants in the high feedback frequency treatment perceive the feedback to be more specific (M = 2.73) than participants in the low feedback frequency treatment (M = 2.39; difference p = 1.3%). This interrelation between feedback frequency and feedback specificity has already been detected in other studies (e.g., Kinicki et al. 2004; Larson 1986). However, in contrast to previous studies, the level of (objective) feedback specificity, being defined as "the level of information presented in feedback messages" (Goodman, Wood, and Hendrickx 2004, p. 248), was held constant. Thus, the question arises as to why the results are nonetheless consistent.

Cognitive research has often referred to "the magic number seven plus or minus two" (Miller 1956) as the limit of information that an individual's working memory can handle and subsequently process. While empirical studies confirm the notion of limited working memory capacity, more recent results suggest that the limit might be rather four or less items (Cowan 1998, 2001; Luck and Vogel 1997; Luck 1998; Ma, Husain, and Bays 2014). If individuals are confronted with more items than they can cognitively process at once, they experience information (or cognitive) overload. Miller (1960) cites approximation as a response to information overload, processing information with less precision, and "cutting categories of discrimination" (Katz and Kahn 1978, p. 451). For feedback recipients suffering from information overload, this could mean that the provided feedback information for multiple facets of their performance is not processed individually, but as an (subjectively) approximated aggregate value, which in turn is perceived (particularly in the retrospective) as less specific since the feedback was memorized and processed in aggregated form.

This approximation effect could be amplified by central tendency bias, i.e., "stimuli with values greater than the category's average tend to be underestimated and stimuli with values less than the average are overestimated" (Allred et al. 2016, p. 1825), as individuals have been shown to exhibit greater central tendency bias when suffering from cognitive overload (Allred et al. 2016). If multiple feedback valence values tend to be perceived as oscillations around the median valence value, since both positive and negative feedback are perceived as less extreme due to the central tendency bias, aggregating this feedback information becomes reasonable, since breaking this information down into more specific feedback information leads to smaller (if any) gain in knowledge. In other words: If all specific feedback information is (about) the same, there is no need to use it instead of an overall feedback value (i.e., an approximation), which favors the perception of less specific feedback.[139]

[139] I use the Kruskal Wallis test to find differences in individuals' most positive (negative) single feedback valence (i.e., the best (worst) feedback on any of the tasks 1-5) and in individuals' range of feedback

Given the setting of the field experiment (e.g., participants are not in an artificial, low-stimulus laboratory environment) and the number feedback information provided simultaneously in the low feedback frequency condition, these participants are more likely to experience information overload than when feedback is distributed in smaller portions over a longer time period (i.e., high feedback frequency treatment). This distributional effect of feedback frequency (if not confounded with feedback quantity) therefore allows me to reveal a novel reason for feedback frequency effectiveness: Frequent feedback enables recipients to receive information in units that they perceive as more precise[140] and thus can be better digested (Goodman, Wood, and Hendrickx 2004), which contributes to the information value of feedback, that depends on recipients' ability to "transform the feedback message to units that are meaningful to them" (Ilgen, Fisher, and Taylor 1979, p. 351).

While previous studies on feedback similarly build on considerations of cognitive resources, they conclude that feedback frequency is positively associated with information overload (e.g., Lam et al. 2011; Holderness, Olsen, and Thornock 2020). However, unlike these laboratory experiments that take place within a narrow time span, which condenses the effect of information overload, the total amount of feedback is held constant in this study and therefore it does not confound the manipulation of feedback frequency with feedback quantity. An overview of the differences (means and significances) and the variable measurements presented are provided in Panel A of Figure 4.16.

---

valence (i.e., best feedback minus worst feedback) between treatments. The results show that none of the three mean scores were significantly different between treatments ($p > 0.1$).

[140] While Ilgen, Fisher, and Taylor (1979) generally emphasize the importance of feedback perception, they explicitly point out that feedback specificity is subjective, so that an individual recipient "receives specific feedback from his or her point of view" (p. 365 f.).

Panel A: Differences in Perception of Feedback Provision by Feedback Frequency

	mean [a]		t-Test	
Item [b]	Low FF	High FF	df	p-value (one-tailed)
Perceived effort of source	5.07	**5.39**	187	0.046
Induced motivation	4.52	**4.83**	187	0.075
Feedback disruptiveness	1.72	**2.18**	187	0.013
Feedback specificity	2.39	**2.73**	187	0.014

Note: Higher values at $p < 0.1$ are marked in bold.
a. *Low FF* = low feedback frequency; *High FF* = high feedback frequency.
b. *Perceived effort of source* = "Do you feel that your tutor made an effort?" (on a 7-point scale, with higher values expressing higher perceived effort of the feedback source); *Induced motivation* = "Did you put extra time into task completion because of the feedback given?" (on a 7-point scale, with higher values expressing higher motivating effect); *Feedback disruptiveness* = "Did you perceive the feedback as an interruption to your daily routine?" (on a 7-point scale, with higher values expressing higher perceived interruption); *Feedback specificity* = "How specific was the feedback?" (adopted from Kinicki et al. 2004, on a 5-point scale, with higher values expressing higher specificity).

Panel B: Differences in Perception of Feedback Provision by Gender Source

	mean [a]		t-Test	
Item [b]	FeSou	MSou	df	p-value (one-tailed)
Competence of source	5.50	5.37	187	0.221
Trustworthiness of source	3.89	3.87	187	0.441
Psychological closeness	**2.44**	2.07	187	0.031
Perceived capability (recipient self-assessment)	**4.74**	4.34	187	0.021
Induced motivation	**4.84**	4.52	187	0.066

Note: Higher values at $p < 0.1$ are marked in bold.
a. *FeSou* = female feedback source; *MSou* = male feedback source.
b. *Competence of source* = "I consider my supervisor to be competent" (adopted from Kinicki et al. 2004; on a 7-point scale, with higher values expressing higher competence of the feedback source); *Trustworthiness of source* = "I can trust what my supervisor says" (adopted from Kinicki et al. 2004; on a 7-point scale, with higher values expressing higher trustworthiness of the feedback source); *Psychological closeness* = "To what extent do you feel psychologically close to your supervisor?" (adapted from Gino and Galinsky 2012; on a 7-point scale, with higher values expressing higher psychological closeness); *Induced motivation* = "Did you put extra time into task completion because of the feedback given?" (on a 7-point scale, with higher values expressing higher motivating effect); *Perceived capability* = "I can't do a good job in this course with my present skills and abilities" (adopted from Kinicki et al. 2004; on a reverse-coded 7-point scale, with higher values expressing higher perceived capability).

Figure 4.16: Differences in the Perception of Feedback Provision arising from the Experimental Manipulations

4.4.2.2 Differences in Perception of Feedback Source due to its Gender

As previously noted, gender role stereotypes influence the perception of leadership behaviors and, consequently, also the perception of feedback provided from an interpersonal source (cf. Eagly and Karau 1991; Eagly and Wood 2012; Lanaj and Hollenbeck 2015). While the agentic attributes projected onto a male supervisor should address source credibility, comprising of its competence and trustworthiness (Ilgen, Fisher, and Taylor 1979), the results, however, demonstrate that participants ascribe credibility to the female feedback source in a similar manner, as there are no significant differences in the attribution of source competence (M = 5.50 for female feedback source, M = 5.37 for male feedback source; difference $p > 10\%$) and source trustworthiness (M = 3.89, M = 3.87, respectively; difference $p > 10\%$) between the two genders. Since, according to feedback literature (Ilgen, Fisher, and Taylor 1979; Kluger and DeNisi 1996; Taylor, Fisher and Ilgen 1984), the credibility of the feedback source determines the (creative) reactions of the feedback recipient, this result could explain to some extent why I am unable to find a creative boost from male feedback source compared to female feedback source.

Complementarily, I find that female feedback elicits relatively more creativity-enhancing effects that can be attributed to the communal female stereotype. Particularly, female feedback evokes perceptions of psychological closeness (M = 2.44) – more than male feedback (M = 2.07; difference $p = 3.1\%$), suggesting that recipients of female feedback are more likely to perceive an environment in which the interpersonal relationship, rather than “faceless” performance, are more pronounced and that they are generally more responsive to (female) feedback (Greller and Herold 1975). However, this effect is not driven by female recipients, at odds with considerations of social group membership in the psychological literature (Sturm et al. 2014) and gender research, which posits that female-to-female interactions convey the greatest social sentiment (Carli 2010).

Surprisingly, female feedback leads recipients to rate their own capability higher (M = 4.74) than male feedback (M = 4.34; difference p = 2.1%), despite both feedback sources are perceived as similarly credible. This effect is further substantiated by the fact that both participant groups assess their performance similarly prior to feedback (M = 56.5% for female source, M = 55.7% for male source; difference p > 10%, see Figure 4.7) and receive feedback with similar valence (M = 43.0%, M = 43.9%, respectively; difference p > 10%, see Figure 4.9), so there is no objective reason for this disparity in perceived capability after feedback other than the perception of the feedback source. Importantly, since perceived capability is "central to what has been termed intrinsic motivation" (Ilgen, Fisher, and Taylor 1979, p. 362), feedback from a female source motivates recipients to engage in creative activities more strongly than from a male source. Again, this is reflected in the difference in participants' indication of how much the feedback motivated them to invest additional time in task completion (M = 4.84 for female feedback source, M = 4.52 for male feedback source), which strives toward significance (difference p = 6.6%).

4.5 Discussion

This experimental field study set out to explore the cause-effect relationship between feedback frequency and creativity. In a world in turmoil and increased competition, where creativity is key, firms accelerate their feedback cycles to avoid employees not recognizing the need to come up with creative ideas. This permeates throughout the organization. Thus, the more frequently feedback is provided from supervisor to employee, the more frequently an employee can derive a meaningful demand for developing new ideas that realign the effectiveness of his or her activities. In addition, the temporal interplay of feedback information allows recipients to verify whether their original ideas are working and still address current challenges, and thus to continuously update and improve the quality of their creative ideas. This study therefore focuses on the frequency of feedback, which has only recently regained paramount importance as a decisive feedback characteristic (e.g., Casas-Arce, Lourenco, and Martinez-Jerez 2017; Holderness, Olsen, and Thornock 2020; Lam et al. 2011).

I find that the frequency of feedback is an effective instrument to elevate the quality of employee creativity. Thereby, I not only extend feedback research by revealing the beneficial effects of feedback frequency for creativity, but also contribute to management accounting research by adding feedback frequency to the short list of known drivers of high-creativity idea generation (Kachelmeier, Reichert, and Williamson 2008; Kachelmeier, Wang, and Williamson 2019). It should be noted that these results do not hold for the quantity of idea generation. Nonetheless, since quality responses are often not initial and directly incentivizing high-creativity idea generation has been proven ineffective (Kachelmeier, Reichert, and Williamson 2008; Kachelmeier, Wang, and Williamson 2019), managers are unlikely to gain their own experiential insights and have little guidance on how to promote the generation of ideas that are most valuable to organization’s ability to innovate, gain competitive advantage, and thus to survive. The findings help managers who want to inspire their employees to be creative and contribute to key organizational success factors by demonstrating that a high frequency of feedback should be given to foster employee creativity.

The supplemental analysis suggests that the positive effects of feedback frequency on employee creativity are due to differences in the perception of feedback arising from its frequency (Ilgen, Fisher, and Taylor 1979). First, frequent feedback increases employees’ intrinsic motivation, as feedback recipients associate the frequency of feedback to the source’s effort. Since these efforts benefit the recipient, frequent feedback induces mechanisms of reciprocity (Fedor and Buckley 1987; Gouldner 1960), the translation of which into greater commitment to working on creative ideas is further supported by survey data. This reasoning is substantiated by the fact that participants were not offered any financial incentives, submission of creative ideas was entirely voluntary, and participants did not face any negative consequences to their final course grade for generating no, few, or poor ideas. Second, participants indicate that the feedback stimulus is more salient when feedback is provided more frequently, which is striking given the stimulus-rich field setting of the experimental study and the relatively low informational value of a feedback message in the high feedback frequency condition. Third, a distributional effect related to feedback frequency is identified that provides recipients with informational units they can digest better and therefore perceive more frequent feedback to be more specific.

While several feedback dimensions related to feedback frequency have been shown to influence recipient creativity, the research design allows me to rule out alternative explanations, as the findings are robust to feedback valence and feedback timing, and as feedback quantity was controlled for. However, independent of the effectiveness of feedback frequency, the results also demonstrate that feedback timing influences recipient's creativity. More specifically, individuals generate more ideas the shorter the postfeedback interval, while there is no effect of feedback timing on the quality of creative ideas. This finding contradicts the notion of creative incubation, which suggests beneficial effects of interruption-free time on the generation of novel ideas (Dodds, Ward, and Smith 2012; Gilhooly 2016; Kachelmeier, Wang, and Williamson 2019). Given the complex nature of a feedback stimulus, which depends on the cognitive-emotional relationship with the feedback provider and requires recipients to recall past behaviors that may be overridden by disruptive events or behaviors, there is a constant threat that facets of the feedback will become increasingly diluted – cognitively and emotionally – over time. Thus, it is not surprising that the results suggest that creative incubation is less pronounced in the context of feedback than when employees receive a more trivial stimulus, such as an instruction ("generate creative ideas for…"), which is more easily retrievable.

As feedback and creativity are highly individual-specific processes, and we know from a number of publications that gender influences the way of perceiving (feedback source) and processing (feedback recipient) information (e.g., Carli 2010; Eagly and Karau 1991), I was interested in better understanding the tension that arise from gender and gender (in)congruence. In contrast to extant gender studies showing differential workplace behavior between men and women, I fail to find feedback recipient gender to make a difference. Also, gender (in)congruence does not interfere with the creative effects of feedback, which opposes suggestions from feedback and psychological research (e.g., Sturm et al. 2014). However, consistent with gender role theory, I find support for my hypothesis that feedback source's gender impacts on creativity.

More specifically, I find that the gender of the feedback source (women) significantly influences the effectiveness of feedback frequency for the quality of employee creativity, regardless of the gender of the recipient. It is remarkable that although the feedback literature emphasizes that "each feedback comes from some source, and [has] its proven effect on individuals' reactions to performance feedback" (Lechermeier and Fassnacht 2018, p. 146), it largely neglects the gender of the feedback source as a crucial source characteristics (see reviews from Alvero, Bucklin, and Austin 2001 and Lechermeier and Fassnacht 2018, which explicitly observe the implications of feedback source and feedback source characteristics), even though "these [gender] roles are salient to the extent that other social roles, particularly family and employment roles, are relatively unimportant." (Eagly and Karau 1991, p. 686). Thus, this study not only addresses this particular gap in the feedback literature, but its findings should extend to other feedback dyads, such as feedback from colleagues (Buckingham and Goodall 2019), (upward) employee feedback (Kim and Kim 2020; Jhun, Bae, and Rhee 2012), or customer feedback (Nasr, Burton, and Gruber 2018), being also influenced by their respective genders.

This result gains importance against the backdrop of an increasing number of women in leadership positions (the German "Gesetz für die gleichberechtigte Teilhabe von Frauen und Männern an Führungspositionen in der Privatwirtschaft und im öffentlichen Dienst" (FüPoG) entered into force in 2015, and the FüPoG II is currently in the process of being passed[141]; Bobe and Kober 2020; Lai et al. 2017). Even more, considering that the participants in this study belong to a younger generation that is likely to be less susceptible to gender stereotypes than the current workforce (see a recent study by the National Centre of Social Research[142]), and the tendency of women to consider themselves as less suited and effective for leadership roles (Paustian-Underdahl, Walker, and Woehr 2014), this strengthens their case for occupying such a role, especially if their job description involves fostering employee creativity.

Since the provision of feedback is an inherent duty of leaders and a powerful lever for influencing employee outcomes (Shea and Howell 1999), my findings also add to the leadership literature in a twofold manner. Fundamentally, the findings nest in an ongoing discussion of gender differences in leadership effectiveness that has

---

[141] See https://www.bmfsfj.de/bmfsfj/service/gesetze/zweites-fuehrungspositionengesetz-fuepog-2-164226 [retrieved 07-21].

[142] See https://www.bsa.natcen.ac.uk/latest-report/british-social-attitudes-30/gender-roles/a-generational-shift-in-attitudes.aspx [retrieved 07-21], which should apply similarly to all Western societies.

evolved from a more descriptive approach of gender differences in leadership styles, as "studies should not be asking whether there is a perceived gender difference in leadership but rather when and why there may be gender differences in perceived leadership effectiveness" (Paustian-Underdahl, Walker, and Woehr 2014, p. 1129). Consistent with results of a recent meta-analysis (Paustian-Underdahl, Walker, and Woehr 2014), I find that female leaders are more effective[143] when they provide feedback frequently. Because many studies of leadership effectiveness are survey-based, it is important to note that this research approach objectively compared the provision of different feedback frequencies as a particular leadership behavior and therefore did not allow for self-selection biases, further accentuating the beneficial effects of female leadership.

More specifically, gaps in leadership research on creativity are filled by demonstrating the positive effect of supervisor feedback frequency on employee creativity and the moderating effect of leader gender on this relationship. This study thus responds to Hughes et al. (2018, p. 564), whose comprehensive review on the effects of leadership on creativity and innovation urges future research to "move beyond the current focus on leader styles to explore the effects of leader characteristics [and] behaviors". In particular, knowledge about the influence of leader gender in the context of creativity is limited, as previous research has been conducted in a male-gendered setting (Reuvers et al. 2008[144]) or focused on a male-favoring leadership style (Wang et al. 2013). Therefore, it is not surprising that previous results generally favor male leaders in enhancing employee creativity. In contrast, the findings

---

[143] While their meta-analysis shows mixed results at an aggregated level, the results clearly suggest women to be more effective when effectiveness is evaluated by others. The discrepancy between self-assessment and external evaluation can be attributed to the gender-stereotypical underestimation of women's (leadership) performance and their generally lower self-esteem (Paustian-Underdahl, Walker, and Woehr 2014).

[144] Actually, Reuvers et al. (2008) examine the interactive effects of transformational leadership and gender on innovative work behaviors.

suggest that frequent feedback from female leaders is more effective in increasing employee creativity than from male leaders, which may indicate the need to reevaluate previous studies of leadership effectiveness on creativity.

The supplemental survey-based analysis of feedback perceptions provides key insights into why a female feedback source amplifies the positive effects of feedback frequency on employee creativity. First, to my surprise and contrary to male stereotypes, I find no differences in the perceptions of source credibility, consisting of source capability and source trustworthiness, between the two (female and male) feedback sources. Since the feedback literature (Ilgen, Fisher, and Taylor 1979; Kluger and DeNisi 1996; Taylor, Fisher, and Ilgen 1984) has pointed out that this attribution is an important characteristic of the feedback source in determining whether employees respond favorably to the feedback stimulus, this may explain to some extent why I do not find that a male feedback source leads to more creative responses.

Second, consistent with suggestions from gender role theory, the data indicate that female feedback is more likely to create a work environment in which personal relationships are paramount. Particularly in the context of creativity, which more than other employee efforts evoke insecurities, negative emotions, and risks of failing, a (female) supervisor who is perceived as psychologically close and sympathetic encourages employees to chase the timid fawn of creativity.

Third, female feedback induces a relatively stronger perceptions of one’s own capability, which is an important contributor to intrinsic motivation (Ilgen, Fisher, and Taylor 1979). Hence, I complement recent research that has shown that perceived capability is not affected by recipient gender or feedback design (Biernat et al. 2020) by demonstrating that it depends on the gender of the feedback source. The descriptive data substantiate that the motivational effect of female feedback source is greater than that of a male feedback source.

In this study, the provided feedback referred to individual performance compared to participants (RPI), consisting of roughly equal proportions of men and women. Gender research has shown that in environments that allow performance comparison with other participants, gender-specific responses depend strongly on group composition: Under such conditions, women not only shy away (Niederle and Vesterlund 2007, 2011) and perform worse than men (Gneezy and Rustichini 2004), but do so particularly when they have to compete against men (Gneezy, Niederle, and Rustichini 2003; Kuhnen and Tymula 2012). Accordingly, female participants may avoid the competitive environment by submitting few or no creative ideas, which is detrimental to their creative performance, especially if they are kept aware of the competitive environment through frequent feedback (recall that each feedback message includes a comparison of one’s performance with that of the other participants). This provides a possible explanation for why no influence of recipient gender on the effectiveness of feedback frequency is found, even though women are generally thought to be more receptive to feedback.

4.6 Managerial Implications

This study provides useful applications for practice. While managers may refrain from providing feedback more frequently due to their limited resources and formal authority (De Stobbeleir, Ashford, and Buyens 2011), this should be especially the case if they strive to foster employee creativity, as previous literature points to negative consequences resulting from evaluations when employees are engaged in creative activities (e.g., Sagiv et al. 2010; Shalley, Zhou, and Oldham 2004). The results update this commonly held view and offer guidance for managers with limited time resources and heterogeneous subordinates. First, speeding up feedback cycles and providing feedback frequently stimulates employees to generate more creative ideas of high quality. Thus, frequent managerial feedback helps organizations to improve their innovative capabilities as, only a fraction (i.e., the best) of all ideas can be implemented within the organization. This particularly holds for female feedback sources, who are stereotypically denied a natural leadership role. Companies should therefore empower female managers and give them a voice – to provide feedback to employees.

Second, the research design defies the notion that more frequent feedback must mean more feedback quantity, and therefore offers an efficient solution for managers to increase their feedback frequency. When managers increase the frequency of their feedback, it is quite sufficient to break the same amount of feedback information into more "digestible" units and present them over multiple points in time to encourage employee creativity, rather than gathering more information about their employees or doing so more frequently to create an "all-encompassing" picture of employee performance with each feedback exchange. From a strategic point of view, particularly in these days of explosion-like increases in available data, such as from Big Data or real-time data sources, this insight advises not to inflate feedback systems, but to slim them down in line with requirements and instead to promote and maintain frequent performance communication.

Third, surprisingly, the results show that recipient gender and feedback valence do not condition the effectiveness of feedback frequency on creativity. Thus, managers should neither shrink from providing frequent feedback if it is negative, as it does not harm employee creativity, nor exclude lower-performing employees from these creative stimuli, as they are equally capable of generating high-creativity ideas. In addition, managers should also not forgo frequent feedback for fear of intimidating female employees (as a male manager) or that the feedback effect will fizzle out on male employees (as a female manager), as gender theory suggests.

Lastly, my findings suggest that managers should not expect creative incubation to increase the quantity of employee creativity, but should give employees the freedom to voice their creative ideas sooner rather than later after receiving feedback. Accordingly, typical inflexible and routine communication sessions (e.g., “We’ll talk about your ideas at our next weekly meeting”) are not conductive to employee creativity.

4.7 Limitations and Future Research

The study involves some limitations that should be addressed in future research. First of all, the results are subject to the usual limitations found in experimental settings. While the nature of the field experiment, unlike laboratory studies, per se allowed for a stimulus-rich environment for the participants, the same is not true with respect to its feedback-richness (Kinicki et al. 2004), as external performance evaluation outside of the experimental setting was rare due to the (digital) course setting. Nonetheless, this should contribute to external validity of the experimental setting, as early suggestions that “in most work settings feedback tends to be much too infrequent” (Ilgen, Fisher, and Taylor 1979, p. 367) should apply more than ever, given the need for more frequent exchanges due to increasingly rapidly changing work environments and the burdensome nature of providing interpersonal feedback. In addition, the feedback design choice adheres to the advice for motivating feedback to “add an increment of information to the recipient over and above the information he or she already has” (Ilgen, Fisher, and Taylor 1979, p. 363). This may explain why firms that have implemented live-time reporting systems are not associated with bursts of creative ideas, and cautions against blindly increasing feedback frequency to raise employees’ (creative) performance. Consequently, further research is needed on the limits of the creative effects of feedback frequency.

Second, given that feedback responses depend on the interaction of multiple feedback dimensions (Ilgen, Fisher, and Taylor 1979), altering other facets of the feedback will provide further assurance about the robustness of the findings regarding the creative implications of feedback frequency. For example, individuals who belong to a negatively stereotyped social group (e.g., women) devalue their feedback because they implicitly expect their 'stigma' will be taken into account when they are evaluated by lower standards for the objective performance level. This effect, referred to as the attributional ambiguity account (Biernat and Danaher 2012), may extend my findings regarding the effect of recipient gender when subjective feedback is provided.

In addition, relative performance information (RPI) was used in this study to stimulate creative responses. Gender research has shown that women perform relatively worse in competitive environments (e.g., Gneezy and Rustichini 2004), especially if men are part of the reference group (Kuhnen and Tymula 2012). Hence, research on the effectiveness of feedback frequency on creativity when absolute feedback is provided or when group composition is varied would complement my results.

Lastly, although the feedback was not provided in person and was largely standardized, the results are only indicative of impersonal sources of feedback. Future research could pick up on this idea as technological advancements and more variety enter into the provision of feedback. For example, designing automated anonymous digital feedback sources, interactive AI-driven feedback systems, or even simpler performance measurement systems might benefit from giving them a feminine appearance (e.g., voice or name), as is already the case with well-known assistants such as Alexa (Amazon), Cortana (Microsoft), Google Assistant (Google), or Siri (Apple) (Abercrombie et al. 2021; Loideain and Adams 2020).

5. Study 3: Feedback as a Management Control and the Influence of Control Use

5.1 Introduction

While the previous two studies in this dissertation address feedback frequency from the perspectives of a dynamic S-O-R framework and feedback theory, respectively, Study 3 provides novel insights into the mechanisms of feedback frequency and its use in practice by examining feedback frequency from the viewpoint of management control systems. In doing so, I deal with a core element of controlling as “feedback is central to the discussion of systems design and implementation in the management accounting literature, especially in the context of management control systems” (Luckett and Eggleton 1991, p. 371). In particular, Luckett and Eggleton (1991) put emphasis on the issue of “optimum feedback frequency” and the ambiguous behavioral effects of feedback frequency. Consistent with definitions of management controls[145] (Grabner and Speckbacher 2016; Tessier and Otley 2012), this study therefore considers the frequency of feedback as a particular management control design choice.

Complementing the findings of the two former studies, the objectives of the present study are twofold. First, I introduce control use (Tessier and Otley 2012) as a moderator for the effectiveness of feedback frequency on employee creativity. Second, using a contingency approach (Grabner 2014; Grabner and Speckbacher 2016), I aim to gain insights into how managers (i.e., the feedback source) respond to creative requirements on their employees (i.e., feedback recipients) in terms of the frequency of providing feedback and the use of feedback.

[145] For instance, refer to Grabner and Speckbacher’s (2016, p. 31) definition of controls as “any process used by managers to direct employee attention and influence employee behavior in ways that increase the probability of achieving organizational goals”.

Since “in management theory and practice, creativity is widely seen as an important source of competitive advantage and business success” (Speckbacher 2017, p. 49), several strands of research have recognized creativity as an important organizational goal, such as the leadership literature (e.g., Hughes et al. 2018) or the service literature (e.g., Coelho and Augusto 2010). Although controls “increase the probability of achieving organizational goals” (Grabner and Speckbacher 2016, p. 31), even management control research has only recently begun to acknowledge the potential of management controls for employee creativity (e.g., Speklé, van Elten, and Widener 2017). This may also be because controls have long been associated with curtailing creative endeavors, as they mostly build on standardization and routines to increase the efficiency of organizations (e.g., Zhou and George 2003). More specifically, psychological and management research (e.g., Amabile 1983; Shalley, Zhou, and Oldham 2004) have argued that controls harm employee creativity by undermining intrinsic motivation and flexibility. Thus, control and creativity have long been viewed as polar opposites by many authors and practitioners.

Gradually, however, “a literature is emerging on a new control paradigm where management control systems are conceptualized not as hinderance but as a facilitator in entrepreneurship and innovation” (Adler and Chen 2011, p. 81; Davila, Foster, and Oyon 2009). Quite contrary to the previously held belief, some management control studies argue that the nature of generating creative ideas “renders control systems useful and makes them valuable” (Speckbacher 2017, p. 50; Adler and Chen 2011; Grabner and Speckbacher 2016) and rather see management control systems as “flexible and dynamic frames adapting and evolving to the unpredictability of innovation, but stable to frame cognitive models, communication patterns, and actions” (Davila, Foster, and Li 2009, p. 327). This less static notion of management control systems guiding creative endeavors, stimulating creativity, and providing pertinent information at each stage of the idea development process (Speckbacher 2017) hence challenges the presumed tension between creativity and control. Yet, management accounting literature identifying control designs that positively impact creativity is still sparse, and skepticism is greatest for controls that are usually formal, critical, and limiting (Speklé, van Elten, and Widener 2017), such as frequent performance feedback.

Given its importance to management control systems (Luckett and Eggleton 1991) and its multiple, sometimes seemingly contradictory, effects on employee behavior (e.g., Fedor and Buckley 1987; Holderness, Olsen, and Thornock 2020), feedback frequency could be said to reflect the discussion of the paradoxical tension between creativity and management controls. Frequent performance feedback ensures that employees receive timely information they need to get creative. Considering that the process of idea generation has special requirements, frequent feedback enables employees to direct their creative efforts more efficiently and effectively, since the desired output cannot be determined ex ante and only through frequent feedback they can continuously reflect on their current creative attempts. Hence, “increased frequency is argued to help develop feelings of task competence by giving individuals the ability to evaluate their performance” (Luckett and Eggleton 1991, p. 385) and employees with a higher number of feedback contacts are better enabled and motivated to come up with new ideas (Adler and Borys 1996; Coelho and Augusto 2010). However, “at the same time, more frequent feedback may decrease feelings of personal control as the individual perceives a more visible use of the feedback mechanism as a mode of controlling behaviour” (Luckett and Eggleton 1991, p. 385). Feedback may also entail directives from the supervisor about thresholds, boundaries, and curtailing behaviors not in line with firm goals. With frequent feedback, this directive is more present and may be perceived more as controlling and monitoring behavior (Deci, Connell, and Ryan 1989), leading to less freedom and intrinsic motivation, eventually at the detriment of the creative activities of employees (e.g., Speklé, van Elten, and Widener 2017; Grabner and Speckbacher 2016). Yet, management control research is silent on whether feedback frequency compromises or encourages employees' creative idea generation.

From a conceptual point of view, Tessier and Otley (2012) propose that the effects of management control systems (e.g., employee creativity) depend on the interplay between the presentation of control (e.g., its frequency) and the use of control (i.e., managerial intention), such as diagnostic and interactive use. Arguably, both managerial intentions offer the potential to influence the effect of control communication on employee creativity as diagnostic use "motivates action, monitors the outcomes of behaviors, and reward goal achievement" and interactive use "frames enquiries and actions by communicating highly charged concerns and encouraging vertical information sharing" (Speklé, van Elten, and Widener 2017, p. 74). However, the large body of management accounting research has not examined the influence of control use for control presentation effectiveness. I therefore address the ambiguity about impact of feedback frequency for employee creativity by drawing on the concept of control use.

This study complements previous research by filling gaps in the management control system literature. First, I contribute to illuminating the paradox between creativity and control (Adler and Chen 2011; Speckbacher 2017) by showing that feedback frequency, a design element of the management control system, positively stimulates employee creativity. While prior management control studies on creativity largely adopt a systems approach (Simons 1995; Speklé, van Elten, and Widener 2017), knowledge about the creative effects of particular control designs in this research field is scarce. Nonetheless, I similarly find that "paradoxically, creativity can flourish in the presence of control" (Speklé, van Elten, and Widener 2017, p. 73).

Second, this study is first to provide an empirical validation of Tessier and Otley's (2012) framework of management control systems. Consistent with their conceptualization, the study's results show that the use of controls (i.e., interactive and diagnostic use) is critical to employee's response to control design (i.e., feedback frequency) and thus to control effectiveness. Given that Tessier and Otley's (2012) incorporation of control use constitutes a revision of Simons' (1995) seminal Levers of Control framework that has shaped management control research in the past decades (see Martyn, Sweeney, and Curtis 2016), it is not surprising that there is lack of considerable knowledge on the implications of control use. By highlighting the importance of control use, a current gap in the management accounting literature is addressed, as "management accounting research offers a rich literature on various control mechanisms and their use (e.g., diagnostic versus interactive), […] to make important contributions to research on the management of creativity" (Speckbacher 2017, p. 51). More specifically, this study empirically demonstrates that the interactive use of controls indeed "provides a focus for employees' creative juices" (Speklé, van Elten and Widener 2017, p. 78) and thus "encourage[s] new ideas to emerge" (Adler and Chen 2011, p. 76) by amplifying the positive impact of feedback frequency on employee creativity.

Third, my results disagree with prior beliefs that the diagnostic use of controls positively affects employee creativity (Cools, Stouthuysen, and Van den Abbeele 2017). Similar to Henri (2006, p. 537), who assumes that "diagnostic and interactive uses create a dynamic tension" that "foster[s] organizational dialogue, stimulate[s] creativity, and focus[es] organizational attention”, such that diagnostic use magnifies the effect of interactive control use, Simons (1995) and Speklé, van Elten, and Widener (2017) view diagnostic and interactive controls as complementary contributors to employee creativity. Contrary to these conjectures, the results indicate that diagnostic use of feedback does not enhance the effects of interactive control use, but actually attenuates the positive effects of feedback frequency on employee creativity. Consequently, with regard to feedback, there is no optimal (non-zero) threshold for diagnostic use of controls.

Fourth, the results substantiate previous findings that creative requirements on employees influence managers’ decisions in designing control system (Grabner 2014; Grabner and Speckbacher 2016). However, I extent these results by revealing that the level of creativity expected from employees in the job also determines managers’ control communication practices (i.e., its frequency) and their use of controls, such that higher (lower) creative demands encourage managers to provide feedback more (less) frequently and to use feedback more (less) interactively and less (more) diagnostically. Thus, managers do not generally use controls more intensively when they expect employees to be creative (Speklé, van Elten, and Widener 2017), but instead change their pattern of control use. This also complements Cools, Stouthuysen, and Van den Abbeele (2017), who found that whether budgets are used interactively or diagnostically reflects the type of creativity expected from employees.

5.2 Conceptual Framework

I derive my theoretical considerations from the framework of Tessier and Otley (2012) conceptualizing management control systems. Their framework represents a revised version of Simons’ (1995) seminal Levers of Control framework to overcome ambiguities and “vague concept definitions” (Tessier and Otley 2012, p. 182).

Nonetheless, the underlying issues of Simons' (1995) framework, how to manage the tension "between top-down direction and bottom-up creativity" (Simons 1995, p. 4) and "how managers can combine innovation and control" (Adler and Chen 2011, p. 63), remain fundamental. In particular, Tessier and Otley (2012) distinguish between managerial intentions (e.g., control use of the addresser of controls) and presentation of controls (i.e., "decisions as to how communicate with employees", p. 181) that affect an employee's (addressee of management control systems) perception of the control and ultimately determine his or her subsequent (creative) response.

Within Tessier and Otley's (2012) framework, feedback is typologized as technical control[146] and can be located within the operational performance control systems "that focus on critical performance variables at an operational level [and] include feedback systems" (Tessier and Otley 2012, p. 180). Nonetheless, the framework "acknowledges the fact that a specific control can have more than one objective (performance and compliance) and can be used at different organisational levels (operational and strategic)" (Tessier and Otley 2012, p. 179). In line with this, feedback may also control at the strategic level (e.g., quarterly feedback from the board to top managers on the development of market share) or address compliance issues (e.g., weekly feedback to employees in the shipping department on return deliveries due to incorrect addresses). This notion contrasts Simons (1995), who proposes a dichotomy of positive (beliefs and interactive) and negative (boundary and diagnostic) controls that either promote creativity or ensure predictability, and "include[s] [feedback systems] in diagnostic controls" (Tessier and Otley 2012, p. 180). Hence, according to Tessier and Otley's (2012) framework feedback is neither inherently good or bad for employee creativity. In this sense, in this study the frequency of feedback is considered

[146] Tessier and Otley (2012, p. 180) consider "goal settings, output controls and cybernetic controls […] as part of technical controls", which are "controls that specify how tasks are to be performed".

as a particular control design choice within the feedback system about control presentation.

Importantly, Tessier and Otley (2012) argue that it is not the controls per se that determine employee behavior, but instead employee perceptions of the controls, which are shaped by their presentation (e.g., its frequency) and managerial intentions (i.e., control use). In other words, it is the interplay between the way managers communicate (e.g., the frequency of feedback) and the way managers use the controls (e.g., interactively or diagnostically) that drives employee reactions (e.g., creativity). Consistent with Tessier and Otley's (2012) emphasis on the importance of employee perception, the feedback literature stresses the centrality of recipient's feedback perception to behavioral responses (e.g., Ilgen, Fisher, and Taylor 1979; Luckett and Eggleton 1991). Similarly, Speklé, van Elten, and Widener (2017, p. 74) acknowledge that "it matters how control is perceived, and that employee outcomes, such as intrinsic motivation and creativity, depend on whether control is viewed as communicating restrictions and limits or whether it is seen as communicating valuable information and making employees believe they have choices in their actions".

Regarding managerial intentions, Tessier and Otley (2012) distinguish between interactive use and diagnostic use of controls by managers, indicating "the intensity of use of controls, they are not control systems per se, rather, they are descriptions of how managers use controls". This allows several controls, such as feedback, to be used interactively and diagnostically. More specifically, an interactive use of controls "facilitate[s] and promote[s] communication and focus[es] attention" (Tessier and Otley 2012, p. 177) and "facilitate[s] ongoing communication between top managers and lower level of management" (Adler and Chen 2011, p. 76). In contrast, controls, when used diagnostically, "are only looked at when deviances from established targets are observed” (Tessier and Otley 2012, p. 177). In this respect, interactive and diagnostic use represent distinct degrees of intensity of control use (Tessier and Otley 2012). For instance, if the organization’s code of conduct is neither regularly discussed (interactive use) nor brought to employees’ attention when violations occur (diagnostic use), the intensity of its use is low. On the contrary, if used intensively, the code of conduct could be used simultaneously in an interactive and a diagnostic manner.

While managerial intentions and presentation together “represent the elements of the framework that managers have an influence on” (Tessier and Otley 2012, p.181), they can be designed independently. This offers “managers to retain some level of flexibility” (Tessier and Otley 2012, p. 181), particularly for controls that may serve more than one purpose, such as feedback (e.g., ensuring compliance with organizational procedures or promoting employee creativity), and allows managers to “focus on only one of these purposes [because] it is not that the presentation differs from the intention, but rather that the presentation excludes some elements of the control” (Tessier and Otley 2012, p. 181).

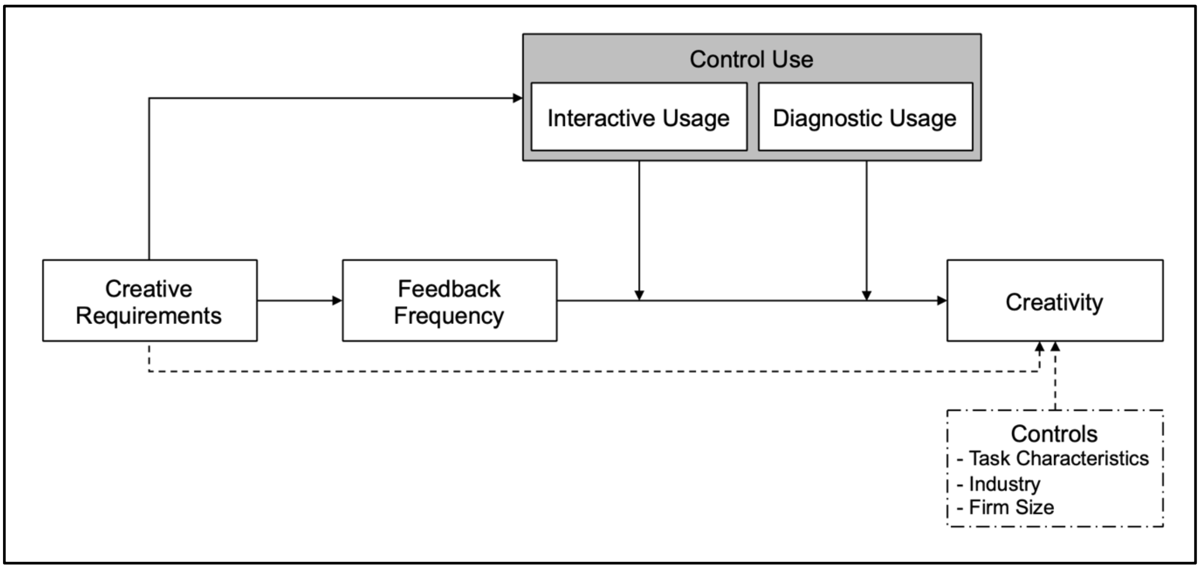


Figure 5.1: Conceptual Model

Thus, in addressing the fundamental tension between control and creativity that is central to the frameworks of Simons (1995) and Tessier and Otley (2012), the question emerges of how feedback frequency (i.e., the presentation of the control “feedback”) and the use of feedback (e.g., interactive or diagnostic) interact in fostering employee creativity. The conceptual model (see Figure 5.1) reflects this question and also incorporates creative requirements on employees to examine whether managers respond to this recently identified contingency factor for management control system design (Grabner 2014; Grabner and Speckbacher 2016) with different control communication practices and control use pattern.

5.3 Hypotheses Development

5.3.1 Feedback Frequency and Creativity

According to Speckbacher (2017), in various strands of literature the tension between control and creativity has long been attributed to two main assumptions. First, controls harm employees’ intrinsic motivation, and because creativity requires employees to be intrinsically motivated (e.g., Amabile 1983; Malik, Choi, and Butt 2019), controls are detrimental to employee creativity. This line of reasoning is supported by motivation crowding theory (Frey and Jegen 2001) or self-determination theory (Gagné and Deci 2005), which suggest adverse effects on intrinsic motivation owing to the fact that controls “such as the setting of goals and targets or performance-based rewards regulate individuals’ extrinsic motivation” (Speckbacher 2017, p. 50). Considering feedback from a control theory perspective (Carver and Scheier 1981; Taylor, Fisher, and Ilgen 1984), these aspects are inherent in the provision of feedback and particularly present if feedback is provided frequently. Similarly, Adler and Chen (2011, p. 66) argue that formal controls (e.g., feedback) require employees to “accept collective goals, conform to pre-given standards and plans, and sacrifice individual interests in order to achieve group goals" and therefore "undermine the intrinsic motivation needed for creativity".

However, feedback has long been recognized in psychological and feedback literature (e.g., Coelho and Augusto 2010; Harackiewicz 1979) as a pivotal source of intrinsic motivation for employees. Especially, "the frequency of feedback should enhance motivation on the task" (Ilgen, Fisher, and Taylor 1979, p. 363) and is "perceived as a limited resource that has value" (Fedor and Buckley 1987, p. 173), reinforcing employees' desire for recompense. Therefore, I expect the negative effects of controls on creativity related to intrinsic motivation to be less pronounced with feedback frequency.

Second, the specific nature of how creative ideas are generated causes controls to undermine creativity (Speckbacher 2017). The rationale behind this is that controls stifle creativity by limiting flexibility at various levels, including cognitive, behavioral, or resource levels. As Speckbacher (2017, p. 51) put it, "creative work exhibits a high level of uncertainty on required inputs, desirable behaviors, and input-output relations, and it is typically impossible to specify the desired outcome", hence "regulating the creative process by constraining behaviors or prescribing results may mislead creative efforts, reduce divergent thinking, and ultimately result in less creative solutions". Following this line of reasoning, feedback risks suppressing employee creativity as "feedback is all too frequently controlling" and "emphasizes how people should behave and implies that the manager is in control" (Deci, Connell, and Ryan 1989, p. 585). Furthermore, controls, such as feedback, require "attention to others, often to hierarchical superiors, rather than to the pleasures of the task itself, creativity requires attention focused on the task rather than on others" (Adler and Chen 2011, p. 66). For instance, take a salesperson who receives weekly feedback on his or her sales numbers (which are not easy to achieve). This employee resembles a driven dog chasing the fulfillment of his or her weekly sales goals (attention to others) rather than acquiring the eye of an eagle perceiving his or her actions (attention to task) in the context of the overall organizational operations, limiting awareness of where and how creative ideas contribute to organizational improvement and the time to work on them.

On the contrary, recent management control research opposes the assumption of incompatibility between creativity and control due to the limitation of flexibility (Adler and Chen 2011; Grabner and Speckbacher 2016). This "new paradigm" (Adler and Chen 2011) envisions management control systems as "flexible and dynamic frames adapting and evolving to the unpredictability of innovation, but stable to frame cognitive models, communication patterns, and actions" (Davila, Foster, and Li 2009, p. 327), underscoring the volatility and uncertainty of the environment in which employees must develop creative ideas. Correspondingly, Speckbacher (2017, p. 51) postulates that "controls take account of the unforeseeable and emergent nature of creative processes and outputs [and] provide continuously updated information on the creative space for creatives' work, keeping an eye on the actual stage of the creative process". This notion of control applies particularly to feedback frequency. Only frequent feedback facilitates managers to provide feedback that adequately addresses the current state and needs within the (ex ante) unpredictable ideation process (e.g., to address novel resource needs that have arisen since the last exchange) so that employees are enabled to constantly recalibrate their creative efforts and respond flexibly to changing demands for creative idea, especially in volatile work environments. Besides, more frequent feedback, ceteris paribus, shortens the time it takes for employees to receive updated or new information, during which they may be working on the basis of outdated or even incorrect information,[147] thus avoiding periods of creative stagnation or misdirected creative efforts. Conclusively, I conjecture that feedback frequency increases flexibility and supports idea generation by virtue of its specific nature and requirements.

Another line of reasoning for the creative effects of feedback frequency can be derived from control tightness[148] (Van der Stede 2001). It has long been argued that tight controls suppress creativity, whereas loose controls "have some unseen benefits, such as in terms of high creativity" (Merchant and Van der Stede 2007, p. 227). Accordingly, feedback frequency should have a negative impact on employee creativity, as high levels of feedback frequency have been associated with tight controls (Campbell, Epstein, and Martinez-Jerez 2011). Nevertheless, recent research points to the potential for tight controls on creativity, as, for example, "tight budgets generate the necessity to rethink everything from scratch" (Speckbacher 2017, p. 50; Cools, Stouthuysen, and Van den Abbeele 2017). I hence propose the following hypothesis:

Hypothesis 1: Feedback frequency is positively related to employee creativity

5.3.2 Moderating Effects of Feedback Use

As mentioned earlier, Tessier and Otley (2012) conceptualize control presentation (i.e., frequency) and control use (i.e., diagnostic and interactive use) as

[147] As the postfeedback interval is also shortened, the risk that the information has become outdated or incorrect since the last feedback is also reduced.

[148] While Tessier and Otley (2012) acknowledge control tightness as important concept in management control research, they do not include it in their framework, as they focus on revising Simons' (1995) framework.

two distinct managerial choices whose interplay determines the employee's (creative) response. According to their framework, interactive and diagnostic control use are also independent from each other, such that feedback (i.e., a control) may be used diagnostically and interactively. Thus, the rationale of the interplay of control use and feedback frequency pursued in this section is that interactive use and diagnostic use add facets to the effects of feedback (e.g., by providing cross-hierarchical information or by guiding employee attention) that accumulate over the number of feedback exchanges resulting in larger (or smaller) differences in the creative outcomes between high and low levels of feedback frequency. In other words, the more intensively feedback is used interactively or diagnostically, the better the employee can exploit on the creative potential of feedback frequency. Consequently, the intensity of interactive and diagnostic use of feedback determines the effects of feedback frequency for employee creativity.

To begin with, interactive control use, such as feedback, "dismantles the functional and hierarchical obstacles between organizational members" (Adler and Chen 2011, p. 76) by increasing the flow of informational through the organization (Abernathy and Brownell 1999; Henri 2006). Not only are employees empowered with more information, but through the interactive use of controls managers also "focus the attention of the entire organization on strategic uncertainties, perceived opportunities, and potential threats" (Adler and Chen 2011, p. 76). Hence, when feedback is used interactively, the more often employees are provided with feedback, the more they "become aware of where potential opportunities and threats may arise, and are [more] motivated to be proactive in searching for new opportunities and guarding against threats" (Speklé, van Elten, and Widener 2017, p. 77). Put differently, through a more intensive interactive use of feedback, each feedback contact offers managers more opportunities to provide the employee with information beyond his or her hierarchical position (Adler and Borys 1996), making each feedback session more effective and increasing differences in cumulative impact of feedback on employee creativity, which varies with the frequency of feedback (i.e., the number of feedback exchanges in a given period).

In addition, interactive control use creates “an environment where people are encouraged to challenge the *status quo*, to engage in debate and dialogue, and to unearth creative and innovative solutions” (Speklé, van Elten, and Widener 2017, p. 78). Similarly, Adler and Chen (2011, p. 76) argue that interactive controls “promote intrinsic motivation by stimulating experimentation with new ideas at all levels and fostering individuals’ innate desire to create and innovate”. Following this line of reasoning, interactive use of feedback increases the impact of feedback frequency on employee creativity by increasing the likelihood that a feedback exchange will trigger a search for a novel idea.

While previous literature consistently suggests that interactive use of feedback enhances the positive effects of feedback frequency on creativity, the effects of diagnostic control use are more controversial in the academic debate. Earlier management control research tends to emphasize the negative consequences for employee creativity arising from diagnostic control use, such as “constrain[ing] innovation and opportunity-seeking to ensure predictable goal achievement” (Simons 1995, p. 91). Furthermore, diagnostic control use underscores existing functional responsibilities and accountabilities (Abernathy and Brownell 1999; Henri 2006), reducing employee motivation and identification with organizational goals (Adler and Chen 2011). This mechanism should apply particularly to the interpersonal control of feedback by having “a built-in power component that preserves and amplifies existing power differences in organizations (e.g., differences in organizational authority and rank)" (Bear et al. 2017, p. 721). As a frequent presentation of feedback raises employee awareness, these (often implicit) facets of diagnostic feedback use thwart the otherwise beneficial effects of feedback frequency on employee creativity.

In contrast, more recent management control research seizes on the potential of diagnostic control, potentially enhancing the creative effects of feedback frequency. Most fundamentally, diagnostic control use "make[s] transparent the organization's goals" (Adler and Chen 2011, p. 75; Speklé, van Elten, and Widener 2017) and thereby provides guidance to employees. If feedback is not used diagnostically, its frequency cannot signal relevance to the employee. While generally, it is expected that managers point more frequently to deviations (i.e., diagnostic use) in critical success factors, if negative (positive) deviations, for example, from the sales target repeatedly do not result in negative (positive) feedback from the manager (i.e., low intensity of diagnostic feedback use), the employee may assume that achieving the sales target is not important (or at least not more important than other performance dimensions) to both the manager and the organization's success, regardless of the frequency of feedback. Consequently, the employee will neither make a last-ditch effort to achieve the goal nor will he or she come up with creative solutions that increase efficiency or effectiveness in achieving this goal. Similarly, if feedback is used diagnostically, frequent feedback can make progress toward organization's goals transparent, which facilitates employee learning, promotes perception of competence, and elicits intrinsic motivation (Adler and Chen 2011; Speklé, van Elten, and Widener 2017), all of which enhances employee creativity (e.g., Amabile 1988). Furthermore, when controls are used diagnostically, they can be "perceived as challenges that only make the problems at hand more interesting, priming individuals to think of unusual solutions and nonstandard approaches" (Speklé, van Elten, and Widener 2017, p. 78). Hence, diagnostic feedback helps employees recognize that their creative efforts are required and, given that diagnostic controls reflect accountability structures, it "motivates employees to take action" (Speklé, van Elten, and Widener 2017, p. 77). Therefore, diagnostic use feedback expands the effectiveness of feedback frequency by making it a lever for the employee's understanding of the organizational goals and how his or her behaviors relate to it, and by causing feedback frequency to correspond with the number of challenges the employee faces, thereby magnifying the difference in effectiveness between low frequency (i.e., little understanding, few opportunities for creative ideas) and high frequency (i.e., better understanding, more opportunities for creative ideas).

Conclusively, I expect the effectiveness of feedback frequency to vary with the intensity that feedback is used by managers (i.e., interactively and diagnostically), causing the employees' creativity to vary as well, and therefore hypothesize:

Hypothesis 2: More intensive interactive use increases the positive relationship between feedback frequency and employee creativity

Hypothesis 3: More intensive diagnostic use increases the positive relationship between feedback frequency and employee creativity

5.3.3 Implications of Creative Requirements

At the latest since the seminal work of Chenhall (2003), management control system research has acknowledged a contingency-based research approach to the design of management control systems. Recently, Grabner and Speckbacher (2016) found that the organization's emphasis on employee creativity is a crucial contingency factor in the design of management control systems, as managers have less task-specific cause-effect knowledge and place more emphasis on intrinsic motivation due to the nature of creativity. While this leads to an increased use of certain controls to regulate the additional risk of dysfunctional employee behavior, at the same time organizations “abstain from the use of (otherwise effective) controls because such controls increase the risk of undermining employee creativity” (Grabner and Speckbacher 2016, p. 31).

More specifically, Simons (1987) demonstrates that firms whose strategy includes a “greater emphasis on fostering individual creativity” (p. 360) are more likely to adopt tight controls owing to the high uncertainty associated with creative activities. Similarly, Henri (2006, p. 536) concludes that innovation capabilities “[impose] additional pressure on the organization’s information processing capacity and more interaction between top management and subordinates is required to increase the flow of information”. In congruence with the previous hypotheses, I therefore argue that managers respond to creative requirements on employees by designing and using the management control system in a way that fosters employee creativity:

Hypothesis 4a: Creative requirements on employees are positively associated with the frequency of feedback they receive

Hypothesis 4b: Creative requirements on employees are positively associated with interactive use of feedback

Hypothesis 4c: Creative requirements on employees are positively associated with diagnostic use of feedback

5.4 Methodology

5.4.1 Sample and Data Collection

An online questionnaire is used to collect data for testing the hypothesized relationships. Since the study seeks to capture responses to control system design choices, the most knowledgeable organizational members are the intended addressees of control systems. Therefore, employees at different organizational levels and in multiple departments are considered as the target population of this study. This is consistent with creativity research notions about the creative potential of organizations in their workers, regardless of their job and position (e.g., Zhou and George 2003).

In developing the survey, a draft version of the questionnaire was pretested for comprehensibility, neutrality, completeness, structure, and length by fourteen academics and members of the target population who were then removed from the main survey. Only minor adjustments were made. I put emphasis on an attractive design of the questionnaire (Churchhill and Iacobucci 2005) and on appropriate incentives (Helgeson, Voss, and Terpening 2002) by sending respondents an exclusive report on the results of the study in order to ensure a high response rate.

Industry	Percentage
Agriculture, forestry and fishing	0.8%
Automotive industry and suppliers	12.5%
Chemical industry	5.0%
Construction	6.5%
Energy and environmental industries	3.5%
Engineering	3.3%
Finance, insurance and real estate	6.8%
Food production	1.5%
Health, medicine and pharmaceuticals	11.3%
Information and communication technology	6.5%
Materials manufacturing and processing	2.3%
Plant and mechanical engineering	7.3%
Retail or wholesale trade	8.3%
Textile production	0.8%
Tourism	1.3%
Transportation and logistics	5.0%
Other manufacturing	3.8%
Other services	2.3%
Other business activities	11.8%
Sales (Mio. EUR)	
< 10	24.8%
10 - 100	25.9%
100 - 1,000	24.0%
> 1,000	25.3%
Employees	
< 100	21.8%
100 - 1,000	27.6%
1,001 - 5,000	21.6%
> 5,000	29.1%

Figure 5.2: Sample Characteristics

A total of 425 responses were received. Ten of the completed questionnaires came from employees of nonprofit companies or from self-employed respondents who were excluded from the study. In addition, some responses were deemed unusable owing to rushing through the questionnaire (eleven) or straight-lining (four). This leaves me with 400 usable observations (adjusted response rate of 30.9%) from employees working in companies of various industries and sizes. The descriptive information in Figure 5.2 presents an industry classification as well as an overview of company sizes in terms of employees and sales. More specifically, participants work in various departments such as marketing and sales (21.5%), manufacturing and production (15.3%), or research and development (9.0%). The proportion of female (38.0%) and of male (61.8%) respondents corresponds to current full-time employment figures in Germany.[149]

I test for potential (non)response bias (Armstrong and Overton 1977). A comparison was made between early (25% of the sample), middle (50%), and late respondents (25%). The results show that all survey constructs did not differ significantly, nor did industry (service vs. non-service) and firm size in terms of sales and employees or employee gender (Kruskal Wallis test 1952).

5.4.2 Assessment of Common Method Bias

Because the data were collected from the same individuals at the same point in time, there is an eligible concern that common method variance may be present (Podsakoff and Organ 1986; Podsakoff et al. 2003). For this reason, I undertook several procedural remedies as suggested in the literature (Podsakoff et al. 2003). First, I counterbalanced the order of questions for the dependent and independent

[149] See https://www.destatis.de/DE/Themen/Arbeit/Arbeitsmarkt/Qualitaet-Arbeit/Dimension-1/teilhabe-frauen-erwerbsleben.html [retrieved 08-21].

variables so that respondents could neither draw conclusions about nor be informed of the underlying conceptual model of the survey. Second, respondents were made aware that there were no right or wrong answers, but were encouraged to respond honestly and comprehensively. Third, anonymity of responses was ensured and communicated, and self-administered questionnaires were used, as opposed to in-person data collection, such as face-to-face interviews. Fourth, where applicable prevalidated measurement scales were used and adaptions were extensively pretested to ensure comprehensibility and to avoid social desirability bias.

Furthermore, I tested for common-method-bias using an exploratory factor analysis that included all items (e.g., Podsakoff and Organ 1986). The non-rotated solution yielded multiple factors, with the first factor not accounting for a majority of the variance (26.1%). In addition, a full collinearity approach was employed (Kock 2015; Latan, Ringle, and Jabbour 2018). Variance inflation factors (VIFs) for all constructs were substantially below the threshold of 3.3. These outcomes suggest that common method variance did not adversely affect the results. It was also controlled for potential common method variance following the marker variable test (Rönkkö and Ylitalo 2011; Limaj and Bernroider 2019) by including a marker variable in the model that is theoretically unrelated to any of the variables. The test (see Figure 5.5) shows that the inclusion does not alter path significance levels, further alleviating concerns that the results are due to common method bias.

5.4.3 Measures

All data were obtained from the collected questionnaire. The multi-item measurement models were reflectively designed on a 7-point Likert-type scale ranging from “does not apply at all” to “fully applies” (Bagozzi 1981). The measurement models are presented in Figure 5.3 with their respective loadings and construct reliabilities.

Where possible, I relied on established scales. In the following, I describe the construct measurements.

Creativity is measured with four items from the widely used Farmer, Tierney, and Kung-McIntyre's (2003) creativity self-assessment scale (e.g., Speklé, van Elten, and Widener 2017). Similar to much of the creativity research (e.g., Baer et al. 2010; Shalley, Gilson, and Blum 2009), a self-report measure was employed for assessing creativity because "an individual's creativity is unlikely to be accurately assessed by any observer" (Coelho, Augusto, and Lages 2011, p. 41). Besides, previous studies have pointed out a strong overlap of self-reported creativity measures with external assessments (e.g., Axtell et al. 2000), but also that "objective measures of creativity are often context- or even organization-specific" (Speklé, van Elten, and Widener 2017, p. 84), which would compromise with the cross-industry and cross-departmental study design. Therefore, the employee him- or herself is the best judge of his or her creativity.

Building on the work of Eisenberger, Rhoades, and Cameron (1999), three items for feedback frequency were derived from established feedback measures included in the Job Diagnostic Survey (Hackman and Oldham 1975) and in the Job Characteristics Inventory (Sims, Szilagyi, and Keller 1976; Dubinsky and Skinner 1984). Specifically, participants are asked for feedback from their direct supervisor so that the power distance between feedback source and recipient is controlled (Ilgen, Fisher, and Taylor 1979). Consistent with prior studies observing feedback frequency and related constructs (e.g., Auh et al. 2019; Dagger, Danaher, and Gibbs 2009; Doney and Cannon 1997; Kinicki et al. 2004; Kuvaas 2011; Mayer, Ehrhart, and Schneider 2009; Seiders et al. 2015), The importance of the recipient's subjective perception of feedback frequency was considered and hence a rigid and quantifiable frequency measure was eschewed: "regardless of the amount of feedback […] a worker really has in his work, it is how much *he perceives that he has* which will affect his reactions" (Hackman and Lawler 1971, p. 264; Ilgen, Fisher, and Taylor 1979). Nevertheless, a redundancy analysis was conducted (Chin 1998; Cheah et al. 2018; Sarstedt, Wilczynski, and Melewar 2013; see chapter 2.1.4.2) to specifically test construct validity. In doing so, the multi-item reflective measurement model of feedback frequency used in this survey is correlated with an objective measure of feedback frequency (adapted from Alvero, Bucklin, and Austin 2001 and Morrison 1993). The results show that convergent validity is established ($r = 0.54$, $p < 0.01$).

Previous measures for diagnostic and interactive controls (Henri 2006; Speklé, van Elten, and Widener 2017; Widener 2006) are infused with conceptual considerations of Simons' (2000) four levers of control, so these measures reflect a control-based perspective. For example, these items implicate that return targets (as part of performance measurement) are a diagnostic control ("Are returns targets used for your job?" in Speklé, van Elten, and Widener 2017, p. 82). This is not consistent with Tessier and Otley's (2012) reconsideration of diagnostic and interactive as concepts that describe the use of controls. Within their framework, return targets (and any other control) can be used interactively and diagnostically so that managers can regularly discuss the content of performance measures (interactive use) as well as pay attention to these performance measures when deviations from target values occur (diagnostic use). Due to their control-oriented approach, the items used so far in management control research hence also lack a content-related link to feedback. I therefore develop two new scales for interactive use and diagnostic use, deriving items from Tessier and Otley's (2012) construct descriptions and definitions, but also incorporating previous scales, in particular the work of Speklé, van Elten, and Widener (2017). As employees are confronted with many feedback sources, I particularly emphasize the supervisor as feedback source for being responsible in designing the control system for the employee. In addition, confirmatory tetrad analysis (CTA-PLS, Gudergan et al. 2008; see chapter 2.1.3.3) was used to empirically ensure correct measurement model specification by following the procedure described for measurement models with fewer than four items. For both measurement models, the vanishing tetrads (based on 5,000 bootstrap samples) support the assumption of reflective measurement models. For creative requirements, three items were adopted from Unsworth, Wall, and Carter (2005).

To rule out alternative explanations, additional data was collected on variables that might interfere with the results. In selecting control variable, I follow previous publications (Grabner and Speckbacher 2016; Speklé, van Elten, and Widener 2017) by controlling for task characteristics (autonomy, output measurability), industry (service vs. non-service), and firm size (in terms of number of employees). Autonomy is measured by converting two items from Grabner and Speckbacher's (2016) delegation scale, which is based on the work of Hage and Aiken (1967) and "captures the degree of employee autonomy in making day-to-day decisions" (Grabner and Speckbacher 2016, p. 36), into a self-assessment measure. Similarly, an item on output measurability was derived from an established scale (Grabner 2014, Grabner and Speckbacher 2016) and transformed into a self-assessment measure. Additionally, a dummy variable was created for industry sector that captures whether the firm operates in a service or non-service industry (Grabner and Speckbacher 2016). Finally, the number of employees was used as proxy for firm size.

For the marker variable, I rely on Miller and Simmering's (2020) blue attitude scale because attitudes are particularly susceptible to common method variance (Podsakoff et al. 2003) and "the affective and evaluative elements inherent in the blue attitude items […] make this marker similarly susceptive to common method variance" (Simmering et al. 2015, p. 487). Along with the strength of this scale being theoretically unrelated to constructs used in accounting or management research, the pretests showed that respondents were disturbed by the content of these items. In response, I adjusted the items by using the city in which my university is located as the subject to be assessed.

Latent Variable	Measures (Standardized Loadings)
Feedback Frequency (α = 0.83, CR = 0.90, AVE = 0.74)	My supervisor often lets me know how well I am doing on my job (0.91)
	I always receive information from my supervisor about the results of my job performance (0.90)
	Anytime, I have the opportunity to get feedback from my supervisor about my job performance (0.76)
Creativity (α = 0.77, CR = 0.85, AVE = 0.59)	Please indicate the extent to which you could be described as a person who…
	…always thinks of other ways to solve problems when he or she runs into obstacles (0.69)
	… copes with several new ideas and problems at the same time (0.68)
	… helps other people develop new ideas (0.84)
	… has lots of new ideas (0.84)
Interactive Use (α = 0.80, CR = 0.88, AVE = 0.71)	Feedback between my supervisor and me usually takes place as a dialogue (0.75)
	The feedback from my supervisor is used for an intensive exchange and discussion of new information (0.88)
	My supervisor and I try to learn from each other through feedback (0.89)
Diagnostic Use (α = 0.70, CR = 0.81, AVE = 0.59)	My supervisor primarily gives feedback when there are deviations from objectives (0.71)
	My supervisor mainly gives feedback when established targets have been missed (0.76)
	My supervisor gives feedback even when there are no problems in achieving targets (rc) (0.84)

Creative Requirements (α = 0.88, CR = 0.93, AVE = 0.81)	My job requires me to have ideas about changing service or facilities for customers (0.89)
	My job requires me to have ideas about changing ways of organizing work (0.90)
	My job requires me to have ideas about changing work goals and objectives (0.91)
Autonomy (α = 0.54, CR = 0.81, AVE = 0.68)	I have great deal of freedom in scheduling work (0.85)
	I make decisions within assigned area of responsibility (0.80)
Output Measurability (single-item)	The desired performance requirements (objectives) towards me are clearly defined
Marker Variable (α = 0.80, CR = 0.88, AVE = 0.71)	I prefer Kassel to other cities (0.79)
	I like the city of Kassel (0.88)
	I hope my next place of residence is Kassel (0.85)
Note: α = Cronbach's alpha; CR = composite reliability; AVE = average variance extracted; rc = reverse coded.	

Figure 5.3: Measurement Models

5.4.4 Data Analysis

To assess the consistency and validity of the measurements, items were analyzed using partial least squares structural equation modeling (PLS-SEM; SmartPLS 3.3, Ringle, Wende, and Becker 2015), following the guidelines of Hair et al. (2012) and Hair et al. (2019) for assessing reflective outer models as outlined in chapter 2.1.4.2. To begin with, the ratio between the sample size and the number of path relationships[150] is sufficiently high to satisfy the "ten times" rule (Barclay, Higgins, and Thompson 1995), indicating that the model estimates are robust. Composite

[150] Actually, the sum of path relationships between latent variables in the inner model and the number of indicators in formative measurements in the outer model are considered when the "ten times" rule is applied. However, since the structural model does not include formative measures, I only mention the number of path relationships.

reliabilities of at least 0.89 indicate good internal consistency (Bagozzi and Yi 1988; Nunnally and Bernstein 1994). Besides, the constructs' Cronbach's alphas are as well sufficiently high.[151] As shown in Figure 5.3, convergent validity is established since all standardized item loadings exceed the minimum threshold of 0.7 (Hulland 1999). Moreover, the reported average variance extracted (AVE) ranges from 0.59 for interactive use to 0.81 for creative requirements, also suggesting convergent validity (Bagozzi and Yi 1988). To demonstrate discriminant validity, three different criteria were used. First, the heterotrait-monotrait ratio (HTMT) of the construct's correlations was determined (Antonetti, Crisafulli, and Maklan 2018; Henseler, Ringle, and Sarstedt 2015; Hogreve, Bilstein, and Hoerner 2019). This compares the average correlation (arithmetic mean) of each combination of indicators from two different latent variables (heterotrait-heteromethod correlations) with the geometric mean of the average correlations between the indicators of each latent variable (monotrait-heteromethod correlations). The resulting ratio, also referred to as disattenuated correlation, estimates the true correlation between two constructs. A bootstrapping with 5,000 samples shows that all 95% confidence intervals of the HTMT do not even include the conservative threshold of 0.85 (Voorhees et al. 2016). Second, the AVE for each construct is larger than the squared correlation with any other respective construct (Fornell and Larcker 1981) (see Figure 5.4). Third, each indicator loads most strongly on the intended construct (Chin 1998) (untabulated). Consequently, discriminant validity is established.

[151] Since the items for autonomy are derived from an established scale and this construct is used only as a control variable, the Cronbach's alpha of this construct, which is just below the threshold, is not considered critical. In addition, while the composite reliability indicates that reliability is established, the Cronbach's alpha is generally considered to be "the lower bound […] for internal consistency reliability" (Hair et al. 2019, p. 15).

	FF	Creativity	Interactive use	Diagnostic use	CRq
Feedback Frequency	*0.86*				
Creativity	0.25	*0.77*			
Interactive use	0.58	0.26	*0.84*		
Diagnostic use	- 0.55	- 0.11	- 0.48	*0.77*	
Creative requirements	0.22	0.45	0.30	- 0.23	*0.90*
Autonomy	0.19	0.28	0.29	- 0.23	0.32
Output measurability	0.27	0.06	0.31	- 0.18	0.07
Firm Size	0.16	0.01	0.04	- 0.11	0.04
Industry	0.07	- 0.03	0.05	- 0.02	- 0.04
Note: Diagonal elements in italics are the square root of constructs' average variance extracted, offdiagonal elements display bivariate correlations between the constructs; FF = feedback frequency; CRq = creative requirements.					

Figure 5.4: Latent Variable Correlations (Fornell-Larcker criterion)

5.5 Results

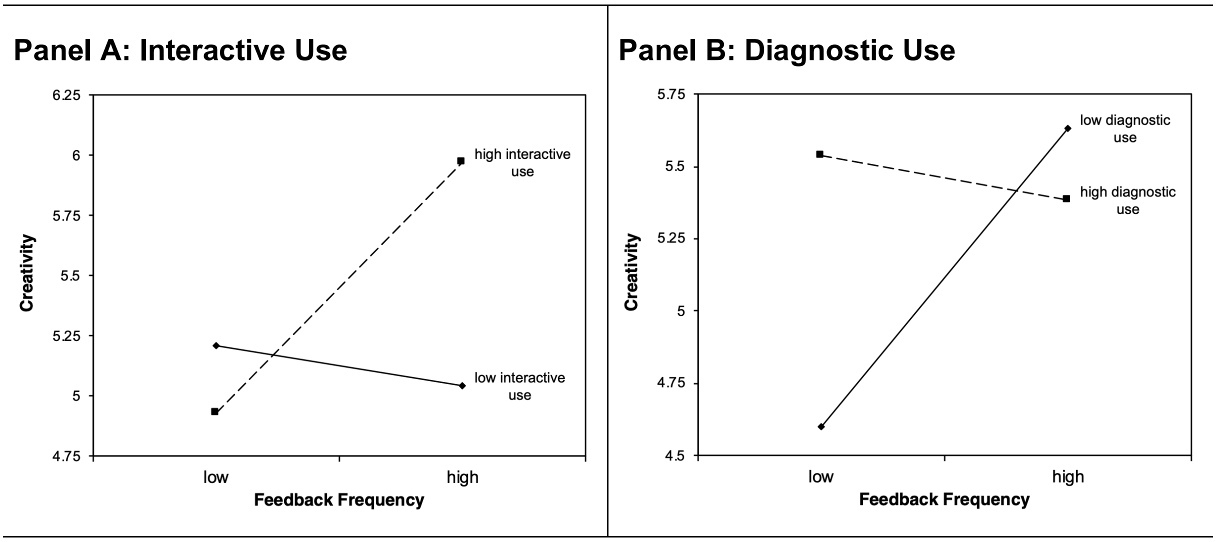
Structural equation modeling (SEM) was employed to test the hypotheses. In particular, variance-based partial least squares (SmartPLS 3.0, Ringle, Wende, and Becker 2015) was used due to its leniency to distributional assumptions. The results of the structural equation analysis provide support for the conceptual model. Figure 5.5 presents an overview of the path coefficients and their respective significance levels for the baseline model, when control variables are included, and for the marker variable test. Neither does the inclusion of control variables or marker variable alter the results, nor do the marker variable relationships indicate that the latent variables share a common variance.

	Hypothesis	Baseline		With Controls		Marker Variable	
		pc	t-value	pc	t-value	pc	t-value
Estimated Paths							
FF → Creativity	H1	0.20 **	3.12	0.21 **	3.22	0.21 **	3.22
FF X Interactive → Creativity	H2	0.11 *	2.28	0.12 *	2.39	0.12 *	2.42
FF X Diagnostic → Creativity	H3	-0.12 *	2.12	-0.12 *	2.39	-0.12 *	2.19
CRq → FF	H4a	0.22 **	4.62	0.22 **	4.58	0.22 **	4.67
CRq → Interactive	H4b	0.30 **	6.03	0.30 **	5.98	0.29 **	5.95
CRq → Diagnostic	H4c	-0.23 **	4.85	-0.23 **	4.93	-0.23 **	4.74
CRq → Creativity		0.41 **	9.41	0.36 **	8.12	0.36 **	7.83
Autonomy → Creativity				0.15 **	2.94	0.16 **	3.09
OutM → Creativity				-0.03	0.57	-0.03	0.60
Firm size → Creativity				-0.01	0.11	-0.00	0.01
Industry → Creativity				-0.03	0.80	-0.03	0.77
Marker → FF						-0.02	0.22
Marker → Creativity						0.05	1.01
Marker → Interactive						0.05	0.82
Marker → Diagnostic						-0.04	0.74
R^2 Creativity		28.4%		30.4%		30.6%	
n		400		400		400	

Note: pc = path coefficient; FF = feedback frequency; Interactive = interactive use; Diagnostic = diagnostic use; CRq = creative requirements; Industry = dummy variable coding for service firm (= 1) and for non-service firm (= 2); OutM = output measurability; *p < .05; **p < .01 (two-tailed test), results of 5,000 bootstrapping samples.

Figure 5.5: Standardized Estimates of Inner Model

More specifically, in H1 I argue that higher feedback frequency increases employee creativity. In line with this hypothesis, the results reveal that feedback frequency positively impacts creativity (H1, $\beta = 0.20$, $p < .01$). Regarding the moderating effects of control use, I find interactive use to significantly increase the relationship between feedback frequency and creativity (H2, $\beta = 0.11$, $p < .05$, $f^2 = 0.02$, see Panel A of Figure 5.6), providing support for H2. Contrary to my expectations, the intensity of diagnostic control use does not increase but decreases the relationship between feedback frequency and creativity (H3, $\beta = -0.12$, $p < .05$, $f^2 = 0.02$, see Panel B of Figure 5.6), confirming a moderating effect of diagnostic control use as hypothesized in H4, yet revealing a reverse direction of the effect. Following Aiken and West (1991) and Kenny (2013), the effect sizes (f^2) of the hypothesized moderations were evaluated finding medium effect sizes that indicate substantial moderation effects of the respective control use.



Note: low (high) values of feedback frequency, interactive use, and diagnostic use refer to 1SD below (above) their respective mean.

Figure 5.6: Moderating Effects of Control Use on the Relationship between Feedback Frequency and Creativity

Furthermore, I examine the effects of creative requirements on feedback frequency and control use. Consistent with H4a and H4b, the analysis shows that creative requirements lead managers to increase the frequency of feedback (H4a, $\beta = 0.22$, $p < .01$) and the intensity of interactive control use (H4b, $\beta = 0.30$, $p < .01$). Surprisingly, managers reduce the intensity of diagnostic control use (H4c, $\beta = -0.23$, $p < .01$) when higher levels of creative requirements are present. Although this finding is inconsistent with the relationship hypothesized in H4c, as it indicates a reverse direction, it fits the results in H3 and my reasoning, which suggests that managers use controls to promote employee creativity when creative requirements are present. A mediation analysis (Zhao, Lynch, and Chen 2010; see chapter 2.1.2.2) shows a significant complementary mediation of the effect of creative requirements on employee creativity by feedback frequency (0.05, $p < .01$).

	VIF [a]	R^2	Q^2
Feedback Frequency	1.77	4.9%	0.03
Creativity	--	28.4%	0.15
Interactive use	1.77	8.7%	0.06
Diagnostic use	1.55	5.4%	0.02
Creative requirements [b]	1.11	--	--

[a] Variance influence factor (VIF) associated with creativity, as this is the only construct in the model with multiple predictors.
[b] Since creative requirements is not used as an endogenous variable, R^2 and Q^2 cannot be calculated.

Figure 5.7: Overview of Latent Variable Statistics

Finally, following the elaborations in chapter 2.1.4.3 the analysis of the structural model reveals that there are no collinearity issues among the predictors of creativity (all variance influence factors are substantially below the threshold of 3) and that all endogenous variables in the model have predictive relevance (Stone-Geisser criterion). These values and the explained variance of all endogenous constructs are shown in Figure 5.7.

5.6 Discussion

For long, management control systems have been accused of impairing employee creativity by undermining intrinsic motivation and curtailing flexibility in creative genesis processes (Speckbacher 2017). This is actually a surprising paradox, considering that the underlying intention of control systems is to influence employee behavior in order to achieve firm goals such as creativity and innovativeness. Recent studies refute this negative view of controls (Adler and Chen 2011; Cools, Stouthuysen, and Van den Abbeele 2017) and show, for instance, through a control package approach, that the intensity of the levers of control (Simons 1995) is positively associated with creativity (Speklé, van Elten, and Widener 2017). Yet, management control literature remains silent on the creative effects of particular controls.

This study therefore set out to examine the effects of feedback on employee creativity, the design of which (e.g., its frequency) is central to management control (Luckett and Eggleton 1991). Specifically, I draw on Tessier and Otley's (2012) revised management control framework, which posits that the interplay of design of control (i.e., feedback frequency) and the use of control (i.e., diagnostic and interactive use) determine employees' creative responses to controls. The results of this cross-industry and cross-departmental study (n = 400) support recent management control studies that challenge the paradox between control and creativity, and extend this strand of literature by highlighting a specific instrument that has the potential to enhance employee creativity, as it shows that feedback frequency positively stimulates employee creativity.

In addition, this study provides empirical evidence that control use determines outcomes of control designs. Whereas, according to Simons (1995), interactive and diagnostic refer to subsets of controls, the conceptual revision of Tessier and Otley (2012) implies that a control is not interactive or diagnostic per se, but can be used by managers as such. Due to the influence of Simons (1995) on management control research in the past decades (see Martyn, Sweeney, and Curtis 2016), this difference in conceptualization may explain to some extent the scarce knowledge on interactive and diagnostic control use (e.g., Henri 2006). The results demonstrate that the effectiveness of feedback frequency (control design) is dependent on whether feedback is used interactively or diagnostically, enhancing or diminishing, respectively, its impact on employee creativity. This finding not only contributes to conceptual management control research, but also provides managers a further lever for controlling employees: Managers may be constrained in designing controls, for instance, due to limited resources, corporate regulations, or corporate culture, yet they can still enhance employees' effectiveness in how they use controls – as any control can be used interactively or diagnostically (Tessier and Otley 2012).

Similar to this question of how managers approach employees with controls, Speklé, van Elten, and Widener (2017, p. 92) suggest that "it would be interesting to determine whether the use of the Levers of Control system is positively associated with creativity in the presence of different leadership styles". In this vein, Klein and Speckbacher (2020, p. 313) state that "it depends on the team leader's leadership style whether the negative or positive effects of using customer-related data in performance evaluations on team creativity will prevail". Considering that control use is similar in nature to facets of leadership styles such as management by exception (Bass 1985), which requires an intensive diagnostic use of controls, or empowering leadership (Kirkman and Rosen 1999), which builds on interactive communication across hierarchical levels, the findings therefore to some degree also offer insight into this intertwined theme in the leadership literature.

Moreover, this study sheds further light on the ambiguity regarding the effects of the use of diagnostic control on employee creativity by finding that diagnostic control use of a particular control (i.e., feedback) does not enhance the positive effects of its interactive use. According to Simons (1995), diagnostic controls, as one of the four Levers of Control, are among the negative forces that constrain employees and ensure predictable goal achievement. However, he adds that "the power of these levers […] does not lie in how each is used alone, but rather in how they complement each other when used together" (Simons 2000, p. 301). Specifically, Henri (2006, p. 537) assumes that the dynamic tension between diagnostic and interactive use promotes creativity by forcing "organizational members to integrate seemingly opposed elements" and by "trigger[ing] the identification of alternative ways of doing things by supporting the identification and synthesis of a variety of viewpoints". Consistent with this reasoning, research has found that the Levers of Control "as a system provides both autonomy support and structure" (Speklé, van Elten, and Widener 2017, p. 74) and "allow[s] experimentation and creativity to flourish, while at the same time providing constraints on employee's behavior" (Mundy 2010, p. 500).

However, when this control systems approach is abandoned and a control perspective is adopted, the results caution against the expectation that dynamic tension stimulates employee creativity, as no interactive effects between diagnostic and interactive use of a given control (i.e., feedback) on employee creativity are found.

This result complements previous findings that diagnostic use of performance management systems inhibits strategic capabilities to innovate while not increasing the positive effects of interactive use of performance management systems (Henri 2006). Consequently, bearing in mind the negative moderating effect on the impact of feedback frequency, there is no optimal (non-zero) threshold for diagnostic feedback use. In other words: If managers want to foster employee creativity, they are constrained in their control design portfolio and may have to substitute other controls for the otherwise positive effects of diagnostically used feedback. This principle, that designing some controls to achieve creative goals requires more intensive use of other controls to compensate for their omitted (non-creative) effects, explains to some extent the findings of Speklé, van Elten, and Widener (2017) on more intensive use of all Levers of Control that promote employee creativity, rather than a single one.[152] The results therefore suggest that managers "do not need to make tradeoffs between whether they want a creative organization or an organization characterized by control" (Speklé, van Elten, and Widener 2017, p. 75) when designing the management control package, they must do so at the control level.

Lastly, my results contribute to the ongoing discussions on contingent management control system design. In practice, it can be observed that firms adapt their management control system depending on the level of creativity they expect from their employees, such as by the level of delegation or the use of predefined targets for performance evaluation (Grabner and Speckbacher 2016) or the design of incentive systems (Grabner 2014). Similarly, I find that managers respond to creative demands

[152] While boundary controls, diagnostic controls, and interactive controls are not significantly directly related to employee creativity, only the effect of beliefs control remains significant when Speklé, van Elten and Widener (2017) decompose their higher-order component construct of intensity of Levers of Control.

placed on their employees with particular control design choices (i.e., feedback frequency). However, while previous research must be “interpreted as a joint test of [..] hypotheses and the assumption of optimality” (Grabner and Speckbacher 2016, p. 41), the results show that the choice on control communication (on average) is also optimal.

This study extends research on the contingency of management control systems by revealing that managers vary not only the design of controls but also their control use when different levels of creativity are expected from their employees. In this regard, I find that the chosen use of control by managers (on average) also promotes employee creativity. Thus, this study complements Cools, Stouthuysen, and Van den Abbeele (2017), who found that interactive and diagnostic control use depend on the type of creativity required from employees. However, the results contrast with those of Speklé, van Elten, and Widener (2017) in that I do not find that managers generally increase their use of controls in order to foster employee creativity, but rather selectively change their control use pattern by using feedback more interactively but less diagnostically. Nonetheless, this is in line with the notions of Grabner and Speckbacher (2016, p. 31), who find that managers use some controls more intensively in the presence of creative requirements, but at the same time “abstain from the use of (otherwise effective) controls”.

5.7 Managerial Implications

This study offers several practical recommendations for managers. First and foremost, managers can foster creativity in their employees by giving them frequent feedback. Importantly, the extent of this positive effect depends on how managers use this management control. Surprisingly, frequent feedback arising from occasions in which managers alert employees to focus their attention on deviations from target values (diagnostic use) is not perceived by employees as opportunities to learn about starting points for novel ideas; instead, they feel monitored and may avoid risks associated with developing creative ideas. By contrast, when managers use frequent feedback as an invitation to discuss with employees and provide them with information across hierarchical levels and departmental boundaries (interactive use), employees feel empowered and are motivated to embark on creative journeys.

Recognizing managers' limited time resources, my findings help them achieve higher levels of efficiency and effectiveness. Since control use reflects the manager’s level of engagement (Tessier and Otley 2012), the results suggest that rather than investing more effort, managers can shift their efforts from a diagnostic use of feedback to a more interactive use if they want to foster creativity in their employees. Arguably, interactive feedback use requires some effort on the part of the manager in preparing for feedback sessions, such as proactively seeking additional information to enrich the feedback conversation. However, this effort is not necessarily obviated when feedback is used diagnostically. Importantly, refraining from diagnostic feedback not only creates free space for managers, but also makes the feedback exchanges more predictable for managers and avoids erratic hustle in response to (positive and negative) deviations from target values. This relieves managers (e.g., by eliminating the need to permanently monitor performance indicators) and prevents unnecessary disruptions in day-to-day business (e.g., through spontaneous “crisis meetings”), so that – in addition to the positive effects on employee creativity – a higher efficiency of the feedback system can be expected when feedback is used interactively. In turn, the space thus gained could be invested in a higher frequency of feedback to increase its effectiveness. Vice versa, this efficiency gain should also materialize on the side of feedback recipients.

5.8 Limitations and Future Research

Regarding the limitations of this study, the research design is subject to the common drawbacks of cross-sectional studies. Besides, this study examines the effects of a control design choice (i.e., feedback frequency) that has been positively associated with employee motivation (e.g., Ilgen, Fisher, and Taylor 1979) and by its nature contributes to organizational flexibility and reactiveness and therefore overcomes aspects of controls that have been criticized for undermining employee creativity (cf. Adler and Chen 2011; Speckbacher 2017). Although the results of Speklé, van Elten, and Widener (2017) support at an abstract level the finding that high levels of feedback frequency augments employees' creative potential, further studies on the creativity effect of specific controls that are not inherently associated with such qualities could underscore this study's key findings.

Further, while previous management control studies have "predominantly focused on organizational-level variables and have ignored individual-level variables" (Adler and Chen 2011, p. 80), a particular strength of the study is to "address this limitation in prior literature" (Adler and Chen 2011, p. 80) by examining the outcomes of management controls and their use at the individual-level, i.e., at the level of the addressee of controls. However, this study must rely on perceptions of management control systems. Yet, both the conceptual framework of this study (Tessier and Otley 2012) and feedback theories (e.g., Ilgen, Fisher, and Taylor 1979; Taylor, Fisher, and Ilgen 1984) emphasize that it is the perception of management control systems that influences employees' (creative) response. In addition, as no industrial differences (e.g., service vs. non-service) regarding the results of the cross-industry, cross-departmental study, and in the design or use of controls (untabulated) are found, the results can likely be read in terms of the management control systems expected in each case (e.g., in a bank, controls should generally be used more intensively diagnostically than in an advertising agency), which contributes to the generalizability of the findings.

Avenues for potential further research encompass the notion of management control as a package (e.g., Grabner and Moers 2013; Simons 1995). Following this approach, Speklé, van Elten, and Widener (2017, p. 77) argue that controls “function as a synergetic package creating dynamic tension by balancing the levers”. More specifically, future studies should examine whether the presence of other controls and/or complementary uses, such as a diagnostically used code of conduct, enhances the positive effects of interactively used control designs, such as frequent feedback on creativity, through the provision of complementary guidance. More than that, this approach would offer mangers another lever for employee creativity in designing the management control package, as controls whose diagnostic use is less detrimental to employee creativity could take the place of diagnostic feedback in providing guidance, increasing the overall potential for employee creativity and hence the overall effectiveness of the control package.

In addition, my results suggest that there is great potential in the management control literature to explore the use of controls, such as diagnostic and interactive use, thereby broadening the view of how controls can be optimally designed. While previous literature has acknowledged its impact for some time (Adler and Chen 2011; Tessier and Otley 2012), scant empirical evidence exists on the effects of control use on control outcomes (e.g., Henri 2006).

6. Conclusion

Crisis-prone, rapidly changing markets, evolving technological advances, and shifting employee demands mandate and facilitate firms to overhaul their feedback systems. At the heart of this contemporary development stands feedback frequency (i.e., how often feedback is provided) to prevent feedback information from becoming obsolete or superseded by disruptive events, to stimulate employees, and to prevent employees from working ineffectively. Not surprisingly, then, latest recommendations for practiticioners (Deloitte 2017; PricewaterhouseCoopers 2015) focus on speeding up feedback cycles, echoing the long-held belief that “the more frequent the feedback, the better” (Ilgen, Fisher, and Taylor 1979, p. 354). However, nascent research on feedback frequency has recently shown that this simplification may fall short (e.g., Casas-Arce, Lourenco, and Martinez-Jerez 2017; Holderness, Olsen, and Thornock 2020) and cautions against the idea that more frequent feedback promotes employee performance in all respects.

Particularly when it comes to nurturing employee creativity, which is an essential organizational resource to remain competitive (Agnihotri et al. 2014a; Anderson, Potocnik, and Zhou 2014), firms face an as yet unresolved dilemma when deciding on the optimal level of feedback frequency due to the specific nature of the idea generation process (Amabile, Goldfarb, and Brackfield 1990; Speckbacher 2017). Frequent feedback equips employees with much-needed stimulation and direction for their creative pursuits, enabling them to continually reorientate on a journey whose destination is unforeseeable and to effectively deploy their (cognitive) resources and efforts. Simultaneously, however, frequent feedback may create a sense of being controlled, undermine flexibility, and compromise intrinsic motivation, which in turn erodes employees’ willingness to engage in creative endeavors. Consequently, the current demands placed on the feedback systems' frequency could prove to be a fatal trap for employee creativity.

Although feedback frequency has a long record of recognition as a pivotal component of management control system design (Luckett and Eggleton 1991) and of the feedback stimulus (Ilgen, Fisher, and Taylor 1979), the literature is silent on whether and how feedback frequency affects employee creativity. The ominous nature of this research gap takes on even greater urgency considering that the frequency of feedback is an inevitable managerial decision – even no feedback has a frequency. Thus, a major goal of this dissertation was to fill this research gap by examining the effects of feedback frequency on employee creativity to contribute to the ongoing academic discussion on the effectiveness of feedback frequency and to provide guidance to managers who yet may be missing out on their employees' creative potential. While the contributions of this dissertation to various strands of literature are discussed in more detail in the respective chapters of each study, I synthesize below the findings of its three empirical studies – two survey-based projects (n = 385; n = 400) and one field experiment (n = 105) – by first summarizing the key findings on the creative consequences of feedback frequency, second offering insights into the complex conversion mechanism of feedback frequency on employee creativity, third presenting critical factors that influence feedback frequency effectiveness, and finally providing implications for management and making suggestions for future research.

6.1 The Implications of Feedback Frequency for Employee Creativity

First and foremost, the three studies consistently show that feedback frequency enhances employee creativity. Thus, the results provide managers with an effective tool to promote the creativity of their employees. Hence, this dissertation complements recent publications that have pointed to the potential of feedback frequency to stimulate employee responses (e.g., Casas-Arce, Lourenco, and Martinez-Jerez 2017; Holderness, Olsen, and Thornock 2020) by being the first to reveal its effects on employee creativity. More specifically, utilizing the methodological advantages of (field) experimental studies, I can also demonstrate an unambiguous cause-effect relationship. Addressing the core of feedback literature, which conceptualizes different dimensions of feedback (e.g., Ilgen, Fisher, and Taylor 1979) that (in theory) can be designed individually, alternative explanations can be ruled out by controlling for dimensions that are often confounded with feedback frequency in practice (and in some previous research), namely feedback timing (Gilhooly 2016; Goomas, Smith and Ludwig 2011; Thornock 2016), feedback quantity (Coelho and Augusto 2010; Fedor and Buckley 1987; Hackman and Oldham 1975), and feedback valence (Kim and Kim 2020; Larson 1986; Zhou 1998). This allows to establish an interference-free cause-effect relationship between feedback frequency and employee creativity.

By considering feedback as an element of the management control system (Grabner and Speckbacher 2016; Luckett and Eggletion 1991; Tessier and Otley 2012), this dissertation contributes to shedding light on the “paradox” between creativity and control (Adler and Chen 2011; Speckbacher 2017) by showing that feedback frequency, a managerial choice in control design, positively stimulates employee creativity. While prior management control studies on creativity largely adopt a systems approach (Simons 1995; Speklé, van Elten, and Widener 2017), knowledge on the creative effects of particular controls is scarce, especially since only recently a “paradigm shift away from the traditional focus on established objectives and stable environments” (Adler and Chen 2011, p. 63) has become an impetus for research on the long-asserted incompatibility of control (e.g., feedback) and employee creativity (e.g., Speklé, van Elten, and Widener 2017; Davila and Ditillo 2017). Thus, also on control level, “paradoxically, creativity can flourish in the presence of control” (Speklé, van Elten, and Widener 2017, p.73).

While the studies base the utilized creativity measures on established scales (Lages and Piercy 2012; Farmer, Tierney, and Kung-McIntyre 2003), I also follow the procedure of Kachelmeier, Wang, and Williamson (2019) to explicitly observe the effects of feedback frequency on the quality of the generated ideas (i.e., high-creativity ideas). The results reveal that feedback frequency positively affects the quality of creativity, and therefore complement recent management accounting research on antecedents of employees’ high-creativity idea generation, by adding feedback frequency to the short list of known determinants (Kachelmeier, Reichert, and Williamson 2008; Kachelmeier, Wang, and Williamson 2019). This is a crucial aspect of feedback frequency as in practice only a fraction of all ideas generated can be implemented in organizations due to resource constraints, and it is particularly unlikely for managers to get their own sense of how to encourage the production of high-creativity ideas, since quality responses are often not initial and cannot be directly incentivized (Kachelmeier, Wang, and Williamson 2019). In addition, due to the specific feedback design, I also address recent calls on investigating the effectiveness of “frequent relative performance information (RPI) feedback, [as] research on the effectiveness of such feedback is limited” (Holderness, Olsen, and Thornock 2020, p. 156).

6.2 Translation Mechanisms of Feedback Frequency

Second, this dissertation takes a look behind the façade of how employees perceive feedback frequency to understand how it changes their creative response.

This not only allows unlocking the black box of intrapersonal processes that link frequent feedback stimuli to behavioral responses, but the moderated mediation research design also facilitates singling out a number of distinct functions of feedback (Ilgen, Fisher, and Taylor 1979). Consistent with the multiple ways in which feedback impacts are described by the broader feedback literature (cf. Agnihotri et al. 2014a; Challagalla and Shervani 1996; Coelho and Augusto 2010; Deci, Connell, and Ryan 1989; Ilgen, Fisher, and Taylor 1979; Mullins, Agnihotri, and Hall, 2020), *information*, *motivation*, *controlling*, and *explanation functions* of feedback frequency for employees are identified that correspond to the complex translation mechanism of feedback frequency into employee creativity. Interestingly, information and motivation functions apply to both customer and supervisor feedback frequency, while controlling and explanation functions are present only if the supervisor is the feedback source.

To start with, role theory (Katz and Kahn 1978) guides me to recognize the importance of the role stressors role conflict and role ambiguity as pivotal mediators of the relationship between feedback frequency and creativity (Coelho, Augusto, and Lages 2011; Solomon et al. 1985). Feedback frequency is identified as powerful lever to reduce employee's role ambiguities (*information function*). By considering the multifacetedness of role ambiguity (Singh 1993), I find that both supervisor and customer feedback frequency help to clarify the expectations of the respective employee's counterpart, but surprisingly not across them, strengthening the point that "feedback is likely to have *localized* effects" (Singh 1993). However, as the research design also allows for feedback interplay, this dissertation provides first empirical evidence for conceptual considerations (Agnihotri et al. 2014a; Challagalla and Shervani 1996; Wilder, Collier, and Barnes 2014) that frequent supervisor feedback enables employees to better understand customer feedback by reducing delay, short-term memory losses between respective feedbacks and strengthening the ability of sensemaking (*explanation function*), supporting recent publications to position leader influence “as a facilitator of the customer-employee interactions” (Dong et al. 2015, p.1373). Contrary to expectations that more frequent feedback originating from customer and supervisor pursuing their own agenda (Yoo and Arnold 2016) provokes inter-sender conflicts, the results do not show a consistently negative relation.

While there seems to be consensus in the literature on the substantial implications of role conflict and role ambiguity for employee creativity (e.g., Bettencourt and Brown 2003; Coelho, Augusto, and Lages 2011; Schepers, Nijssen, and van der Heijden 2016), the results allow for a more nuanced understanding of previous findings by showing that these effects change with the level of the frequency of feedback. Regarding the implications of creativity-stimulating role conflict, which represents a situation in which employees generally have to decide which role partner’s needs are given preferential consideration, the results show that this positive relation might turn into a negative relation at high levels of supervisor feedback frequency (*controlling function*). This evidence thereby substantiates theoretical considerations that higher supervisor feedback frequency puts employees into a behavioral corset and thereby hinders creativity as “feedback is all too frequently controlling; it emphasizes how people should behave and implies that the manager is in control” (Deci, Connell, and Ryan 1989, p.585), complementing research on leader behaviors and employee creativity (cf. Amabile et al. 2004). Furthermore, whereas previous research (e.g., Bettencourt and Brown 2003; Coelho, Augusto, and Lages 2011) consistently acknowledges the negative influence of role ambiguity on the employee’s search for novel ideas, the results show that this negative relationship is magnified by increasing levels of feedback frequency (*motivation function*). Drawing on notions of the ability of feedback to foster intrinsic motivation (e.g., Coelho and Augusto 2010), it can be concluded that employees who are not motivated to generate novel ideas by sufficiently frequent feedback will not engage in creative behaviors regardless of the level of any of their role ambiguity facets.

Moreover, supplemental analyses, drawing on feedback research suggestions, provide further insights into the mechanisms by which feedback frequency operates. I find that recipients perceive more frequent feedback as additional effort from their supervisors and as a greater disruption to their daily routine. This suggests that frequent feedback induces motivation through reciprocity (Fedor and Buckley 1987; Gouldner 1960), and, consistent with theoretical considerations (Ilgen, Fisher, and Taylor 1979; Kluger and DeNisi 1996), that a frequent feedback stimulus is more salient and less likely to be overshadowed by other stimuli in the recipient's environment. Surprisingly, I also find that high levels of feedback frequency help employees process the feedback in units they can digest better. This result contrasts with previous lines of reasoning that claim frequent feedback causes information overload, e.g., Lam et al. (2011) and Holderness, Olsen, and Thornock (2020), who, however, do not control for the quantity of feedback.

6.3 Determinants of Feedback Frequency Effectiveness

Third, since the studies in this dissertation draw on multiple theories to explore the creative effects of feedback frequency, a plethora of factors that influence the effectiveness of feedback frequency emerges. These various viewpoints and factors offer major and specific contributions to the feedback, service, and management control literature, but also to gender and leadership research, that will be elaborated in the next paragraphs. In addition, these factors are also of interest to the feedback literature on an abstract level, as they pave the way to fuse two opposing theoretical camps that have argued for both a positive (e.g., Ilgen, Fisher, and Taylor 1979;

Holderness, Olsen, and Thornock 2020) and a negative (e.g., Casas-Arce, Lourenco, and Martínez-Jerez 2017; Deci, Connell, and Ryan 1989) relationship between feedback frequency and employee performance, respectively. In other words, if managers are mindful of these factors when designing feedback, they can determine whether frequent feedback promotes or inhibits employee performance (i.e., creativity). In brief, I find use of feedback (interactive and diagnostic), feedback recipient characteristics (employee experience), feedback source (customer or supervisor), and feedback source characteristics (gender) moderate the implications of feedback frequency on employee creativity.

Drawing from a management control perspective on feedback frequency, I find, consistent with conceptual considerations (Tessier and Otley 2012), that *interactive feedback use* and *diagnostic feedback use* (i.e., control use) are critical to employee's response to feedback frequency (i.e., control design) and thus to their creativity (i.e., employee effectiveness). Thereby, I am first to operationalize and empirically validate Tessier and Otley's (2012) framework of management control systems, and respond to a recent call in management accounting research to examine "a rich literature on various control mechanisms and their use (e.g., diagnostic versus interactive), […] to make important contributions to research on the management of creativity" (Speckbacher 2017, p. 51). More specifically, this dissertation empirically demonstrates that feedback, when used interactively, indeed "provides a focus for employees' creative juices" (Speklé, van Elten, and Widener 2017, p. 78) and "encourage[s] new ideas to emerge" (Adler and Chen 2011, p. 76) by amplifying the positive impact of feedback frequency on employee creativity. However, the results disagree with prior beliefs that the diagnostic use of controls positively affects employee creativity (Cools, Stouthuysen, and Van den Abbeele 2017; Speklé, van

Elten, and Widener 2017), as diagnostic use of feedback attenuates the positive effects of feedback frequency on employee creativity and fails to create a dynamic tension with the interactive use of feedback that stimulates creativity (Henri 2006). On an abstract level, this dissertation hence extends conceptual research on management control systems suggesting that for their overall management controls system, firms “do not need to make tradeoffs between whether they want a creative organization or an organization characterized by control” (Speklé, van Elten, and Widener 2017, p. 75), as the results show that this does not hold true at the control level and that firms nevertheless have to counterbalance the benefits of diagnostic control use with its creative harm.

In addition, this dissertation contributes to contingency-based management control research (Chenhall 2003) by substantiating previous findings that creative requirements on employees influence managers’ decisions in designing control system (Grabner 2014; Grabner and Speckbacher 2016), such as the frequency with which feedback is provided to employees. However, these results are extended by revealing that the level of creativity expected of employees in their job also determines managers’ use of controls, such that higher (lower) creative demands encourage managers to use feedback more (less) interactively and less (more) diagnostically. Thus, managers do not generally use controls more intensively when they expect employees to be creative (Speklé, van Elten, and Widener 2017), but rather selectively change their control use pattern by using feedback more interactively but less diagnostically, which also proves to be optimal for fostering employee creativity.

Next, I find that *employee experience* moderates feedback frequency effectiveness. Despite employee experience is not explicit part of the underlying theory (Katz and Kahn 1978) that determined research model design, state-of-the art procedures were applied to identify unobserved heterogeneity. More specifically, the joint analysis with FIMIX-PLS and PLS-POS according to Hair et al. (2016) and Matthews et al. (2016) uncovered the importance of employee experience in the influence of feedback frequency on employee creativity. To the best of my knowledge, no study has yet and fully completed the process of transforming unobserved heterogeneity “into observed heterogeneity for future studies”[153] (Becker et al. 2013, p. 668). Surprisingly, the relevance of the frequency of feedback as a driver of employee creativity does not diminish with more experience, as key conceptual feedback studies indicate (Anseel et al. 2015; Ilgen, Fisher, and Taylor 1979); rather, more experienced employees approach creative activities in different ways, and different sources of feedback change their meaning to employee creativity as they develop their own experiences. Nevertheless, these results are consistent with previous creativity studies (e.g., Agnihotri et al. 2014a; Amabile et al. 2004), which posit that employee creativity depends on individual domain- and creativity-relevant skills and abilities that, according to human capital theory (Becker 1964), change with years of experience. Interestingly, these findings also refute the simplistic stereotype that more experienced employees are less creative (Ng and Feldman 2008, 2013).

Furthermore, based on the consideration that the supervisor is not the only source of feedback for an employee and that employees are constantly exposed to multiple feedback sources (Greller and Herold 1975), the discussion of feedback frequency effectiveness is enriched by observing the simultaneous influence of *customer feedback frequency* and *supervisor feedback frequency* on employee creativity. In doing so, this dissertation goes beyond previous feedback research, as

[153] See chapter 2.1.2.4 for an exemplary compendium of alternative applications of these procedures in research practice, which are, however, insufficient in terms of their methodological accuracy.

most studies concentrate on the supervisor as feedback source and are also limited to only one source of feedback (see the reviews from Alvero, Bucklin, and Austin 2001 and Lechermeier and Fassnacht 2018; for an exception, see Siahtiri 2018). The results show that feedback frequency from customers not only impacts employee creativity, but its effects are inverse to supervisor feedback frequency considering employee's level of experience and, as previously mentioned, lacks controlling and explanation function. More specifically, among (in)experienced employees, customer feedback frequency (enhances) attenuates employee creativity. This contradicts the intuitive assumption that employees need a baseline understanding of internal processes and have to find their way around their jobs before they can translate external stimuli from outside the organizational unit into creative ideas. Even more, the decision to introduce the customer as a complementary feedback source allows to detect an interaction effect between customer and supervisor feedback (see previous section on explanation function), elevating the importance of a multi-feedback-source approach and raises the suspicion that the extensive research on supervisor feedback effectiveness (see Alvero, Bucklin, and Austin 2001 and Lechermeier and Fassnacht 2018) is biased at the level of customer feedback frequency.

By examining the joint effect of supervisor feedback frequency and customer feedback frequency on the creativity of employees who provide services to customers (i.e., they are service employees), this dissertation also contributes to a large strand of literature that centers around fostering service employee creativity (for an overview see Schepers, Nijssen, and van der Heijden 2016) and assigns "utmost importance" (Coelho and Augusto 2010, p.427) to understanding service employee creativity and its antecedents. Furthermore, incorporating the customer as a provider of feedback source in the research design goes beyond previous studies in this literature stream that have only recognized a single feedback source (Agnihotri et al. 2014a; Coelho and Augusto 2010; Johlke and Duhan 2000; for an exception, see Siahtiri 2018), despite service literature has made much of involving the customer in the service generation process and from a service-dominant logic perspective in particular, it seems mandatory and inherently appropriate for the service employee to receive feedback from his or her customers (Vargo and Lusch 2008a). Consequently, the results add more color to unambiguously positive arguments in service research suggesting that “codesign fosters creative thinking” (Trischler et al. 2018, p.89), such as that present in the literature on customer involvement, customer participation, and co-creation (Chang and Taylor 2016; Storey and Larbig 2018), and echoes recent indications of more nuanced and cautious customer engagements in the service process (e.g., Brodie et al. 2011; Cabiddu, Frau, and Lombardo 2019). This study uncovers the risks involved when translating general conceptual considerations which argue that “the customer is always a co-creator of value” (FP6, Vargo and Lusch 2008a, p. 7) into allowing customers to provide frequent feedback to every service employee. Because the conceptual model singles out creativity as one element of the co-creation process, the results allow for a clearly defined addition to previously documented value co-destructive social interactions (Cabiddu, Frau, and Lombardo 2019).

As an additional moderator of the creativity effects of feedback frequency, *gender* is introduced as a critical characteristic of the feedback source. I was interested in better understanding the tension that arise from gender and gender (in)congruence since feedback and creativity are highly individual-specific processes, and we know from a number of publications that gender influences the way of perceiving (feedback source) and processing (feedback recipient) information (e.g., Carli 2010; Eagly and Karau 1991). Contrary to gender role theory predictions, I unexpectedly discover that recipient gender and gender (in)congruence are irrelevant to the relationship between feedback frequency and recipient creativity. However, I find feedback from a female supervisor to be more powerful, and surprisingly so regardless of the recipient's gender, which contradicts suggestions from feedback theory that male leaders are perceived as more competent, increasing their feedback frequency effectiveness, but is consistent with arguments from gender role theory that female behaviors are more likely to be perceived as being helpful (Eagly and Karau 1991; Eagly and Wood 2012; Lanaj and Hollenbeck 2015).

Although "these [gender] roles are salient to the extent that other social roles, particularly family and employment roles, are relatively unimportant." (Eagly and Karau 1991, p. 686) and feedback research has repeatedly emphasized the importance of the feedback source characteristics (e.g., Holderness, Olsen, and Thornock 2017; Steelman and Rutkowski 2004), gender has received scant attention so far (see Alvero, Bucklin, and Austin 2001 and Lechermeier and Fassnacht 2018). Filling this gap in feedback literature gains importance against the backdrop of an increasing number of women in leadership positions (e.g., Bobe and Kober 2020; Lai et al. 2017) and the tendency of women to view themselves as less suited and effective for leadership roles (Paustian-Underdahl, Walker, and Woehr 2014), which strengthens their case for occupying such a role, especially if their job description includes fostering employee creativity.

Beyond, the results also add to an ongoing discussion about leadership effectiveness, since the provision of feedback is an inherent duty of leaders and a powerful lever for influencing employee outcomes (Shea and Howell 1999). Fundamentally, the findings dovetail with an ongoing discussion of gender differences in leadership effectiveness that has evolved from a more descriptive approach to gender differences in leadership styles (Paustian-Underdahl, Walker, and Woehr

2014). More specifically, I fill gaps in leadership research on creativity by demonstrating the positive influence of supervisor feedback frequency on employee creativity and the moderating effect of leader gender on this relationship. In doing so, this dissertation echos Hughes et al. (2018, p. 564), whose comprehensive review of the effects of leadership on creativity and innovation urges future research to "move beyond the current focus on leader styles to explore the effects of leader characteristics [and] behaviors". Whereas the approach of examining feedback frequency as particular leadership behavior is novel, the results on the influence of gender on leadership effectiveness contrast previous findings (Reuvers et al. 2008; Wang et al. 2013), in which, however, the research settings were biased toward male leaders, suggesting that the widely accepted assumption that male leaders are more effective in enhancing employee creativity needs to be reevaluated.

6.4 Managerial Implications

From a managerial perspective, the three studies provide a plethora of practical propositions that advocate the positive potential inherent to the frequency of feedback as an effective management lever for elevating employee creativity. This is important for any supervisor to recognize, as it is an inevitable decision regarding how often feedback should be provided, and as fostering employee creativity is essential for organizations to remain competitive (Agnihotri et al. 2014a; Anderson, Potocnik, and Zhou 2014).

In practice, supervisors refrain from providing feedback more frequently due to their formal authority (De Stobbeleir, Ashford, and Buyens 2011) especially when they strive to foster employee creativity, as previous literature points to negative consequences resulting from evaluations when employees are engaged in creative activities (e.g., Sagiv et al. 2010; Shalley, Zhou, and Oldham 2004). Moreover, realizing the time-pressured work contexts that many supervisors live in, they may hesitate to provide feedback more often because of resource restrictions (e.g., time, cognitive capability). My findings update this commonly held view on feedback frequency and offer guidance for supervisors with limited time resources and heterogeneous subordinates.

Most importantly, supervisors can directly control employee creativity by adjusting the frequency of feedback they provide to their employee. Thus, supervisors command a powerful design element of their control package, as more frequent feedback leads to more creativity. This particularly applies to female feedback sources, who are stereotypically denied a natural leadership role. Companies should therefore empower female supervisors and give them a voice – to provide feedback to employees frequently.

Nonetheless, the extent of the positive effect of feedback frequency on employee creativity depends on how supervisors use the feedback. When supervisors use feedback as an invitation to discuss with employees and provide them with information across hierarchical levels and departmental boundaries (interactive feedback use), it elevates the effects of feedback frequency; whereas feedback arising from deviations from target values (diagnostic feedback use) reverses its positive effects, as employees feel monitored and may avoid risks related to generating creative ideas. In offering a way out of supervisors' dilemma of limited capacities, the results suggest that rather than investing more effort in using feedback to fostering employee creativity, supervisors can shift their engagement from a diagnostic use of feedback to a more interactive use.

This resource-neutral recommendation is complemented by an efficient solution for supervisors to increase their feedback frequency, as the findings object to the notion that more frequent feedback must mean more feedback quantity. Thus, when supervisors increase the frequency of their feedback, breaking the same amount of feedback information into more “digestible” units and present them over multiple points in time is quite sufficient to stimulate employee creativity. From a strategic point of view, this insight advises, especially in these days of exponentially increasing data availability, for example through Big Data or real-time data sources, not to inflate feedback systems, but to slim them down in line with requirements and instead to promote and maintain frequent performance communication.

The implications for how to best exploit employees’ creative potential extend to confronting appropriate employees with frequent feedback. Surprisingly, the results show that recipient gender and feedback valence do not condition the effectiveness of feedback frequency on creativity. Thus, supervisors should neither shrink from providing frequent feedback if it is negative, as it does not harm employee creativity, nor exclude lower-performing employees from these creative stimuli, since they are equally capable of generating ideas. In addition, supervisors should also not forgo frequent feedback for fear of intimidating female employees (as a male supervisor) or that the feedback effect will fizzle out on male employees (as a female supervisor), as gender theory suggests.

However, the results clearly guide supervisors to concentrate feedback capacities on experienced employees. Not only does this recommendation require no additional effort on the part of supervisors, as they need to provide feedback to inexperienced employees correspondingly less frequently, and is thus resource-neutral, but it also optimizes the aggregate creativity level in a twofold manner: It liberates employees whose creativity is hampered by frequent feedback exchanges and nourishes employees who are responsive to frequent feedback stimuli. Therefore, to provide frequent feedback to employees effectively, it is essential for supervisors to be aware of the employee’s experience.

The results also help supervisors find better solutions for shaping employees’ feedback environment, since customer feedback activities have also been found to contribute to employee creativity. Surprisingly, these are only beneficial for a selection of employees. Counterintuitively, supervisors should motivate less experienced employees to get into contact with their customers, and vice versa, by breaking down organizational and technological barriers that restrict customer exchange (Giebelhausen et al. 2014), cultivating a feedback-seeking climate (Auh et al. 2019) towards customers, and ensuring sufficient resources, education, and information on how to enquire for feedback in the most appropriate manner (Auh et al. 2019).

6.5 Avenues for further Research

The studies and findings of this dissertation open up avenues for further research. To begin with, feedback research has repeatedly noted that feedback dimensions interact to influence the effectiveness of feedback (e.g., Ilgen, Fisher, and Taylor 1979; Kinicki et al. 2004). It is not presumptuous, but reasonable, to assert that the feedback dimensions that have been identified as influential on the creative impact of feedback frequency represent an incomplete list. From the theoretically innumerable possible combinations, inspired from gender role theory, I would like to focus briefly on the type of feedback information. Despite the fact that extant gender research has emphasized that “feedback has different effects for females and males” (Wozniak 2012, p. 170), women desire more frequent communication, and are more receptive to feedback (Burton and Hoobler 2006; Sturm et al. 2014; Röhner and Schütz 2020), the results fail to show that recipient gender makes a difference to the creative effects of feedback frequency. It is suspected that this is partly due to the relative performance information provided to participants, since women perform relatively worse in competitive environments (e.g., Gneezy and Rustichini 2004), especially when – as in case of my study – men are part of the reference group (Kuhnen and Tymula 2012), so that these two effects cancel each other out. I therefore suggest further research in the influence of recipient gender on feedback frequency effectiveness, for instance, by providing absolute feedback or varying the composition of the reference group (e.g., only women). Similarly, testing for the attributional ambiguity account (Biernat and Danaher 2012), when subjective feedback is provided, may extend my findings.

Second, this dissertation focused on two key sources of feedback that are close to employees, namely supervisor and customer, which implies that it is still unkown how the feedback dynamics would evolve in the presence of additional feedback sources, for example, by recognizing different types of customers or internal counterparts. Along these lines, a promising research impulse springs from recent studies in information technology journals (e.g., Krancher, Luther, and Jost 2018) that deal with automated feedback sources to reflect the ongoing digitization revolution. Automated feedback sources might compete with my emphasis on human feedback sources as their speed and frequency, ubiquitous availability, but also their inherent inability to build inter-human relationships distinguish human and machine feedback sources, especially since I find that feminine attributes and mechanisms of interpersonal reciprocity play a vital role in how the feedback stimulus is transformed into a creative employee response. Therefore, one avenue for future research might be to juxtapose in-person human feedback with remote (via internet, email, or video-call) human, quasi-human, and machine feedback to explore their impact on "the last human stronghold" (Beniaminy 2020) of creativity. In parallel, another avenue might be to explore how inhuman feedback sources can simulate and adopt positive human characteristics. At the most rudimentary level, giving them a feminine appearance (e.g., voice or name), as is already the case with popular assistants such as Alexa (Amazon), Cortana (Microsoft), Google Assistant (Google), or Siri (Apple) (Abercrombie et al. 2021; Loideain and Adams 2020), could be beneficial for the creative effects of (frequent) feedback from automated anonymous digital feedback sources, interactive AI-driven feedback systems, or even simpler inhuman performance measurement systems.

Third, a more holistic view of the innovation process and the preceding creativity process, the results of which serve as input to the innovation process (Hughes et al. 2018), raises the question of which creative ideas initiate innovations and benefit the organization, as creative ideas per se are only crude material whose potential has yet to be extracted though their implementation. Still, with respect to better understanding how to stimulate creativity, it can be argued to be less important whether feedback frequency induces an objectively novel idea and more important whether employees invest cognitive and emotional resources and are enabled to think outside “his or her box”. In addition, I find that feedback frequency raises the quality of generated ideas, that should particular offer potential for organizations to succeed (Kachelmeier, Reichert, and Williamson 2008). Nevertheless, firms face the threat of (valuable) creative ideas becoming ineffective if they fail at the hurdle of selection for implementation[154] or if innovations derived from a creative idea do not benefit the organization, which is referred to as the “creativity/innovation maximization fallacy” (Anderson, Potocnik, and Zhou 2014; Kimberly 1981). The notion that the generation of ideas can be stimulated by frequent feedback from multiple sources may offer a starting point for future research to overcome this fallacy: Ideas that emerge from customer feedback should have some degree of difference from those that emerge

[154] Not surprisingly, a strand of creativity literature is devoted studying the process of selecting the ideas of highest quality from the sea of ideas generated (e.g., Faure 2004; Putman and Paulus 2009; Rietzschel, Nijstad, and Stroebe 2006, 2010).

from supervisor feedback, given that both sources observe (and thus provide feedback on) different employee behaviors. These ideas, in turn, address more likely issues that relate back to either of the sources. Since innovation is an inter-personal process (Hughes et al. 2018) and its outcome does not benefit all parties equally (Anderson and King 1993), it is worth investigating whether supervisors, who pursue a different agenda than customers according to the “three-cornered fight” (Bateson 1985), exploit their hierarchical power and pull strings within the organization, causing ideas originating from frequent supervisor feedback to have a higher probability of succeeding in the innovation process; and whether these ideas have a similarly positive impact on the organization as ideas that originate from frequent customer feedback, from whose implementation potentially numerous customers could benefit. Future studies are thus encouraged to take a more comprehensive perspective on the interaction between feedback frequency and the innovation process.

Finally, the view of management control as a package (e.g., Grabner and Moers 2013; Simons 1995) holds great potential for further investigations. My results demonstrate that managers need to counterbalance the creativity effects of a particular control (e.g., feedback) with some of its controlling effects. In particular, managers have to forgo the (otherwise) beneficial effects of diagnostic feedback use if they want to optimize their employees’ creativity. A management control package approach would mean identifying another control that substitutes for these lost exemplarily positive effects, but whose use is less detrimental to employee creativity. While the findings of Speklé, van Elten, and Widener (2017) implicitly show that firms engage in such counterbalancing activities, there is hardly any evidence on their optimality. Even more, previous research suggests interactive effects between controls (Speklé, van Elten, and Widener 2017), but research on this potential to efficiently elevate employee creativity is still sparse (for an exception see Henri 2006). The paucity of research on the interactive effects of controls on employee creativity becomes even more apparent when reflecting on the fact that my study clearly indicates that control use is pivotal to the creative effects of controls, considerably complexifying the struggle to determine the ideal design of the control package. Consequently, the interplay of controls and their use opens up a wide field for researchers to make important contributions to our current understanding of management control effectiveness, and provides managers with a lever to increase employee creativity and the overall effectiveness of the control package.

References

Abercrombie, Gavin, Amanda Cercas Curry, Mugdha Pandya, and Verena Rieser (2021), “Alexa, Google, Siri: What are Your Pronouns? Gender and Anthropomorphism in the Design and Perception of Conversational Assistants,” *arXiv preprint arXiv:2106.02578*.

Abernethy, Margaret A. and Peter Brownell (1999), “The Role of Budgets in Organizations facing strategic Change: An exploratory Study,” *Accounting, Organizations and Society*, 24 (3), 189-204.

Adler, Paul S., and Bryan Borys (1996), “Two Types of Bureaucracy: Enabling and Coercive,” *Administrative Science Quarterly*, 41 (1), 61-89.

------, and Clara Xiaoling Chen (2011), “Combining Creativity and Control: Understanding individual Motivation in large-scale collaborative Creativity,” *Accounting, Organizations and Society*, 36 (2), 63-85.

Agnihotri, Raj, Adam A. Rapp, James “Mick” Andzulis, and Colin B. Gabler (2014a), “Examining the Drivers and Performance Implications of Boundary Spanner Creativity,” *Journal of Service Research*, 17 (2), 164-81.

------, Kevin J. Trainor, Michael T. Krush, and Sukumarakurup Krishnakumar (2014b), “Satisfied and Productive Boundary Spanners: A Model of Resiliency and Customer Expectations,” *Journal of Services Research*, 14 (2), 57-74.

Aguinis, Herman, James C. Beaty, Robert J. Boik, and Charles A. Pierce (2005), “Effect Size and Power in Assessing Moderating Effects of Categorical Variables Using Multiple Regression: A 30-year Review,” *Journal of Applied Psychology*, 90 (1), 94-107.

Aiken, Leona S. and Stephen G. West (1991), *Multiple Regression: Testing and Interpreting Interactions*. Newbury Park, CA: SAGE.

Akaike, Hirotogu (1973), “Information Theory and an Extension of the Maximum Likelihood Principle,” in *Second International Symposium on Information Theory*, B.N. Petrov, and F. Csáki, eds. Budapest: Akadémiai Kiadó, 267-81.

Allred, Sarah R., L. Elizabeth Crawford, Sean Duffy, and John Smith (2016), “Working Memory and Spatial Judgments: Cognitive Load Increases the Central Tendency Bias,” *Psychonomic Bulletin & Review*, 23 (6), 1825-31.

Alvero, Alicia M., Barbara Bucklin, and John Austin (2001), “An Objective Review of the Effectiveness and Essential Characteristics of Performance Feedback in Organizational Settings (1985-1998),” *Journal of Organizational Behavior*, 21 (1), 3-29.

Amabile, Teresa M. (1983), “Social Psychology of Creativity: A Componential Conceptualization,” *Journal of Personality and Social Psychology*, 45 (2), 357-76.

------ (1988), “A Model of Creativity and Innovation in Organizations,” in: *Research in Organizational Behavior*, Vol. 10, B. M. Staw, and L. L. Cummings, eds. Greenwich, CT: JAI Press, 123-67.

------ (1996), *Creativity in Context: Update to the Social Psychology of Creativity*. Boulder, CO: Westview Press.

------, Phyllis Goldfarb, and Shereen C. Brackfield (1990), “Social Influences on Creativity: Evaluation, Coaction, and Surveillance,” *Creativity Research Journal*, 3, 6-21.

------, Elizabeth A. Schatzel, Giovanni B. Moneta, and Steven J. Kramer (2004), “Leader Behaviors and the Work Environment for Creativity: Perceived Leader Support,” *The Leadership Quarterly*, 15 (1), 5-32.

Amaratunga, Dilanthi, David Baldry, Marjan Sarshar, and Rita Newton (2002), "Quantitative and qualitative Research in the Built Environment: Application of "mixed" Research Approach," *Work Study*, 51 (1), 17-31.

Ammons, Robert B. (1956), "Effects of Knowledge of Performance: A Survey and Tentative Theoretical Formulation," *The Journal of General Psychology*, 54 (2), 279-99.

Anderson, Kristin J., and Campbell Leaper (1998), "Meta-Analyses of Gender Effects on Conversational Interruption: Who, What, When, Where, and How," *Sex Roles*, 39 (3), 225-52.

Anderson, Neil and Nigel King (1993), "Innovation in Organizations," in: *International Review of Industrial and Organizational Psychology*, C. L. Cooper and I. T. Robertson, eds. London, U.K.: Wiley, 1-34.

------, Kristina Potočnik, and Jing Zhou (2014), "Innovation and Creativity in Organizations: A state-of-the-science Review, Prospective Commentary, and Guiding Framework," *Journal of Management*, 40 (5), 1297-333.

Andersson, Ulf, Alvaro Cuervo-Cazurra, and Bo Bernhard Nielsen (2014), "From the Editors: Explaining Interaction Effects within and across Levels of Analysis," *Journal of International Business Studies*, 45, 1063-71.

Andiola, Lindsay M (2014), "Performance Feedback in the Audit Environment: A Review and Synthesis of Research on the Behavioral Effects," *Journal of Accounting Literature*, 33 (1-2), 1-36.

Anseel, Frederik, Adam S. Beatty, Winny Shen, Filip Lievens, and Paul R. Sackett (2015), "How are We Doing After 30 Years? A Meta-Analytic Review of the Antecedents and Outcomes of Feedback-Seeking Behaviour," *Journal of Management*, 41 (1), 318-48.

Antonetti, Paolo, Benedetta Crisafulli, and Stan Maklan (2018), "Too good to be true? Boundary Conditions to the Use of downward Social Comparisons in Service Recovery," *Journal of Service Research*, 21 (4), 438-55.

Archer, John, and Barbara Lloyd (2002), *Sex and Gender (2nd ed.)*. Cambridge: Cambridge University Press.

Armstrong, J. Scott and Terry S. Overton (1977), "Estimating Nonresponse Bias in Mail Surveys," *Journal of Marketing Research*, 14 (3), 396-402.

Auh, Seigyoung, Bulent Menguc, Pinar Imer, and Aypar Uslu (2019), "Frontline Employee Feedback-Seeking Behavior: How is It Formed and When Does It Matter?," *Journal of Service Research*, 22 (1), 44-59.

Axtell, Carolyn M., David J. Holman, Kerrie L. Unsworth, Toby D. Wall, Patrick E. Waterson, and E. Harrington (2000), "Shopfloor Innovation: Facilitating the Suggestion and Implementation of Ideas," *Journal of Occupational and Organizational Psychology*, 73 (3), 265-85.

Baer, Markus, Roger T. A. Leenders, Greg R. Oldham, and Abhijeet K. Vadera (2010), "Win or Lose the Battle for Creativity: The Power and Perils of intergroup Competition," *Academy of Management Journal*, 53 (4), 827-45.

Bagozzi, Richard P. (1981), "Evaluating structural Equation Models with unobservable Variables and Measurement Error: A Comment," *Journal of Marketing Research*, 18 (3), 375-81.

------ (1984), "A Prospectus for Theory Construction in Marketing," *Journal of Marketing*, 48 (1), 11-29.

------ (2011), "Measurement and Meaning in Information Systems and Organizational Research: Methodological and Philosophical Foundations," *MIS Quarterly*, 35 (2), 261-92.

------, and Youjae Yi (1988), "On the Evaluation of Structural Equation Models," *Journal of the Academy of Marketing Science*, 16 (1), 74-94.

------, Youjae Yi, and Lynn W. Phillips (1991), "Assessing Construct Validity in Organizational Research," *Administrative Science Quarterly*, 36 (3), 421-58.

Baier, Alexandra, Brent J. Davis, Tarek Jaber-Lopez, Michael Seidl, (2018), "Gender, Competition and the Effect of Feedback and Task: An Experiment," *Working Paper Forschungsförderung*, 62, Hans-Böckler-Stiftung, Düsseldorf.

Balafoutas, Loukas, Rudolf Kerschbamer, and Matthias Sutter (2012), "Distributional Preferences and competitive Behavior," *Journal of Economic Behavior & Organization*, 83 (1), 125-35.

Barclay, Donald, Christopher Higgins, and Ronald Thompson (2015), "The partial least squares (PLS) Approach to casual Modeling: Personal Computer Adoption and Use as an Illustration," *Technology Studies*, 2 (2), 285-309.

Baron, Reuben M., and David A. Kenny (1986), "The Moderator-Mediator Variable Distinction in Social Psychological Research: Conceptual, Strategic and Statistical Considerations," *Journal of Personality and Social Psychology*, 51 (6), 1173-82.

Bass, Bernard M. (1985), "Leadership: Good, better, best," *Organizational Dynamics*, 13 (3), 26-40.

Bateman, Thomas S., and Dennis W. Organ (1983), "Job Satisfaction and the Good Soldier: The Relationship between Affect and Employee "Citizenship"," *Academy of Management Journal*, 26 (4), 587-95.

Bateson, John E. G. (1985), "Perceived Control and the Service Encounter," in *The Service Encounter: Managing Employee/Customer Interaction in Service Businesses*, John A. Czepiel, Michael R. Solomon, and Carol F. Surprenant, eds. Lexington, Mass.: Lexington Books, 67-82.

Baxter, Ryan J., Kip Holderness, and David A. Wood (2016), "The Effects of Gamification on Corporate Compliance Training: A Field Experiment of True Office Anti-Corruption Training Programs," Available at SSRN: https://ssrn.com/abstract=2766683 or http://dx.doi.org/10.2139/ssrn.2766683.

Bear, Julia B., Lily Cushenbery, Manuel London, and Gary D. Sherman (2017), "Performance Feedback, Power Retention, and the Gender Gap in Leadership," *The Leadership Quarterly*, 28 (6), 721-40.

Bechtel, Nathan T., Heather M. McGee, Bradley E. Huitema, and Alyce M. Dickinson (2015), "The Effects of the Temporal Placement of Feedback on Performance," *The Psychological Record*, 65 (3), 425-34.

Becker, Gary S. (1964), *Human Capital: A Theoretical and Empirical Analysis with Specific Reference to Education*. New York: Columbia University Press.

Becker, Jan-Michael, Kristina Klein, and Martin Wetzels (2012), "Hierarchical Latent Variable Models in PLS-SEM: Guidelines for Using Reflective-Formative Type Models," *Long Range Planning*, 45 (5-6), 359-94.

------, Christian M. Ringle, Marko Sarstedt, and Franziska Völckner (2015), "How Collinearity Affects Mixture Regression Results," *Marketing Letters*, 26 (4), 643-59.

------, Arun Rai, Christian M. Ringle, and Franziska Völckner (2013), "Discovering Unobserved Heterogeneity in Structural Equation Models to Avert Validity Threats," *MIS Quarterly*, 37 (3), 665-94.

Bell, Simon J. and James A. Luddington (2006), "Coping with Customer Complaints," *Journal of Service Research*, 8 (3), 221-33.

Beniaminy, Israel (2020), "Creativity – The Last Human Stronghold?," *The Future of Things* (available at https://thefutureofthings.com/3695-creativity-the-last-human-stronghold [retrieved 03-21]).

Berger, Samuel M., and Timothy D. Ludwig (2007), "Reducing Warehouse Employee Errors Using Voice-Assisted Technology That Provided Immediate Feedback," *Journal of Organizational Behavior Management*, 27 (1), 1-31.

Bettencourt, Lance A. and Stephen W. Brown (2003), "Role Stressors and Customer-Oriented Boundary-Spanning Behaviors in Service Organizations," *Journal of the Academy of Marketing Science*, 31 (4), 394-408.

------, Stephen W. Brown, and Scott B. MacKenzie (2005), "Customer-Oriented Boundary-Spanning Behaviors: Test of a Social Exchange Model of Antecedents," *Journal of Retailing*, 81 (2), 141-57.

Biernat, Monica, and Kelly Danaher (2012), "Interpreting and Reacting to Feedback in Stereotype-Relevant Performance Domains," *Journal of Experimental Social Psychology*, 48 (1), 271-76.

------, Molly Carnes, Amarette Filut, and Anna Kaatz (2020), "Gender, Race, and Grant Reviews: Translating and Responding to Research Feedback," *Personality and Social Psychology Bulletin*, 46 (1), 140-54.

Bilodeau, Ina M. (1966), "Information Feedback," *Acquisition of Skill*, New York: Academic Press, 55, 255-96.

Bisbe, Josep, Joan-Manuel Batista-Foguet, and Robert Chenhall (2007), "Defining Management Accounting Constructs: A Methodological Note on the Risks of Conceptual Misspecification," *Accounting, Organizations and Society*, 32 (7-8), 789-820.

Blanca, Maria J., Rafael Alarcón, Jaume Arnau, Roser Bono, and Rebecca Bendayan (2017), "Non-Normal Data: Is ANOVA still a valid Option?," *Psicothema*, 29 (4), 552-57.

------, Rafael Alarcón, Jaume Arnau, Roser Bono, and Rebecca Bendayan (2018), "Effect of Variance Ratio on ANOVA Robustness: Might 1.5 be the Limit?," *Behavior Research Methods*, 50 (3), 937-62.

Blake, Robert R., and Jane S. Mouton (1985), "Don't Let the Norms Stifle Creativity," *Personnel*, 62, 28-33.

Bloomfield, Robert, Mark W. Nelson, and Eugene Soltes (2016), "Gathering Data for Archival, Field, Survey, and Experimental Accounting Research," *Journal of Accounting Research*, 54 (2), 341-95.

Bobe, Belete J., and Ralph Kober (2020), "Does gender matter? The Association between Gender and the Use of Management Control Systems and Performance Measures," *Accounting & Finance*, 60 (3), 2063-98.

Bollen, Kenneth A. (1989), *Structural Equations with Latent Variables*. New York, NY: Wiley.

------ (2011), "Evaluating Effect, Composite, and Causal Indicators in Structural Equation Models," *MIS Quarterly*, 35 (2), 359-72.

------, and Shawn Bauldry (2011), "Three Cs in Measurement Models: Causal Indicators, Composite Indicators, and Covariates," *Psychological Methods*, 16 (3), 265-84.

Boneau, C. Alan (1960), "The Effects of Violations of Assumptions Underlying the t Test," *Psychological Bulletin*, 57 (1), 49-64.

Borna, Shaheen, and Gwendolen White (2003), ""Sex" and "Gender": Two Confused and Confusing Concepts in the "Women in Corporate Management" Literature," *Journal of Business Ethics*, 47 (2), 89-99.

Bozdogan, Hamparsum (1987), "Model Selection and Akaike's Information Criterion (AIC): The General Theory and its Analytical Extensions," *Psychometrika*, 52 (3), 345-70.

------ (1994), “Mixture-Model Cluster Analysis Using Model Selection Criteria and a New Informational Measure of Complexity,” *Proceedings of the first US/Japan conference on the frontiers of statistical modeling: An informational approach*, 2, 69-113.

Brodie, Roderick J., Linda D. Hollebeek, Biljana Jurić, and Ana Ilić (2011), “Customer Engagement: Conceptual Domain, Fundamental Propositions, and Implications for Research,” *Journal of Service Research*, 14 (3), 252-71.

Brookins, Craig C., Thabiti M. Anyabwile, and Rupert Nacoste (1996), “Exploring the Links between Racial Identity Attitudes and Psychological Feelings of Closeness in African American College Students,” *Journal of Applied Social Psychology*, 26 (3), 243-64.

Buckingham, Marcus, and Ashley Goodall (2019), “The Feedback Fallacy,” *Harvard Business Review*, 97 (2), 92-101.

Burton, James P., and Jenny M. Hoobler (2006), “Subordinate Self-Esteem and Abusive Supervision,” *Journal of Managerial Issues*, 18, 340-55.

Butler, Judith (2011), *Gender Trouble: Feminism and the Subversion of Identity*. New York: Routledge.

Cabiddu, Francesca, Morena Frau, and Sebastiano Lombardo (2019), “Toxic Collaborations: Co-Destroying Value in the B2B Context,” *Journal of Service Research*, 22 (3), 241-55.

Campbell, Dennis, Marc J. Epstein, and F. Asis Martinez-Jerez (2011), “The Learning Effects of Monitoring,” *The Accounting Review*, 86 (6), 1909-34.

Campbell, Donald T., and Julian C. Stanley (1963), *Experimental and Quasi-Experimental Designs for Research*. Boston, MA: Houghton Mifflin Company.

Carli, Linda L. (1989), “Gender Differences in Interaction Style and Influence,” *Journal of Personality and Social Psychology*, 56 (4), 565-76.

Carli, Linda L. (2010), “Gender and Group Behavior,” in: *Handbook of Gender Research in Psychology*, J. Chrisler, and D. McCreary, eds.New York, NY: Springer, 337-58.

Carte, Traci A., and Craig J. Russell (2003), “In Pursuit of Moderation: Nine Common Errors and Their Solutions,” *MIS Quarterly*, 27 (3), 479-501.

Carver, Charles S. and Michael F. Scheier (1981): *Attention and Self-Regulation: A Control-Theory Approach to Human Behavior*. New York: Springer-Verlag.

Casas-Arce, Pablo, Sofia M. Lourenço, and Asís Martínez-Jerez (2017), “The Performance Effect of Feedback Frequency and Detail: Evidence from a Field Experiment in Customer Satisfaction,” *Journal of Accounting Research*, 55 (5), 1051-88.

Cassel, Claes, Peter Hackl, and Anders H. Westlund (1999), “Robustness of Partial Least-Squares Method for Estimating Latent Variable Quality Structures,” *Journal of Applied Statistics*, 26 (4), 435-46.

Cenfetelli, Ronald T., and Geneviève Bassellier (2009), “Interpretation of Formative Measurement in Information Systems Research,” *MIS Quarterly*, 33 (4), 689-707.

Cepeda, Gabriel C., Christian Nitzl, and José L. Roldán (2017), “Mediation Analyses in Partial Least Squares Structural Equation Modeling: Guidelines and Empirical Examples,” in *Partial Least Squares Path Modeling: Basic Concepts, Methodological Issues and Applications*, H. Latan and R. Noonan, eds. Cham, Switzerland: Springer International, 173-95.

Challagalla, Goutam N. and Tasadduq A. Shervani (1996), “Dimensions and Types of Supervisory Control: Effects on Salesperson Performance and Satisfaction,” *Journal of Marketing*, 60 (1), 89-105.

Chang, Woojung and Steven A. Taylor (2016), “The Effectiveness of Customer Participation in New Product Development: A Meta-Analysis”, *Journal of Marketing*, 80 (1), 47-64.

Cheah, Jun-Hwa, Marko Sarstedt, Christian M. Ringle, T. Ramayah, and Hiram Ting (2018), “Convergent Validity Assessment of Formatively Measured Constructs in PLS-SEM: On Using Single-Item Versus Multi-Item Measures in Redundancy Analyses,” *International Journal of Contemporary Hospitality Management*, 30 (11), 3192-210.

Chenhall, Robert 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), 127-68.

Chenhall, Robert H. (2012), “Developing an Organizational Perspective to Management Accounting,” *Journal of Management Accounting Research*, 24 (1), 65-76.

Chin, Wynne W. (1998), “The Partial Least Squares Approach to Structural Equation Modeling,” in *Modern Methods for Business Research*, G.A. Marcoulides, eds. Mahwah, NJ: Lawrence Erlbaum Associates, 295-336.

------ (2003), “A Permutation Procedure for Multi-Group Comparison of PLS Models,” *PLS and Related Methods: Proceedings of the International Symposium PLS*, 3 (1), 33-43.

------, and Jens Dibbern (2010), “An Introduction to a Permutation Based Procedure for Multi-Group PLS Analysis: Results of Tests of Differences on Simulated Data and a Cross Cultural Analysis of the Sourcing of Information System Services between Germany and the USA,” in *Handbook of Partial Least Squares*, V.V.

Esposito, W.W. Chin, J. Henseler, and H. Wang, eds. Berlin, Heidelberg: Springer, 171-93.

Churchill, Gilbert A. Jr. (1979), “A Paradigm for Developing Better Measures of Marketing Constructs,” *Journal of Marketing Research*, 16 (1), 64-73.

------, and Dawn Iacobucci (2005), *Marketing Research: Methodological Foundations*. New York: Dryden Press.

Coelho, Filipe and Mário Augusto (2010), “Job Characteristics and the Creativity of Frontline Service Employees,” *Journal of Service Research*, 13 (4), 426-38.

------, ------, and Luis Filipe Lages (2011), “Contextual Factors and the Creativity of Frontline Employees: The Mediating Effects of Role Stress and Intrinsic Motivation,” *Journal of Retailing*, 87 (1), 31-45.

Cohen, Jacob (1978), “Partialed Products are Interactions; Partialed Powers are Curve Components,” *Psychological Bulletin*, 85 (4), 858-66.

------, (1988), *Statistical Power Analysis for the Behavioral Sciences*. Academic Press.

Collis, Jill, and Roger Hussey (2013), *Business Research: A practical Guide for undergraduate and postgraduate Students*. Basingstoke: Palgrave Macmillan.

Cook, Thomas D., Donald T. Campbell, and A. Day (1979), *Quasi-Experimentation: Design & Analysis Issues for Field Settings* (Vol. 351). Boston, MA: Houghton Mifflin.

Cools, Martine, Kristof Stouthuysen, and Alexandra Van den Abbeele (2017), “Management Control for stimulating different Types of Creativity: The Role of Budgets,” *Journal of Management Accounting Research*, 29 (3), 1-21.

Costa, Paul T., and Robert R. McCrae (1992), *Revised NEO Personality Inventory (NEO PI-R) and NEO Five-Factor Inventory (NEO-FFI): Professional Manual*. Odessa, FL: Psychological Assessment Resources.

Cowan, Nelson (1998), “Visual and Auditory Working Memory Capacity,” *Trends in Cognitive Sciences*, 2 (3), 77-80.

------ (2001), “The Magical Number 4 in Short-Term Memory: A Reconsideration of Mental Storage Capacity,” *Behavioral and Brain Sciences*, 24 (1), 87-114.

Cronbach, Lee J. (1987), “Statistical Tests for Moderator Variables: Flaws in Analysis Recently Proposed,” *Psychological Bulletin*, 102 (3), 414-17.

Crowder, Robert G. (2014), *Principles of Learning and Memory*. New York: Psychology Press.

Dagger, Tracey S., Peter J. Danaher, and Brian J. Gibbs (2009), “How often versus how long: The Interplay of Contact Frequency and Relationship Duration in customer-reported Service Relationship Strength,” *Journal of Service Research*, 11 (4), 371-88.

Davila, Antonio, and Angelo Ditillo (2017), “Management Control Systems for creative Teams: Managing stylistic Creativity in Fashion Companies,” *Journal of Management Accounting Research*, 29 (3), 27-47.

------, George Foster, and Mu Li (2009), “Reasons for Management Control Systems Adoption: Insights from Product Development Systems choice by early-stage entrepreneurial companies." *Accounting, Organizations and Society*, 34 (3-4), 322-47.

------, George Foster, and Daniel Oyon (2009), “Accounting and Control, Entrepreneurship and Innovation: Venturing into new Research Opportunities,” *European Accounting Review*, 18 (2), 281-311.

Davison, Anthony C., and David Victor Hinkley (1997), *Bootstrap Methods and Their Application*. Cambridge: Cambridge University Press.

De Castro, Rochelle, Dana Sambuco, Peter A. Ubel, Abigail Stewart, and Reshma Jagsi (2013), “Batting 300 is Good: Perspectives of Faculty Researchers and their Mentors on Rejection, Resilience, and Persistence in academic medical Careers," *Academic Medicine: Journal of the Association of American Medical Colleges*, 88 (4), 497-504.

De Stobbeleir, Katleen E. M., Susan J. Ashford, and Dirk Buyens (2011), "Self-Regulation of Creativity at Work: The Role of feedback-seeking Behavior in creative Performance," *Academy of Management Journal*, 54 (4), 811-31.

Deci, Edward L. (1975), *Intrinsic motivation*. New York, NY: Plenum.

------, James P. Connell, and Richard M. Ryan (1989), "Self-Determination in a Work Organization," *Journal of Applied Psychology*, 74 (4), 580-90.

Deloitte (2017), "Rewriting the Rules for the digital Age. 2017 Deloitte global human capital trends" (available at https://www2.deloitte.com/content/dam/Deloitte/global/Documents/About-Deloitte/central-europe/ce-global-human-capital-trends.pdf [retrieved 09-21]).

Diamantopoulos, Adamantios (2006), "The Error Term in formative Measurement Models: Interpretation and Modeling Implications," *Journal of Modelling in Management*, 1 (1), 7-17.

------ (2011), "Incorporating Formative Measures into Covariance-Based Structural Equation Models," *MIS Quarterly*, 35, 335-58.

------, Marko Sarstedt, Christoph Fuchs, Petra Wilczynski, and Sebastian Kaiser (2012), "Guidelines for Choosing between Multi-Item and Single-Item Scales for Construct Measurement: A Predictive Validity Perspective," *Journal of the Academy of Marketing Science*, 40 (3), 434-49.

------, and Heidi M. Winklhofer (2001), "Index Construction with Formative Indicators: An Alternative to Scale Development," *Journal of Marketing Research*, 38 (2), 269-77.

Dijkstra, Theo K., and Jörg Henseler (2015), "Consistent Partial Least Squares Path Modeling," *MIS Quarterly*, 39 (2), 297-316.

Dipboye, Robert L. (1985), "Some Neglected Variables in Research on Discrimination in Appraisals," *Academy of Management Review*, 10 (1), 116-27.

Dodds, Rebecca A., Thomas B. Ward, and Steven M. Smith (2012), "Incubation in Problem Solving and Creativity," in: *The Creativity Research Handbook*, Vol. 3, M. A. Runco, eds. New York, NY: Hampton Press, 251-84.

Dohmen, Thomas, and Armin Falk (2011), "Performance Pay and Multidimensional Sorting: Productivity, Preferences, and Gender," *American Economic Review*, 101 (2), 556-90.

Dokko, Gina, Aimée A. Kane, and Marco Tortoriello (2013), "One of Us or One of My Friends: How Social Identity and Tie Strength Shape the Creative Generativity of Boundary-Spanning Ties," *Organization Studies*, 35 (5), 703-26.

Doney, Patricia M. and Joseph P. Cannon (1997), "An Examination of the Nature of Trust in Buyer-Seller Relationships," *Journal of Marketing*, 61 (2), 35-51.

Dong, Y., Hui Liao, Aichia Chuang, Jing Zhou, and Elizabeth M. Campbell (2015), "Fostering Employee Service Creativity: Joint Effects of Customer Empowering Behaviors and Supervisory Empowering Leadership," *Journal of Applied Psychology*, 100 (5), 1364-80.

Döring, Nicola, and Jürgen Bortz (2016), *Forschungsmethoden und Evaluation*. Wiesbaden: Springer.

Drolet, Aimee L., and Donald G. Morrison (2001), "Do We Really Need Multiple-Item Measures in Service Research?," *Journal of Service Research*, 3 (3), 196-204.

Dubinsky, Alan J., and Steven J. Skinner (1984), "Impact of Job Characteristics on Retail Salespeople's Reactions to their Jobs," *Journal of Retailing*, 60 (2), 35-62.

Eagly, Alice H., and Maureen Crowley (1986), "Gender and Helping Behavior: A Meta-analytic Review of the Social Psychological Literature," *Psychological Bulletin*, 100 (3), 283-308.

------, Mary C. Johannesen-Schmidt, and Marloes L. van Engen (2003), "Transformational, Transactional, and Laissez-Faire Leadership Styles: A Meta-Analysis Comparing Women and Men," *Psychological Bulletin*, 129 (4), 569-91.

------, and Steven J. Karau (1991), "Gender and the Emergence of Leaders: A Meta-Analysis," *Journal of Personality and Social Psychology*, 60 (5), 685-710.

------, and Wendy Wood (2012), "Social Role Theory," in: *Handbook of Theories in Social Psychology*, P. van Lange, A. Kruglanski, and E. T. Higgins, eds. Los Angeles: SAGE Publications Ltd, 458-76.

Easterby-Smith, Mark, Richard Thorpe, and Paul R. Jackson (2012), *Management Research*. London: Sage.

Edmondson, Amy C. (1999), "Psychological Safety and Learning Behavior in Work Teams," *Administrative Science Quarterly*, 44, 350-83.

Efron, Bradley, and Robert J. Tibshirani (1986), "Bootstrap Methods for Standard Errors, Confidence Intervals, and Other Measures of Statistical Accuracy," *Statistical Science*, 1 (1), 54-75.

------, and ------ (1994), *An Introduction to the Bootstrap*. Boca Raton, FL: Chapman & Hall.

Eisenberger, Robert, Linda Rhoades, and Judy Cameron (1999), "Does Pay for Performance Increase or Decrease Perceived Self-Determination and Intrinsic Motivation?," *Journal of Personality and Social Psychology*, 77 (5), 1026-40.

Engel, James F., Roger D. Blackwell, and Paul W. Miniard (1995), *Consumer Behavior*, 8th ed. Fort Worth: Dryden Press.

Ertac, Seda, and Balazs Szentes (2011), "The Effect of Information on Gender Differences in Competitiveness: Experimental Evidence", Working Paper, No. 1104, Koç University-TÜSİAD Economic Research Forum (ERF), Istanbul.

Eva, Nathan, Hannah Meacham, Alexander Newman, Gary Schwarz, and Tse Leng Tham (2019), "Is Coworker Feedback More Important than Supervisor Feedback for Increasing Innovative Behavior?," *Human Resource Management*, 58 (4), 383-96.

Eyring, Henry, and V. G. Narayanan (2018), "Performance Effects of Setting a High Reference Point for Peer-Performance Comparison," *Journal of Accounting Research*, 56 (2), 581-615.

Falk, R. Frank, and Nancy B. Miller (1992), *A Primer for Soft Modeling*. Akron: University of Akron Press.

Farmer, Steven M., Pamela Tierney, and Kate Kung-McIntyre (2003), "Employee Creativity in Taiwan: An Application of Role Identity Theory," *Academy of Management Journal*, 46 (5), 618-30.

Faure, Corinne (2004), "Beyond Brainstorming: Effects of Different Group Procedures on Selection of Ideas and Satisfaction with the Process," *Journal of Creative Behavior*, 38, 13-34.

Fedor, Donald B. and M. Ronald Buckley (1987), "Providing Feedback to Organizational Members: A Reconsideration," *Journal of Business and Psychology*, 2, 171-81.

------, Walter D. Davis, John M. Maslyn, and Kieran Mathieson (2001), "Performance Improvement Efforts in Response to Negative Feedback: The Roles of Source Power and Recipient Self-Esteem," *Journal of Management*, 27 (1), 79-97.

Fletcher, Clive (1999), “The Implications of Research on Gender Differences on self-assessments and 360-degree appraisal,” *Human Resource Management Journal*, 9 (1), 39-46.

Floyd, Eric, and John A. List (2016), “Using Field Experiments in Accounting and Finance,” *Journal of Accounting Research*, 54 (2), 437-75.

Fornell, Claes (1982), *A Second Generation of Multivariate Analysis: Measurement and Evaluation*. Vol. 2. Greenwood.

------, and Fred L. Bookstein (1982), “Two Structural Equation Models: LISREL and PLS Applied to Consumer Exit-Voice Theory,” *Journal of Marketing Research*, 19 (4), 440-52.

------, and David F. Larcker (1981), “Evaluating Structural Equation Models with Unobservable Variables and Measurement Error,” *Journal of Marketing Research*, 18 (1), 39-50.

------, and David F. Larcker (1987), “A Second Generation of Multivariate Analysis: Classification of Methods and Implications for Marketing Research Review of Marketing,” *Review of Marketing*, 51, 407-50.

Fotaki, Marianna, Beverly Dawn Metcalfe, and Nancy Harding (2014), “Writing Materiality into Management and Organization Studies through and with Luce Irigaray,” *Human Relations*, 67 (10), 1239-63.

Franke, George, and Marko Sarstedt (2019), “Heuristics Versus Statistics in Discriminant Validity Testing: A Comparison of Four Procedures,” *Internet Research*, 29 (3), 430-47.

Frey, Bruno S. and Reto Jegen (2001), “Motivation Crowding Theory,” *Journal of Economic Surveys*, 15 (5), 589-611.

Frieze, Irene Hanson, and Man Yu Li (2010), “Gender, Aggression, and Prosocial Behavior,” in: in: *Handbook of Gender Research in Psychology*, J. Chrisler, and D. McCreary, eds. Springer, New York, NY, 311-35.

Fuchs, Christoph, and Adamantios Diamantopoulos (2009), “Using Single-Item Measures for Construct Measurement in Management Research: Conceptual Issues and Application Guidelines,” *Die Betriebswirtschaft*, 69 (2), 195-210.

Fugate, Douglas L., and Joanna Phillips (2010), “Product Gender Perceptions and Antecedents of Product Gender Congruence,” *Journal of Consumer Marketing*, 27 (3), 251-61.

Gagné, Marylène and Edward L. Deci (2005), “Self-determination Theory and Work Motivation,” *Journal of Organizational Behavior*, 26 (4), 331-62.

Gardner, Donald G., Larry L. Cummings, Randall B. Dunham and Jon L. Pierce (1998), “Single-Item Versus Multiple-Item Measurement Sales: An Empirical Comparison,” *Educational and Psychological Measurement*, 58 (6), 898-915.

Gardner, Wendi L., and Shira Gabriel (2004), “Gender Differences in Relational and Collective Interdependence: Implications for Self-Views, Social Behavior, and Subjective Well-Being,” in: *The Psychology of Gender*, A. H. Eagly, A. E. Beall, and R. J. Sternberg, eds. New York, NY: The Guilford Press, 169-91.

Gefen, David, Edward E. Rigdon, and Detmar Straub (2011), “Editor's Comment: An Update and Extension to SEM Guidelines for Administrative and Social Science Research,” *MIS Quarterly*, 35 (2), iii-xiv.

Geisser, Seymour (1974), “A Predictive Approach to the Random Effect Model,” *Biometrika*, 61 (1), 101-7.

Giebelhausen, Michael, Stacey G. Robinson, Nancy J. Sirianni, and Michael K. Brady (2014), “Touch versus Tech: When Technology Functions as a Barrier or a Benefit to Service Encounters,” *Journal of Marketing*, 78 (4), 113-24.

Gilhooly, Ken J. (2016), “Incubation and Intuition in Creative Problem Solving,” *Frontiers in Psychology*, 7 (July), Article 1076.

Gill, David, and Victoria Prowse (2014), “Gender differences and dynamics in competition: The role of luck,” *Quantitative Economics,* 5(2), 351-76.

Gippel, Jennifer, Tom Smith, and Yushu Zhu (2015), “Endogeneity in Accounting and Finance Research: Natural Experiments as a State-of-the-Art Solution,” *Abacus*, 51 (2), 143-68.

Glass, Gene V., Percy D. Peckham, and James R. Sanders (1972), “Consequences of Failure to Meet Assumptions Underlying the Fixed Effects Analyses of Variance and Covariance,” *Review of Educational Research*, 42 (3), 237-88.

Gneezy, Uri, Muriel Niederle, and Aldo Rustichini (2003), “Performance in Competitive Environments: Gender Differences,” *The Quarterly Journal of Economics*, 118 (3), 1049-74.

------, and Aldo Rustichini (2004), “Gender and Competition at a young Age,” *American Economic Review*, 94 (2), 377-81.

Goodhue, Dale L., William Lewis, and Ron Thompson (2012), “Does PLS Have Advantages for Small Sample Size or Non-Normal Data?,” *MIS Quarterly*, 36 (3), 981-1001.

Goodman, Jodi S., Robert E. Wood, and Margaretha Hendrickx (2004), “Feedback Specificity, Exploration, and Learning,” *Journal of Applied Psychology*, 89 (2), 248-62.

Goomas, David T., and Timothy David Ludwig (2017), “Computerized immediate Feedback increases Product Recall Efficiency due to interlocking Contingencies in Food Manufacturing,” *Journal of Organizational Behavior Management*, 37 (1), 96-106.

------, Stuart M. Smith, and Timothy D. Ludwig (2011), “Business Activity Monitoring: Real-time Group Goals and Feedback Using an Overhead Scoreboard in a Distribution Center,” *Journal of Organizational Behavior Management*, 31 (3), 196-09.

Gouldner, Alvin W. (1960), “The Norm of Reciprocity: A Preliminary Statement,” *American Sociological Review*, 161-78.

Gow, Ian D., David F. Larcker, and Peter C. Reiss (2016), “Causal Inference in Accounting Research,” *Journal of Accounting Research*, 54 (2), 477-523.

Grabner, Isabella (2014), “Incentive System Design in Creativity-Dependent Firms,” *The Accounting Review*, 89 (5), 1729-50.

------, and Frank Moers (2013), “Management Control as a System or a Package? Conceptual and empirical Issues,” *Accounting, Organizations and Society*, 38 (6-7), 407-19.

------, and Gerhard Speckbacher (2016), “The Cost of Creativity: A Control Perspective,” *Accounting, Organizations and Society*, 48, 31-42.

Graham, Katarina A., Scott B. Dust, and Jonathan C. Ziegert (2018), “Supervisor-Employee Power Distance Incompatibility, Gender Similarity, and Relationship Conflict: A Test of Interpersonal Interaction Theory,” *Journal of Applied Psychology*, 103 (3), 334-46.

Greller, Martin M., and David M. Herold (1975), “Sources of Feedback: A Preliminary Investigation,” *Organizational Behavior & Human Performance*, 13 (2), 244-56.

Grewal, Rajdeep, Murali Chandrashekaran, Jean L. Johnson, and Girish Mallapragada (2013), “Environments, Unobserved Heterogeneity, and the Effect of Market Orientation on Outcomes for High-Tech Firms,” *Journal of the Academy of Marketing Science*, 41 (2), 206-33.

Gudergan, Siegfried P., Christian M. Ringle, Sven Wende, and Alexander Will (2008), “Confirmatory Tetrad Analysis in PLS Path Modeling,” J*ournal of Business Research*, 61 (12), 1238-49.

Guilford, Joy Paul (1950), “Creativity,” *American Psychologist*, 5, 444-54.

Hackman, J. Richard and Edward E. Lawler (1971), “Employee Reactions to Job Characteristics,” *Journal of Applied Psychology*, 55 (3), 259-86.

------, and Greg R. Oldham (1975), “Development of the Job Diagnostic Survey,” *Journal of Applied Psychology*, 60 (2), 159-70.

Hage, Jerald and Michael Aiken (1967), “Relationship of Centralization to other structural Properties,” *Administrative Science Quarterly*, 72-92.

Hahn, Carsten, Michael D. Johnson, Andreas Herrmann, and Frank Huber (2002), “Capturing Customer Heterogeneity Using a Finite Mixture PLS Approach,” *Schmalenbach Business Review*, 54 (3), 243-69.

Hair, Joseph F., Jr., William C. Black, Barry J. Babin, Rolph E. Anderson, and Ronald L. Tatham (2010), *Multivariate Data Analysis: A Global Perspective (7th ed.)*. Upper Saddle River, New Jersey: Pearson Prentice Hall Publishing.

------, G. Tomas M. Hult, Christian M. Ringle, and Marko Sarstedt (2017), *A Primer on Partial Least Squares Structural Equation Modeling (PLS-SEM) (2nd ed.)*. Thousand Oaks, CA: Sage Publications.

------, Lucy M. Matthews, Ryan L. Matthews, and Marko Sarstedt (2017), “PLS-SEM or CB-SEM: Updated Guidelines on Which Method to Use,” *International Journal of Multivariate Data Analysis*, 1 (2), 107-23.

------, Christian M. Ringle, and Marko Sarstedt (2011), “PLS-SEM: Indeed a Silver Bullet,” *Journal of Marketing Theory and Practice*, 19 (2), 139-52.

------, Christian M. Ringle, and Marko Sarstedt (2013), "Partial Least Squares Structural Equation Modeling: Rigorous Applications, Better Results and Higher Acceptance," *Long Range Planning*, 46 (1-2), 1-12.

------, Jeff J. Risher, Marko Sarstedt, and Christian M. Ringle (2019), "When to Use and How to Report the Results of PLS-SEM," *European Business Review*, 31 (1), 2-24.

------, Marko Sarstedt, Lucy M. Matthews, and Christian M. Ringle (2016), "Identifying and Treating Unobserved Heterogeneity with FIMIX-PLS: Part I – Method," *European Business Review*, 28 (1), 63-76.

------, Marko Sarstedt, Christian M. Ringle, Siegfried P. Gudergan (2017), *Advanced Issues in Partial Least Squares Structural Equation Modeling*. Thousand Oaks, CA: Sage Publications.

------, Marko Sarstedt, Christian M. Ringle, and Jeannette A. Mena (2012), "An Assessment of the Use of Partial Least Squares Structural Equation Modeling in Marketing Research," *Journal of the Academy of Marketing Science*, 40 (3), 414-33.

Hancock, Adrienne B., and Benjamin A. Rubin (2015), "Influence of Communication Partner's Gender on Language," *Journal of Language and Social Psychology*, 34 (1), 46-64.

Hannan, R. Lynn, Gregory P. McPhee, Andrew H. Newman, and Ivo D. Tafkov (2013), "The Effect of Relative Performance Information on Performance and Effort Allocation in a Multi-task Environment," *The Accounting Review*, 88 (2), 553-75.

Harackiewicz, Judith M. (1979), "The Effects of Reward Contingency and Performance Feedback on intrinsic Motivation," *Journal of Personality and Social Psychology*, 37 (8), 1352-63.

Hardacre, Stephanie L., and Emina Subašić (2018), "Whose Issue is it anyway? The Effects of Leader Gender and Equality Message Framing on Men's and Women's Mobilization toward Workplace Gender Equality," *Frontiers in Psychology*, 9, No. 2497.

Harrison, Glenn W., and John A. List (2004), "Field Experiments," *Journal of Economic Literature*, 42 (4), 1009-55.

Hartline, Michael D., and O. C. Ferrell (1996), "The Management of Customer-Contact Service Employees: An Empirical Investigation," *Journal of Marketing*, 60 (4), 52-70.

Hartmann, Frank G. H., and Frank Moers (1999), "Testing Contingency Hypotheses in Budgetary Research: An Evaluation of the Use of Moderated Regression Analysis," *Accounting, Organizations and Society*, 24, 291-316.

Harwell, Michael R., Elaine N. Rubinstein, William S. Hayes, and Corley C. Olds (1992), "Summarizing Monte Carlo Results in Methodological Research: The One- and Two-Factor Fixed Effects ANOVA Cases," *Journal of Educational Statistics*, 17 (4), 315-39.

Hayes, Andrew F. (2017), *Introduction to Mediation, Moderation, and Conditional Process Analysis: A Regression-Based Approach*. New York, NY: The Guilford Press.

Hecht, Gary, Andrew H. Newman, and Ivo D. Tafkov (2019), "Managers' Strategic Use of Discretion Over Relative Performance Information Provision and Implications for Team-Members' Effort," *Management Accounting Research*, 45, 1-18.

Heilman, Madeline E. (1983), "Sex Bias in Work Settings: The Lack of Fit Model," *Research in Organizational Behavior*, 5, 269-98.

------, and Julie J. Chen (2005), “Same Behavior, Different Consequences: Reactions to Men's and Women's Altruistic Citizenship Behavior,” *Journal of Applied Psychology*, 90 (3), 431-41.

Heinz, Matthias, Steffen Juranek, and Holger A. Rau (2012), „Do Women behave more reciprocally than Men? Gender Differences in real effort Dictator Games,” *Journal of Economic Behavior & Organization*, 83 (1), 105-10.

Helgeson, James G., Kevin E. Voss, and Willbann D. Terpening (2002), “Determinants of Mail-Survey Response: Survey Design Factors and Respondent Factors,” *Psychology & Marketing*, 19 (3), 303-28.

Henri, Jean-François (2006), “Management Control Systems and Strategy: A resource-based Perspective,” *Accounting, Organizations and Society*, 31 (6), 529-58.

Henseler, Jörg (2010), “On the Convergence of the Partial Least Squares Path Modeling Algorithm,” *Computational Statistics*, 25 (1), 107-20.

------ (2012), “PLS-MGA: A Non-Parametric Approach to Partial Least Squares-Based Multi-Group Analysis,” in *Challenges at the Interface of Data Analysis, Computer Science, and Optimization*, W.A. Gaul, A. Geyer-Schulz, L. Schmidt-Thieme, J. Kunze, eds. Berlin: Springer, 495-501.

------, and Wynne W. Chin (2010), “A Comparison of Approaches for the Analysis of Interaction Effects between Latent Variables Using Partial Least Squares Path Modeling,” *Structural Equation Modeling*, 17 (1), 82-109.

------, Theo K. Dijkstra, Marko Sarstedt, Christian M. Ringle, Adamntios Diamantopoulos, Detmar W. Straub, David J. Ketchen Jr., Joseph F. Hair Jr., G. Tomas M. Hult, and Roger J. Calantone (2014), “Common Beliefs and Reality About PLS: Comments on Rönkkö and Evermann (2013),” *Organizational Research Methods*, 17 (2), 182-209.

------, and Georg Fassott (2010), “Testing Moderating Effects in PLS Path Models: An Illustration of Available Procedures,” in *Handbook of Partial Least Squares*, V.V. Esposito, W.W. Chin, J. Henseler, and H. Wang, eds. Berlin, Heidelberg: Springer, 713-35.

------, Geoffrey Hubona, Pauline Ash Ray (2016), “Using PLS Path Modeling in New Technology Research: Updated Guidelines,” *Industrial Management & Data Systems*, 116 (1), 2-20.

------, Christian M. Ringle, and Marko Sarstedt (2015), “A New Criterion for Assessing Discriminant Validity in Variance-Based Structural Equation Modeling,” *Journal of the Academy of Marketing Science*, 43, 115-35.

------, Christian M. Ringle, Rudolf R. Sinkovics (2009), “The Use of Partial Least Squares Path Modeling in International Marketing,” in *Advances in International Marketing*, Vol. 20, R.R. Sinkovics, P.N. Ghauri, eds. Bingley: Emerald Group Publishing Limited, 277-320.

------, and Marko Sarstedt (2013), “Goodness-Of-Fit Indices for Partial Least Squares Path Modeling,” *Computational Statistics*, 28 (2), 565-80.

Heskett, James L., Thomas O. Jones, Gary W. Loveman, W. Earl Sasser, and Leonard A. Schlesinger (2008), “Putting the service-profit chain to work,” *Harvard Business Review*, 86 (7-8), 118-29.

Hildebrand, Christian, Gerald Häubl, Andreas Herrmann, and Jan R. Landwehr (2013), “When Social Media can be Bad for You: Community Feedback Stifles Consumer Creativity and Reduces Satisfaction with Self-Designed Products,” *Information Systems Research*, 24 (1), 14-29.

Hoever, Inga J., Jing Zhou, and Daan van Knippenberg (2018), Different Strokes for Different Teams: The Contingent Effects of Positive and Negative Feedback on the Creativity of Informationally Homogeneous and Diverse Teams," *Academy of Management Journal*, 61 (6), 2159-81.

Hogreve, Jens, Anja Iseke, Klaus Derfuss, and Tönnjes Eller (2017), "The Service–Profit Chain: A Meta-Analytic Test of a Comprehensive Theoretical Framework," *Journal of Marketing*, 81 (3), 41-61.

------, Nicola Bilstein, and Kathrin Hoerner (2019), "Service Recovery on Stage: Effects of Social Media Recovery on Virtually Present Others," *Journal of Service Research*, 22 (4), 421-39.

Holderness, D. Kip, Jr., Kari Joseph Olsen, and Todd A. Thornock (2017), "Who Are You to Tell Me That?! The Moderating Effect of Performance Feedback Source and Psychological Entitlement on Individual Performance," *Journal of Management Accounting Research*, 29 (2), 33-46.

------, Kari Joseph Olsen, and Todd A. Thornock (2020), "Assigned versus Chosen Relative Performance Information: The Effect of Feedback Frequency on Performance," *Journal of Management Accounting Research*, 32 (1), 137-58.

Homburg, Christian, and Hans Baumgartner (1995), "Beurteilung von Kausalmodellen – Bestandsaufnahme und Anwendungsempfehlungen," *Marketing: Zeitschrift für Forschung und Praxis*, 17 (3), 162-76.

Hon, Alice H., Wilco W. Chan, and Lin Lu (2013), "Overcoming Work-Related Stress and Promoting Employee Creativity in Hotel Industry: The Role of Task Feedback from Supervisor," *International Journal of Hospitality Management*, 33, 416-24.

Houde, Sébastien, Annika Todd, Anant Sudarshan, June A. Flora, K. Carrie Armel (2013), "Real-Time Feedback and Electricity Consumption: A Field Experiment Assessing the Potential for Savings and Persistence," *The Energy Journal*, 34 (1), 87-102.

Hsieh, Jung-Kuei, Hung-Chang Chiu, Chih-Ping Wei, Hsiuju Rebecca Yen, and Yu-Chun Cheng (2013), “A Practical Perspective on the Classification of Service Innovations,” *Journal of Services Marketing*, 27 (5), 371-84.

Hughes, David J., Allan Lee, Amy Wei Tian, Alexander Newman, and Alison Legood (2018), “Leadership, Creativity, and Innovation: A critical Review and practical Recommendations,” *The Leadership Quarterly*, 29 (5), 549-69.

Hulland, John (1999), “Use of Partial Least Squares (PLS) in Strategic Management Research: A Review of Four Recent Studies,” *Strategic Management Journal*, 20 (2), 195-204.

Iacobucci, Dawn, Neela Saldanha, and Xiaoyan Deng (2007), “A Mediation on Mediation: Evidence That Structural Equation Models Perform Better Than Regression,” *Journal of Consumer Psychology*, 17 (2), 140-54.

Ilgen, Daniel R., Cynthia D. Fisher, and M. Susan Taylor (1979), “Consequences of Individual Feedback on Behavior in Organizations,” *Journal of Applied Psychology*, 64 (4), 349-71.

Irwin, Julie R., and Gary H. McClelland (2001), “Misleading Heuristics and Moderated Multiple Regression Models,” *Journal of Marketing Research*, 38 (1), 100-9.

Jacoby, Jacob (2002), “Stimulus-Organism-Response Reconsidered: An Evolutionary Step in Modeling (Consumer) Behavior,” *Journal of Consumer Psychology*, 12 (1), 51-57.

Janka, Marc and Thomas W. Guenther (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), 131-61.

Jarvis, Cheryl Burke, Scott B. MacKenzie, and Philip M. Podsakoff (2003), “A Critical Review of Construct Indicators and Measurement Model Misspecification in

Marketing and Consumer Research," *Journal of Consumer Research*, 30 (2), 199-218.

Jedidi, Kamel, Harsharanjeet S. Jagpal, and Wayne S. DeSarbo (1997), "STEMM: A General Finite Mixture Structural Equation Model," *Journal of Classification*, 14 (1), 23-50.

Jhun, Soomyung, Zong-Tae Bae, and Seung-Yoon Rhee (2012), "Performance Change of Managers in Two Different Uses of Upward Feedback: A Longitudinal Study in Korea," *The International Journal of Human Resource Management*, 23 (20), 4246-64.

Johlke, Mark C. and Dale F. Duhan (2000), "Supervisor Communication Practices and Service Employee Job Outcomes," *Journal of Service Research*, 3 (2), 154-65.

Johnson, Cathryn, Jody Clay-Warner, and Stephanie J. Funk (1996), "Effects of Authority Structures and Gender on Interaction in Same Sex Groups," *Social Psychology Quarterly*, 59 (3), 221-36.

Jollineau, S. Jane, and Robert M. Bowen (2021), "A Practical Guide to using Mediation and Moderation Analyses in Accounting Research," Available at SSRN: https://ssrn.com/abstract=3786380.

Jones, Bethany A., Walter Pierre Bouman, Emma Haycraft, and Jon Arcelus (2019), "The Gender Congruence and Life Satisfaction Scale (GCLS): Development and Validation of a Scale to Measure Outcomes from Transgender Health Services," *International Journal of Transgenderism*, 20 (1), 63-80.

Jöreskog, Karl G. (1971), "Simultaneous Factor Analysis in Several Populations," *Psychometrika*, 36 (4), 409-26.

------, and Arthur S. Goldberger (1975), "Estimation of a Model with Multiple Indicators and Multiple Causes of a Single Latent Variable," *Journal of the American Statistical Association*, 70 (351), 631-39.

Jost, John T., and Aaron C. Kay (2005), “Exposure to Benevolent Sexism and Complementary Gender Stereotypes: Consequences for Specific and Diffuse Forms of System Justification,” *Journal of Personality and Social Psychology*, 88 (3), 498-509.

Judge, Timothy A., Carl J. Thoresen, Joyce E. Bono, and Gregory K. Patton (2001), “The Job Satisfaction-Job Performance Relationship: A Qualitative and Quantitative Review,” *Psychological Bulletin*, 127 (3), 376-407.

Kaas, Leo, and Christian Manger (2012), “Ethnic Discrimination in Germany’s Labour Market: A Field Experiment,” *German Economic Review*, 13 (1), 1-20.

Kachelmeier, Steven J., Bernhard E. Reichert, and Michael G. Williamson (2008), "Measuring and Motivating Quantity, Creativity, or Both,” *Journal of Accounting Research*, 46 (2), 341-73.

------, Laura W. Wang, and Michael G. Williamson (2019), “Incentivizing the Creative Process: From Initial Quantity to Eventual Creativity,” *The Accounting Review*, 94 (2), 249-66.

Kamas, Linda, and Anne Preston (2012), “The Importance of being confident; gender, career choice, and willingness to compete,” *Journal of Economic Behavior & Organization*, 83 (1), 82-97.

Katz, Daniel, and Robert L. Kahn (1978), *The social psychology of organizations*. New York: Wiley.

Kenny, David A. (2013), “Moderator Variables.,” (accessed March 24, 2021), [available at http://davidakenny.net/cm/moderation.html].

Kerlinger, Fred N., and Howard B. Lee (2000), *Foundations of Behavioral Research*. Fort Worth, TX: Harcourt College Publishers.

Khan, Gohar F., Marko Sarstedt, Wen-Lung Shiau, Joseph F. Hair Jr., Christian M. Ringle, and Martin Fritze (2019), “Methodological Research on Partial Least

Squares Structural Equation Modeling (PLS-SEM),” *Internet Research*, 29 (3), 407-29.

Kim, Yeun Joon, and Junha Kim (2020), “Does Negative Feedback Benefit (or Harm) Recipient Creativity? The Role of the Direction of Feedback Flow,” *Academy of Management Journal*, 63 (2), 584-612.

Kimberly, John R. (1981), “Managerial Innovation,” in: Handbook of Organizational Design, P. C. Nystrom and W. H. Starbuck, eds. Oxford, U.K.: Oxford University Press, 84-104.

Kinicki, Angelo J., Gregory E. Prussia, Bin J. Wu, and Frances M. McKee-Ryan (2004), “A Covariance Structure Analysis of Employees' Response to Performance Feedback,” *Journal of Applied Psychology*, 89 (6), 1057-69.

Kirby, Richard (2009), “Development of a Real-Time Performance Measurement and Feedback System for Alpine Skiers,” *Sports Technology*, 2 (1-2), 43-52.

Kirkman, Bradley L. and Benson Rosen (1999), “Beyond Self-management: Antecedents and Consequences of Team Empowerment,” *Academy of Management Journal*, 42 (1), 58-74.

Klein, Aleksandra and Gerhard Speckbacher (2020), “Does using Accounting Data in Performance Evaluations spoil Team Creativity? The Role of Leadership Behavior,” *The Accounting Review*, 95 (4), 313-30.

Klein, Christian, Bernhard Zwergel, and Sebastian Heiden (2009), “On the Existence of Sports Sentiment: The Relation between Football Match Results and Stock Index Returns in Europe,” *Review of Managerial Science*, 3 (3), 191-208.

Kluger, Avraham N., and Angelo DeNisi (1996), “The Effects of Feedback Interventions on Performance: A historical Review, a Meta-Analysis, and a preliminary Feedback Intervention Theory,” *Psychological Bulletin*, 119 (2), 254-84.

Kock, Ned (2015), “Common Method Bias in PLS-SEM: A Full Collinearity Assessment Approach,” *International Journal of e-Collaboration*, 11 (4), 1-10.

Krancher, Oliver, Pascal Luther, and Marc Jost (2018), “Key Affordances of Platform-as-a-Service: Self-Organization and Continuous Feedback,” *Journal of Management Information Systems*, 35 (3), 776-812.

Kruskal, William H., and W. Allen Wallis (1952), “Use of Ranks in One-Criterion Variance Analysis,” *Journal of the American statistical Association*, 47 (260), 583-621.

Kuegler, Maurice, Stefan Smolnik, and Gerald Kane (2015), “What’s in IT for Employees? Understanding the Relationship between Use and Performance in Enterprise Social Software”, *The Journal of Strategic Information Systems*, 24 (2), 90-112.

Kuhnen, Camelia M., and Agnieszka Tymula (2012), “Feedback, Self-esteem, and Performance in Organizations,” *Management Science*, 58 (1), 94-113.

Kumar, Piyush, Mayukh Dass, and Omer Topaloglu (2014), “Understanding the Drivers of Job Satisfaction of Frontline Service Employees: Learning From “Lost Employees”,” *Journal of Service Research*, 17 (4), 367-80.

Kuvaas, Bård (2011), “The interactive Role of Performance Appraisal Reactions and regular Feedback,” *Journal of Managerial Psychology*, 26 (2), 123-37.

Lages, Cristiana R., and Nigel F. Piercy (2012), “Key Drivers of Frontline Employee Generation of Ideas for Customer Service Improvement,” *Journal of Service Research*, 15 (2), 215-30.

Laher, Sumaya, and Sarah Croxford (2013), “Men are from Mars, Women are from Venus: Exploring Gender Differences in Personality in the South African Context,” *SA Journal of Human Resource Management*, 11 (1), 1-8.

Lai, Karen M., Bin Srinidhi, Ferdinand A. Gul, and Judy S. Tsui (2017), “Board Gender Diversity, Auditor Fees, and Auditor Choice,” *Contemporary Accounting Research*, 34 (3), 1681-714.

Lam, Chak Fu, D. Scott DeRue, Elizabeth P. Karam, and John R. Hollenbeck (2011), “The impact of feedback frequency on learning and task performance: Challenging the “more is better” assumption”, *Organizational Behavior and Human Decision Processes*, 116 (2), 217-228.

Lanaj, Klodiana, and John R. Hollenbeck (2015), “Leadership Over-Emergence in Self-Managing Teams: The Role of Gender and Countervailing Biases,” *Academy of Management Journal*, 58 (5), 1476-94.

Larson, James R., Jr. (1986), “Supervisors' Performance Feedback to Subordinates: The Impact of Subordinate Performance Valence and Outcome Dependence,” *Organizational Behavior and Human Decision Processes*, 37 (3), 391-408.

Latan, Hengky, Christian M. Ringle, and Charbel Jose Chiappetta Jabbour, (2018), “Whistleblowing Intentions among public Accountants in Indonesia: Testing for the moderation Effects,” *Journal of Business Ethics*, 152 (2), 573-88.

Leaper, Campbell, and Melanie M. Ayres (2007), “A Meta-Analytic Review of Gender Variations in Adults' Language Use: Talkativeness, Affiliative Speech, and Assertive Speech,” *Personality and Social Psychology Review*, 11 (4), 328-63.

Lechermeier, Jonas, and Martin Fassnacht (2018), “How do Performance Feedback Characteristics influence Recipients’ Reactions? A state-of-the-art Review on Feedback Source, Timing, and Valence Effects,” *Management Review Quarterly*, 68 (2), 145-93.

Lee, Sun Young, Marko Pitesa, Stefan Thau, and Madan M. Pillutla (2015), “Discrimination in Selection Decisions: Integrating Stereotype Fit and Interdependence Theories,” *Academy of Management Journal*, 58 (3), 789-812.

Levene, Howard (1960), "Robust Tests for Equality of Variances. Contributions to Probability and Statistics," in: *Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling*, I. Olkin, eds. Palo Alto, CA: Stanford University Press, 278-92.

Levitt, Steven D., and John A. List (2009), "Field Experiments in Economics: The Past, the Present, and the Future," *European Economic Review*, 53 (1), 1-18.

Li, Shelley Xin, and Tatiana Sandino (2018), "Effects of an Information Sharing System on Employee Creativity, Engagement, and Performance," *Journal of Accounting Research*, 56 (2), 713-47.

Limaj, Everist and Edward W. Bernroider (2019), "The Roles of absorptive Capacity and cultural Balance for exploratory and exploitative Innovation in SMEs," *Journal of Business Research*, 94, 137-53.

Lindquist, Everet Franklin (1953), *Design and Analysis of Experiments in Psychology and Education*. Boston, MA: Houghton Mifflin.

Lix, Lisa M., Joanne C. Keselman, and Harvey J. Keselman (1996), "Consequences of Assumption Violations Revisited: A Quantitative Review of Alternatives to the One-Way Analysis of Variance F test," *Review of Educational Research*, 66 (4), 579-619.

Lohmöller, Jan-Bernd (1989), "Predictive vs. Structural Modeling: Pls vs. ml," in *Latent Variable Path Modeling with Partial Least Squares*, Heidelberg: Physica, 199-226.

Loideain, Nóra Ni, and Rachel Adams (2020), "From Alexa to Siri and the GDPR: The Gendering of Virtual Personal Assistants and the Role of Data Protection Impact Assessments," *Computer Law & Security Review*, 36, 1-14.

Luck, Steven J. (1998), "Response from Luck and Vogel (response to commentary by Nelson Cowan)," *Trends in Cognitive Sciences*, 2, 78-80.

------, and Edward K. Vogel (1997), “The Capacity of Visual Working Memory for Features and Conjunctions,” *Nature*, 390 (6657), 279-281.

Luckett, Peter F., and Ian R. Eggleton (1991), “Feedback and Management Accounting: A Review of Research into Behavioural Consequences”, *Accounting, Organizations and Society*, 16 (4), 371-94.

Ludwig, Timothy D., and David T. Goomas (2009), “Real-Time Performance Monitoring, Goal-Setting, and Feedback for Fork-Lift Drivers in a Distribution Centre,” *Journal of Occupational and Organizational Psychology*, 82 (2), 391-403.

Luft, Joan, and Michael D. Shields (2003), “Mapping Management Accounting: Graphics and Guidelines for Theory-Consistent Empirical Research,” *Accounting, Organizations and Society*, 28 (2-3), 169-249.

Luke, Molli M., and Mark Alavosius (2011), “Adherence with Universal Precautions after Immediate, Personalized Performance Feedback,” *Journal of Applied Behavior Analysis*, 44 (4), 967-71.

Lurie, Nicholas H., and Jayashankar M. Swaminathan (2009), “Is timely Information always better? The Effect of Feedback Frequency on Decision Making,” *Organizational Behavior and Human Decision Processes*, 108 (2), 315-29.

Ma, Wei Ji, Masud Husain, and Paul M. Bays (2014), “Changing Concepts of Working Memory,” *Nature Neuroscience*, 17 (3), 347-356.

MacCallum, Robert C., and Michael W. Browne (1993), “The Use of Causal Indicators in Covariance Structure Models: Some Practical Issues,” *Psychological Bulletin*, 114 (3), 533-41.

MacKenzie, Scott B., Philip M. Podsakoff, and Nathan P. Podsakoff (2011), “Construct Measurement and Validation Procedures in MIS and Behavioral Research: Integrating New and Existing Techniques,” *MIS Quarterly*, 35 (2), 293-334.

MacKinnon, David P., Amanda J. Fairchild, and Matthew S. Fritz (2007), “Mediation Analysis,” *Annual Review of Psychology*, 58, 593-614.

Major, Brenda, and Pamela J. Sawyer (2009), “Attributions to Discrimination: Antecedents and Consequences,” in: *Handbook of Prejudice, Stereotyping, and Discrimination*, Todd D. Nelson, eds. New York: Psychology Press, 89-110.

Malhotra, Neeru and Anna-Lena Ackfeldt (2016), “Internal Communication and Prosocial Service Behaviors of Front-Line Employees: Investigating Mediating Mechanisms,” *Journal of Business Research*, 69 (10), 4132-39.

Malik, Muhammad Abdur Rahman, Jin Nam Choi, and Arif Nazir Butt (2019), “Distinct Effects of intrinsic Motivation and extrinsic Rewards on radical and incremental Creativity: The moderating Role of Goal Orientations,” *Journal of Organizational Behavior*, 40 (9-10), 1013-26.

Mancha, Ruben, Mark T. Leung, Jan Clark, and Minghe Sun (2014), “Finite Mixture Partial Least Squares for Segmentation and Behavioral Characterization of Auction Bidders,” *Decision Support Systems*, 57, 200-11.

Marcoulides, George A., Wynne W. Chin, and Carol Saunders (2012), “When Imprecise Statistical Statements Become Problematic: A Response to Goodhue, Lewis, and Thompson,” *MIS Quarterly*, 36 (3), 717-28.

------, and Carol Saunders (2006), “PLS: A Silver Bullet?,” *MIS Quarterly*, 30 (2), iii-iix.

Markey, Rob, Fred Reichheld, and Andreas Dullweber (2009), “Closing the Customer Feedback Loop,” *Harvard Business Review*, 87 (12), 43-47.

Martyn, Patricia, Breda Sweeney, and Emer Curtis (2016), “Strategy and Control: 25 Years of empirical Use of Simons’ Levers of Control Framework,” *Journal of Accounting & Organizational Change*, 12 (3), 281-324.

Matthews, Lucy M., Alex R. Zablah, Joe F. Hair, and Greg W. Marshall (2016), “Increased Engagement or Reduced Exhaustion: Which Accounts for the Effect of Job Resources on Salesperson Job Outcomes?," *Journal of Marketing Theory and Practice*, 24 (3), 249-64.

Mayer, David M., Mark G. Ehrhart, and Benjamin Schneider (2009), "Service Attribute Boundary Conditions of the Service Climate-Customer Satisfaction Link," *Academy of Management Journal*, 52 (5), 1034-50.

McClelland, Gary H., and Charles M. Judd (1993) "Statistical Difficulties of Detecting Interactions and Moderator Effects," *Psychological Bulletin*, 114 (2), 376-90.

McCrae, Robert R., and Oliver P. John (1992), "An Introduction to the Five-Factor Model and its Applications," *Journal of Personality*, 60, 175-215.

McCrae, Robert R., Antonio Terracciano, and 78 Members of the Personality Profiles of Cultures Project (2005), "Universal Features of Personality Traits From the Observer's Perspective: Data From 50 Cultures," *Journal of Personality and Social Psychology*, 88 (3), 547-61.

Mendes, Wendy B., Shannon McCoy, Brenda Major, and Jim Blascovich (2008), "How attributional Ambiguity shapes physiological and emotional Responses to social Rejection and Acceptance," *Journal of Personality and Social Psychology*, 94 (2), 278-91.

Merchant, Kenneth A. and Wim A. Van der Stede (2007), *Management Control Systems: Performance Measurement, Evaluation and Incentives*. London: Pearson Education.

Metcalfe, Janet, Nate Kornell, and Bridgid Finn (2009), "Delayed versus Immediate Feedback in Children's and Adults' Vocabulary Learning," *Memory & Cognition*, 37 (8), 1077-87.

Miller, George A (1956), "The Magic Number Seven Plus or Minus Two: Some Limits on our Capacity to Process Information," *Psychological Review*, 63 (2), 81-97.

Miller, James G. (1960), “Information Input Overload and Psychopathology,” *American Journal of Psychiatry*, 116 (8), 695-704.

Miller, Rowland S., Dan Perlman, and Sharon Brehm (2007), “Intimate Relationships,” in: Handbook of Intercultural Communication, H. Kotthoff, and H. Spencer-Oatey, eds. Berlin: Mouton de Gruyter, 341-62.

Miller, Brian K. and Marcia Simmering (2020), “Attitude Toward the Color Blue: An "" Ideal"" Ideal Marker Variable,” *Academy of Management Proceedings*, 2020 (1), p. 15040. Briarcliff Manor, NY: Academy of Management.

Moore, Kristin Anderson, Tamara G. Halle, Sharon Vandivere, and Carrie L. Mariner (2002), “Scaling Back Survey Scales: How Short Is Too Short?,” *Sociological Methods & Research*, 30 (4), 530-67.

Morrison, Elizabeth W. (1993), “Newcomer Information Seeking: Exploring Types, Modes, Sources, and Outcomes,” *Academy of Management Journal*, 36 (3), 557-89.

Motro, Daphna, and Aleksander P. Ellis (2017), “Boys, don’t Cry: Gender and Reactions to Negative Performance Feedback,” *Journal of Applied Psychology*, 102 (2), 227-35.

Mullins, Ryan, Raj Agnihotri, and Zachary Hall (2020), “The Ambidextrous Sales Force: Aligning Salesperson Polychronicity and Selling Contexts for Sales-Service Behaviors and Customer Value”, *Journal of Service Research*, 23 (1), 33-52.

Mundy, Julia (2010), “Creating dynamic Tensions through a balanced Use of Management Control Systems,” *Accounting, Organizations and Society*, 35, 499-523.

Nakata, Tatsuya (2015), “Effects of Feedback Timing on Second Language Vocabulary Learning: Does Delaying Feedback Increase Learning?,” *Language Teaching Research*, 19 (4), 416-34.

Nasr, Linda, Jamie Burton, and Thorsten Gruber (2018), "Developing a Deeper Understanding of Positive Customer Feedback," *Journal of Services Marketing*, 32 (2), 142-60.

Ng, Sylvia C., Carolin Plewa, and Jillian C. Sweeney (2016), "Professional service providers' resource integration styles (PRO-RIS) facilitating customer experiences", *Journal of Service Research*, 19 (4), 380-95.

Ng, Thomas W. H., and Daniel C. Feldman (2008), "The Relationship of Age to Ten Dimensions of Job Performance," *Journal of Applied Psychology*, 93 (2), 392-423.

------, and Daniel C. Feldman (2013), "A Meta-Analysis of the Relationships of Age and Tenure with Innovation-Related Behaviour," *Journal of Occupational and Organizational Psychology*, 86 (4), 585-616.

Niederle, Muriel, and Lise Vesterlund (2007), "Do Women shy away from Competition? Do Men compete too much?," *The Quarterly Journal of Economics*, 122 (3), 1067-101.

------, and Lise Vesterlund (2007), "Gender and Competition," *Annual Review of Economics*, 3 (1), 601-30.

Nitzl, Christian (2016), "The Use of Partial Least Squares Structural Equation Modelling (PLS-SEM) in Management Accounting Research: Directions for Future Theory Development," *Journal of Accounting Literature*, 37, 19-35.

------, and Bernhard Hirsch (2016), "The Drivers of a Superior's Trust Formation in His Subordinate: The Manager-Management Accountant Example," *Journal of Accounting & Organizational Change*, 12 (4), 472-503.

------, Jose L. Roldán, and Gabriel C. Cepeda (2016), "Mediation Analysis in Partial Least Squares Path Modeling: Helping Researchers Discuss More

Sophisticated Models," *Industrial Management & Data Systems*, 119 (9), 1849-64.

Nohria, Nitin, and Ranjay Gulati (1996), "Is Slack Good or Bad for Innovation?," *Academy of Management Journal*, 39 (5), 1245-64.

Nunnally, Jum C., and Ira Bernstein (1994), *Psychometric Theory (3rd ed.)*. New York: McGraw-Hill.

Oldham, Greg R. and Anne Cummings (1996), "Employee Creativity: Personal and Contextual Factors at Work," *The Academy of Management Journal*, 39 (3), 607-34.

Ostrom, Amy L., A. Parasuraman, David E. Bowen, Lia Patricio, and Christopher Voss (2015), "Service Research Priorities in a Rapidly Changing Context," *Journal of Service Research*, 18 (2), 127-59.

Ozkok, Ozlem, Simon J. Bell, Jagdip Singh, and Kwanghui Lim (2019), "Frontline Knowledge Networks in open Collaboration Models for Service Innovations," *AMS Review*, 9 (3), 268-88.

Papa, Frank J., David Aldrich, and Randall E. Schumacker (1999), "The Effects of Immediate Online Feedback upon Diagnostic Performance," *Academic Medicine*, 74 (10), 16-18.

Park, Lora E., Cheryl L. Kondrak, Deborah E. Ward, and Lindsey Streamer (2018), "Positive Feedback from Male Authority Figures boosts Women's Math Outcomes," *Personality and Social Psychology Bulletin*, *44*, 359-83.

Paustian-Underdahl, Samantha C., Lisa S. Walker, and David J. Woehr (2014), "Gender and Perceptions of Leadership Effectiveness: A Meta-Analysis of Contextual Moderators," *Journal of Applied Psychology*, 99 (6), 1129-45.

Pichler, Shaun, Gerard Beenen, and Stephen Wood (2020), "Feedback Frequency and Appraisal Reactions: A meta-analytic Test of Moderators," *The International Journal of Human Resource Management*, 31 (17), 2238-63.

Podsakoff, Philip M., Scott B. MacKenzie, Jeong-Yeon Lee, Nathan P. Podsakoff (2003), "Common Method Biases in Behavioral Research: A Critical Review of the Literature and Recommended Remedies," *Journal of Applied Psychology*, 88 (5),879-903.

------, and Dennis W. Organ (1986), "Self-reports in Organizational Research: Problems and Prospects," *Journal of Management*, 12 (4) 531-44.

Powell, Gary N., Jeffrey H. Greenhaus, Tammy D. Allen, and Russel E. Johnson (2019), "Introduction to Special Topic Forum: Advancing and Expanding Work-Life Theory from Multiple Perspectives," *Academy of Management Review*, 44 (1), 54-71.

Preacher, Kristopher J., and Andrew F. Hayes (2004), "SPSS and SAS Procedures for Estimating Indirect Effects in Simple Mediation Models," *Behavior Research Methods, Instruments, & Computers*, 36 (4), 717-31.

------, and Andrew F. Hayes (2008), "Asymptotic and Resampling Strategies for Assessing and Comparing Indirect Effects in Multiple Mediator Models," *Behavior Research Methods*, 40 (3), 879-91.

Presslee, Aadam, Thomas W. Vance, and R. Alan Webb (2013), "The Effects of Reward Type on Employee Goal Setting, Goal Commitment, and Performance," *The Accounting Review*, 88 (5), 1805-31.

PricewaterhouseCoopers (2015), "Performance Management. Change is on the way but will it be enough?" (available at https://www.pwc.com.au/people-business/assets/publications/performance-management-mar15.pdf [retrieved 09-21]).

Putman, Vicky L., and Paul B. Paulus (2009), “Brainstorming, Brainstorming Rules and Decision Making,” *The Journal of Creative Behavior*, 43 (1), 29-40.

Ramaswamy, Venkatram, Wayne S. DeSarbo, David J. Reibstein, and William T. Robinson (1993), “An Empirical Pooling Approach for Estimating Marketing Mix Elasticities with PIMS Data,” *Marketing Science*, 12 (1), 103-24.

Ratzmann, Martin, Siegfried P. Gudergan, and Ricarda Bouncken (2016), “Capturing Heterogeneity and PLS-SEM Prediction Ability: Alliance Governance and Innovation,” *Journal of Business Research*, 69 (10), 4593-603.

Reuvers, Mark, Marloes L. van Engen, Claartje J. Vinkenburg, and Elisabeth Wilson-Evered (2008), “Transformational Leadership and innovative Work Behaviour: Exploring the Relevance of Gender Differences,” *Creativity and Innovation Management*, 17 (3), 227-44.

Rietzschel, Eric F., Bernard A. Nijstad, and Wolfgang Stroebe (2006), Productivity is not Enough: A Comparison of Interactive and Nominal Brainstorming Groups on Idea Generation and Selection,” *Journal of Experimental Social Psychology*, 42, 244-51.

------, Bernard A. Nijstad, and Wolfgang Stroebe (2010), “The Selection of Creative Ideas after Individual Idea Generation: Choosing between Creativity and Impact,” *British Journal of Psychology*, 101 (1), 47-68.

Rigdon, Edward E. (2012), “Rethinking Partial Least Squares Path Modeling: In Praise of Simple Methods,” *Long Range Planning*, 45 (5-6), 341-58.

------ (2014a), “Comment on “Improper Use of Endogenous Formative Variables”,” *Journal of Business Research*, 67 (1), 2800-02.

------ (2014b), “Rethinking Partial Least Squares Path Modeling: Breaking Chains and Forging Ahead,” *Long Range Planning*, 47 (3), 161-67.

------ (2016), “Choosing PLS Path Modeling as Analytical Method in European Management Research: A Realist Perspective,” *European Management Journal*, 34 (6), 598-605.

------, Christian M. Ringle, and Marko Sarstedt (2010), “Structural Modeling of Heterogeneous Data with Partial Least Squares,” in *Review of Marketing Research*, Vol. 7, N.K. Malhotra, eds. Bingley: Emerald Group Publishing Limited, 255-96.

------, Christian M. Ringle, Marko Sarstedt, and Siegfried P. Gudergan (2011), “Assessing Heterogeneity in Customer Satisfaction Studies: Across Industry Similarities and Within Industry Differences,” in *Measurement and Research Methods in International Marketing*, M. Sarstedt, M. Schwaiger, and C.R. Taylor, eds. Bingley: Emerald Group Publishing Limited, 169-94.

Ringle, Christian M., Marko Sarstedt, and Erik A. Mooi (2010), “Response-Based Segmentation Using Finite Mixture Partial Least Squares,” in *Data Mining*, R. Stahlbock, S. Crone, S. Lessmann, eds. Boston, MA: Springer, 19-49.

------, Marko Sarstedt, Rainer Schlittgen, and Charles R. Taylor (2013), “PLS Path Modeling and Evolutionary Segmentation,” *Journal of Business Research*, 66 (9), 1318-24.

------, Marko Sarstedt, and Detmar W. Straub (2012), “Editor's Comments: A Critical Look at the Use of PLS-SEM in “MIS Quarterly”,” *MIS Quarterly*, 36 (1), iii-xiv.

------, Sven Wende, and Jan-Michael Becker (2015), “SmartPLS 3,” Boenningsteadt: SmartPLS GmbH, www.smartpls.com.

------, Sven Wende, and Alexander Will (2010), “Finite Mixture Partial Least Squares Analysis: Methodology and Numerical Examples,” in *Handbook of Partial Least Squares*, V.V. Esposito, W.W. Chin, J. Henseler, and H. Wang, eds. Berlin, Heidelberg: Springer, 195-218.

Rizzo, John R., Robert J. House, and Sidney I. Lirtzman (1970), “Role Conflict and Ambiguity in Complex Organizations,” *Administrative Science Quarterly*, 15 (2), 150-63.

Roberts, Tomi-Ann, and Susan Nolen-Hoeksema (1989), “Sex Differences in Reactions to evaluative Feedback,” *Sex Roles*, 21 (11-12), 725-47.

Rodgers, Waymond, and Andrés Guiral (2011), “Potential Model Misspecification Bias: Formative Indicators Enhancing Theory for Accounting Researchers,” *The International Journal of Accounting*, 46 (1), 25-50.

Röhner, Jessica, and Astrid Schütz (2020), *Psychologie der Kommunikation*. Berlin, Heidelberg: Springer.

Romero, Jaime, Daniel Ruiz-Equihua, Sandra M. C. Loureiro, and Luis V. Casaló (2021), “Smart Speaker Recommendations: Impact of Gender Congruence and Amount of Information on Users' Engagement and Choice,” *Frontiers in Psychology*, 12, 1037-46.

Rönkkö, Mikko, and Joerg Evermann (2013), “A Critical Examination of Common Beliefs About Partial Least Squares Path Modeling,” *Organizational Research Methods*, 16 (3), 425-48.

------, and Jukka Ylitalo (2011), “PLS Marker Variable Approach to Diagnosing and Controlling for Method Variance,” *ICIS 2011 Proceedings*, 8.

Rossiter, John R. (2002), “The C-OAR-SE Procedure for Scale Development in Marketing,” *International Journal of Research in Marketing*, 19 (4), 305-35.

Sagiv, Lilach, Sharon Arieli, Jacob Goldenberg, and Ayalla Goldschmidt (2010), “Structure and Freedom in Creativity: The Interplay between externally imposed Structure and personal cognitive Style,” *Journal of Organizational Behavior*, 31 (8), 1086-110.

Sarstedt, Marko, Jan-Michael Becker, Christian M. Ringle, and Manfred Schwaiger (2011), “Uncovering and Treating Unobserved Heterogeneity with FIMIX-PLS: Which Model Selection Criterion Provides an Appropriate Number of Segments?,” *Schmalenbach Business Review*, 63 (1), 34-62.

------, Paul Bengart, Abdel Monim Shaltoni, and Sebastian Lehmann (2018), “The Use of Sampling Methods in Advertising Research: A Gap between Theory and Practice,” *International Journal of Advertising*, 37 (4), 650-63.

------, Adamantios Diamantopoulos, and Thomas Salzberger (2016), “Should We Use Single Items? Better Not,” *Journal of Business Research*, 69 (8), 3199-203.

------, Adamantios Diamantopoulos, Thomas Salzberger, and Petra Baumgartner (2016), “Selecting Single Items to Measure Doubly Concrete Constructs: A Cautionary Tale,” *Journal of Business Research*, 69 (8), 3159-67.

------, Jörg Henseler, and Christian M. Ringle (2011), “Multigroup Analysis in Partial Least Squares (PLS) Path Modeling: Alternative Methods and Empirical Results,” in *Measurement and Research Methods in International Marketing*, M. Sarstedt, M. Schwaiger, and C.R. Taylor, eds. Bingley: Emerald Group Publishing Limited, 195-218.

------, and Erik Mooi (2014), *A Concise Guide to Market Research: The Process, Data, and Methods Using IBM SPSS Statistics*. Berlin: Springer.

------, and Christian M. Ringle (2010), “Treating Unobserved Heterogeneity in PLS Path Modeling: A Comparison of FIMIX-PLS With Different Data Analysis Strategies,” *Journal of Applied Statistics*, 37 (8), 1299-318.

------, Christian M. Ringle, and Joseph F. Hair Jr. (2017), “Partial Least Squares Structural Equation Modeling”, in *Handbook of Market Research*, C. Homburg, M. Klarmann, and A. Vomberg, eds. Heidelberg: Springer, 1-40.

------, Christian M. Ringle, Jörg Henseler, and Joseph F. Hair Jr. (2014), "On the Emancipation of PLS-SEM: A Commentary on Rigdon (2012)," *Long Range Planning*, 47 (3), 154-160.

------, and Petra Wilczynski (2009), "More for Less? A Comparison of Single-Item and Multi-Item Measures," *Die Betriebswirtschaft*, 69 (2), 211-27.

------, Petra Wilczynski, and T.C. Melewar, (2013), "Measuring Reputation in Global Markets—A Comparison of Reputation Measures' Convergent and Criterion Validities," *Journal of World Business*, 48 (3), 329-39.

Satterthwaite, Franklin E. (1946), "An Approximate Distribution of Estimates of Variance Components," *Biometrics Bulletin*, 2 (6), 110-14.

Schepers, Jeroen J.L., Edwin J. Nijssen, and Gielis A.H. van der Heijden (2016), "Innovation in the Frontline: Exploring the Relationship between Role Conflict, Ideas for Improvement, and Employee Service Performance," *International Journal of Research in Marketing*, 33 (4), 797-817.

Schneider, Benjamin and David E. Bowen (2019), "Perspectives on the Organizational Context of Frontlines: A Commentary," *Journal of Service Research*, 22 (1), 3-7.

Schnell, Rainer, Paul B. Hill, and Elke Esser (2018), *Methoden der Empirischen Sozialforschung* (11., überarb. Aufl.), Berlin: De Gruyter Oldenbourg.

Schwarz, Gideon (1978), "Estimating the Dimension of a Model," *Annals of Statistics*, 6 (2), 461-64.

Seiders, Kathleen, Andrea Godfrey Flynn, Leonard L. Berry, and Kelly L. Haws (2015), "Motivating Customers to adhere to Expert Advice in professional Services: A medical Service Context," *Journal of Service Research*, 18 (1), 39-58.

Shadish, William R., Thomas D. Cook, and Donald T. Campbell (2002), *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*, Boston, MA: Houghton Mifflin.

Shalley, Christina E., and Lucy L. Gilson (2004), “What Leaders need to Know: A Review of Social and Contextual Factors that can Foster or Hinder Creativity,” *The Leadership Quarterly*, 15 (1), 33-53.

------, Lucy L. Gilson, and Terry C. Blum (2009), “Interactive Effects of Growth need Strength, Work Context, and Job Complexity on self-reported creative Performance,” *Academy of Management Journal*, 52 (3), 489-505.

------, Jing Zhou, and Greg R. Oldham (2004), “The Effects of Personal and Contextual Characteristics on Creativity: Where should we go from here?,” *Journal of Management*, 30 (6), 933-58.

Shea, Christine M., and Jane M. Howell (1999), “Charismatic Leadership and Task Feedback: A Laboratory Study of their Effects on Self-Efficacy and Task Performance,” *The Leadership Quarterly*, 10 (3), 375-96.

Shimada, Koji, Yuki Ochi, Takuya Matsumoto, Hiroshi Matsugi, and Takao Awata (2014), “An Empirical Study of Electric Power Demand Control by Real-Time Feedback of Consumption Levels: Case of Nushima Island Households,” *Procedia Technology*, 18, 53-57.

Shmueli, Galit (2010), “To Explain or to Predict?,” *Statistical Science*, 25 (3), 289-310.

------, Soumya Ray, Juan Manuel Velasquez Estrada, and Suneel Babu Chatla (2016), “The Elephant in the Room: Evaluating the Predictive Performance of PLS Models,” *Journal of Business Research*, 69 (10), 4552-64.

Shrout, Patrick E., and Niall Bolger (2002), “Mediation in Experimental and Nonexperimental Studies: New Procedures and Recommendations,” *Psychological Methods*, 7 (4), 422-45.

Siahtiri, Vida (2018), “Innovation at the Service Encounter in Knowledge Intensive Business Services: Antecedents and Boundary Conditions,” *Journal of Product Innovation Management*, 35 (5), 742-62.

Sijbom, Roy B. L., Frederik Anseel, Michiel Crommelinck, and Alain De Beuckelaer (2018), “Why Seeking Feedback From Diverse Sources May Not Be Sufficient for Stimulating Creativity: The Role of Performance Dynamism and Creative Time Pressure,” *Journal of Organizational Behavior*, 39 (3), 355-68.

Simmering, Marcia J., Christie M. Fuller, Hettie A. Richardson, Yasemin Ocal, and Guclu M. Atinc (2015), “Marker variable Choice, Reporting, and Interpretation in the Detection of common Method Variance: A Review and Demonstration,” *Organizational Research Methods*, 18 (3), 473-511.

Simons, Robert (1987), “Accounting Control Systems and Business Strategy: An Empirical Analysis,” *Accounting, Organizations and Society*, 12, 357-74.

------ (1995), *Levers of Control*. Boston, MA: Harvard Business School Press.

------ (2000), *Performance Measurement and Control Systems*. Upper Saddle River, NJ: Prentice Hall.

Sims, Henry P., Andrew D. Szilagyi, and Robert T. Keller (1976), “The Measurement of Job Characteristics,” *Academy of Management Journal*, 19 (2), 195-212.

Singh, Jagdip (1993), “Boundary Role Ambiguity: Facets, Determinants, and Impacts,” *Journal of Marketing*, 57 (2), 11-31.

------ (2000), “Performance Productivity and Quality of Frontline Employees in Service Organizations,” *Journal of Marketing*, 64 (2), 15-34.

------, Garima Sharma, James Hill, and Andrew Schnackenberg (2013), “Organizational Agility: What It is, What It is Not, and Why It Matters,” *Academy of Management Proceedings*, 2013 (1), 1-40.

Sobel, Michael E. (1982), "Asymptotic Confidence Intervals for Indirect Effects in Structural Equation Models," *Sociological Methodology*, 13, 290-312.

Solomon, Michael R., Carol Surprenant, John A. Czepiel, and Evelyn G. Gutman (1985), "A Role Theory Perspective on Dyadic Interactions: The Service Encounter," *Journal of Marketing*, 49 (1), 99-111.

Speckbacher, Gerhard (2017), "Creativity Research in Management Accounting: A Commentary", *Journal of Management Accounting Research*, 29 (3), 49-54.

Speklé, Roland F., Hilco J. van Elten, and Sally K. Widener (2017), "Creativity and Control: A Paradox-Evidence from the Levers of Control Framework," *Behavioral Research in Accounting*, 29 (2), 73-96.

Sprinkle, Geoffrey B. (2003), "Perspectives on Experimental Research in Managerial Accounting," *Accounting, Organizations and Society*, 28 (2-3), 287-318.

Squillacciotti, Silvia (2010), "Prediction-Oriented Classification in PLS Path Modeling," in *Handbook of Partial Least Squares*, V.V. Esposito, W.W. Chin, J. Henseler, and H. Wang, eds. Berlin, Heidelberg: Springer, 219-33.

Steelman, Lisa A., and Kelly A. Rutkowski (2004), "Moderators of Employee Reactions to negative Feedback," *Journal of Managerial Psychology*, 19 (1), 6-18.

Stevens, Stanley Smith (1946), "On the Theory of Scales of Measurement," *Science*, 103 (2684), 677-80.

Stone, Mervyn (1974), "Cross-Validation and Multinomial Prediction," *Biometrika*, 61 (3), 509-15.

Storey, Chris, and Christine Larbig (2018), "Absorbing Customer Knowledge: How Customer Involvement Enables Service Design Success," *Journal of Service Research*, 21 (1), 101-18.

Stroop, J. Ridley (1935), "Studies of Interference in Serial Verbal Reactions, *Journal of Experimental Psychology*, 18, 643-62.

Sturm, Rachel E., Scott N. Taylor, Leanne E. Atwater, and Phillip W. Braddy (2014), "Leader Self-Awareness: An Examination and Implications of Women's Underprediction," *Journal of Organizational Behavior*, 35 (5), 657-77.

Swoboda, Bernhard, Cathrin Puchert, and Dirk Morschett (2016), "Explaining the Differing Effects of Corporate Reputation Across Nations: A Multilevel Analysis," *Journal of the Academy of Marketing Science*, 44 (4), 454-73.

Tafkov, Ivo D. (2013), "Private and Public Relative Performance Information under Different Compensation Contracts," *The Accounting Review*, 88 (1), 327-50.

Taylor, M. Susan, Cynthia D. Fisher, and Daniel R. Ilgen (1984), "Individual's Reactions to Performance Feedback in Organizations: A Control Theory Perspective," *Research in Personnel and Human Resources Management*, 2, 81-124.

Taylor, Scott N., and Jacqueline N. Hood (2011), "It may not be what you think: Gender Differences in predicting emotional and social Competence," *Human Relations*, 64 (5), 627-52.

Tenenhaus, Michel, Silvano Amato, and Vincenzo Esposito Vinzi (2004), "A Global Goodness-Of-Fit Index for PLS Structural Equation Modelling," *Proceedings of the XLII SIS Scientific Meeting*, 1, 739-42.

------, Vincenzo Esposito Vinzi, Yves-Marie Chatelin, and Carlo Lauro (2005), "PLS Path Modeling," *Computational Statistics and Data Analysis*, 48 (1), 159-205.

Tessier, Sophie and David Otley (2012), "A Conceptual Development of Simons' Levers of Control Framework," *Management Accounting Research*, 23 (3), 171-85.

Thiele, Kai Oliver, Marko Sarstedt, and Christian M. Ringle (2015), "A Comparative Evaluation of New and Established Methods for Structural Equation Modeling," *Proceedings of the 2015 Academy of Marketing Science Annual Conference*. Denver, CO: Academy of Marketing.

Thornock, Todd A. (2016), "How The Timing of Performance Feedback Impacts Individual Performance," *Accounting, Organizations and Society*, 55, 1-11.

Tierney, Pamela, and Steven M. Farmer (2002), "Creative Self-Efficacy: Its Potential Antecedents and Relationship to Creative Performance," *Academy of Management Journal*, 45 (6), 1137-48.

------, Steven M. Farmer, and George B. Graen (1999), "An Examination of Leadership and Employee Creativity: The Relevance of Traits and Relationships," *Personnel Psychology*, 52 (3), 591-620.

Trischler, Jakob, Simon J. Pervan, Stephen J. Kelly, and Don R. Scott (2018), "The Value of Codesign: The Effect of Customer Involvement in Service," *Journal of Service Research*, 21 (1), 75-100.

Uhl-Bien, Mary, and Michael Arena (2018), "Leadership for organizational Adaptability: A theoretical Synthesis and integrative Framework," *The Leadership Quarterly*, 29 (1), 89-104.

Unger, Rhoda K. (1979), "Toward a Redefinition of Sex and Gender," *American Psychologist*, 34, 1085-94.

Unsworth, Kerrie L., Toby D. Wall, and Angela Carter (2005), "Creative Requirement: A neglected Construct in the Study of Employee Creativity?," *Group & Organization Management*, 30 (5), 541-60.

Van der Borgh, Michel, Ad de Jong, and Edwin J. Nijssen (2019), "Balancing Frontliners' Customer- and Coworker-Directed Behaviors when Serving Business Customers," *Journal of Service Research*, 22 (3), 323-44.

Van der Stede, Wim A. (2001), "Measuring 'tight budgetary control'," *Management Accounting Research*, 12 (1), 119-37.

Van Veldhuizen, Roel (2017), "Gender Differences in Tournament Choices: Risk Preferences, Overconfidence or Competitiveness?," Discussion Paper, No. 14, Ludwig Maximilians-Universität München und Humboldt-Universität zu Berlin, Collaborative Research Center Transregio 190.

Vargo, Stephen L. and Robert F. Lusch (2008a), “Service-Dominant Logic: Continuing the Evolution,” *Journal of the Academy of Marketing Science*, 36 (1), 1-10.

------ (2008b), “Why “service”?,” *Journal of the Academy of marketing Science*, 36 (1), 25-38.

Vescio, Theresa K., Sarah J. Gervais, Mark Snyder, and Ann Hoover (2005), “Power and the Creation of patronizing Environments: The stereotype-based Behaviors of the Powerful and their Effects on female Performance in masculine Domains,” *Journal of Personality and Social Psychology*, 88 (4), 658-72.

Voorhees, Clay M., Michael K. Brady, Roger Calantone, and Edward Ramirez (2016), “Discriminant Validity Testing in Marketing: An Analysis, Causes for Concern, and Proposed Remedies,” *Journal of the Academy of Marketing Science*, 44, 119-34.

Wang, Laura W. (2017), “Recognizing the Best: The Productive and Counterproductive Effects of Relative Performance Recognition,” *Contemporary Accounting Research*, 34 (2), 966-90.

Wallas, Graham (1926), *The Art of Thought*. New York: Harcourt.

Wang, An-Chih, Jack Ting-Ju Chiang, Chou-Yu Tsai, Tzu-Ting Lin, and Bor-Shiuan Cheng (2013), “Gender makes the Difference: The Moderating Role of Leader Gender on the Relationship between Leadership Styles and Subordinate Performance,” *Organizational Behavior and Human Decision Processes*, 122 (2), 101-13.

Weisberg, Robert W. (1993), “Creativity: Beyond the Myth of Genius,” New York, NY: W. H. Freeman and Company.

Welch, Bernhard L. (1947), “The Generalization of Student's' Problem when Several Different Population Variances Are Involved,” *Biometrika*, 34 (1/2), 28-35.

West, Michael A. (2002), “Sparkling Fountains or Stagnant Ponds: An Integrative Model of Creativity and Innovation Implementation in Work Groups,” *Applied Psychology*, 51 (3), 355-87.

Wetzels, Martin, Gaby Odekerken-Schröder, and Claudia Van Oppen (2009), “Using PLS Path Modeling for Assessing Hierarchical Construct Models: Guidelines and Empirical Illustration,” *MIS Quarterly*, 33 (1), 177-95.

Whitney, Dean (2018), “Effects of Evaluation on Creativity,” in: *Creativity in Context: Updates to the Social Psychology of Creativity*, Teresa M. Amaible, eds. London, New York: Routledge, 131-52.

Widener, Sally K. (2006), “Associations between strategic Resource Importance and Performance Measure Use: The Impact on Firm Performance,” *Management Accounting Research*, 17 (4), 433-57.

Wilden, Ralf, and Siegfried P. Gudergan (2015), “The Impact of Dynamic Capabilities on Operational Marketing and Technological Capabilities: Investigating the Role of Environmental Turbulence,” *Journal of the Academy of Marketing Science*, 43 (2), 181-99.

Wilder, Kelly M., Joel E. Collier, and Donald C. Barnes (2014), “Tailoring to Customers’ Needs: Understanding How to Promote an Adaptive Service Experience with Frontline Employees,” *Journal of Service Research*, 17 (4), 446-59.

Wold, Herman O.A. (1982), “Soft Modeling: The Basic Design and Some Extensions”, in *Systems under Indirect Observations: Part II*, K.G. Jöreskog, and H.O.A. Wold, eds. North-Holland, Amsterdam, 1-54.

Wood, Wendy, and Alice H. Eagly (2002), "A Cross-Cultural Analysis of the Behavior of Women and Men: Implications for the Origins of Sex Differences," *Psychological Bulletin*, 128 (5), 699-727.

Wozniak, David (2012), "Gender Differences in a Market with relative Performance Feedback: Professional Tennis Players," *Journal of Economic Behavior & Organization*, 83 (1), 158-71.

------, William T. Harbaugh, and Ulrich Mayr (2016), "The Effect of Feedback on Gender Differences in Competitive Choices," Available at SSRN: https://ssrn.com/abstract=1976073 or http://dx.doi.org/10.2139/ssrn.1976073.

Wu, Amery D., and Bruno D. Zumbo (2008), "Understanding and Using Mediators and Moderators, " *Social Indicators Research*, 87 (3), 367-392.

Wu, Huiping, and Shing-On Leung (2017), "Can Likert Scales Be Treated as Interval Scales? - A Simulation Study," *Journal of Social Service Research*, 43 (4), 527-32.

Yoo, Jaewon and Todd J. Arnold (2016), "Frontline Employee Customer-Oriented Attitude in the Presence of Job Demands and Resources: The Influence Upon Deep and Surface Acting," *Journal of Service Research*, 19 (1), 102-17.

Young, S. Mark, Fei Du, Kelsey Kay Dworkis, and Kari Joseph Olsen (2016), "It's All about All of Us: The Rise of Narcissism and Its Implications for Management Control System Research," *Journal of Management Accounting Research*, 28 (1), 39-55.

Zeithaml, Valarie A., Leonard L. Berry, and A. Parasuraman (1988), "Communication and Control Processes in the Delivery of Service Quality," *Journal of Marketing*, 52 (2), 35-48.

Zhao, Xinshu, John G. Lynch, and Qimei Chen (2010), “Reconsidering Baron and Kenny: Myths and Truths about Mediation Analysis,” *Journal of Consumer Research*, 37 (3), 197-206.

Zhou, Jing (1998), “Feedback Valence, Feedback Style, Task Autonomy, and Achievement Orientation: Interactive Effects of Creative Performance,” *Journal of Applied Psychology*, 83, 261-76.

------, and Jennifer M. George (2001), “When Job Dissatisfaction Leads to Creativity: Encouraging the Expression of Voice,” *Academy of Management Journal*, 44 (4), 682-96.

------, and Jennifer M. George (2003), “Awakening Employee Creativity: The Role of Leader Emotional Intelligence,” *The Leadership Quarterly*, 14 (4-5), 545-68.