

Koppers, Sebastian

*An Investigation of Performance Feedback as a Management Practice –
Results from a Synthesis of Empirical Evidence and a Field Experiment*

Dissertation
for obtaining the degree of Doctor of Business and Economics
(Doctor rerum politicarum - Dr. rer. pol.)

at WHU – Otto Beisheim School of Management

July 12th, 2020

First Advisor: *Prof. Dr. Sascha L. Schmidt*

Second Advisor: *Prof. Dr. Christoph Hienerth*

Acknowledgements

When asked about his most important advice to doctoral students, Nobel laureate Richard Thaler (2017b) recommended to “make your research about the world, not the literature”. While working on the early stages of this dissertation, Thaler’s recommendation has struck a chord with me and guided my thinking through the numerous opportunities and challenges throughout this experience that will serve me across my lifetime. As a doctoral student at the Center for Sports and Management (CSM) at WHU – Otto Beisheim School of Management, I had the privilege to be part of a research community that intuitively lives up to this recommendation and strives for real-world applicability while encouraging me to pursue my own research interests at all times. Completing this dissertation would not have been possible without the continuous support, invaluable (performance) feedback and honest guidance from several people, to whom I would like to express my gratitude:

- To my first supervisor Prof. Dr. Sascha L. Schmidt, for taking on the role as a mentor, for never hesitating to invest in my development while allowing me to define my own research path and for trusting in my decisions with his support always in-reach.
- To my second supervisor Prof. Dr. Christoph Hienerth, for encouraging me to conduct this dissertation project in my own way and for providing valuable thoughts on how to align the research across all three papers.
- To Jun.-Prof. Dr. Dominik Schreyer, for his valuable support and honesty, for asking the right questions and for unselfishly allowing me to learn from the experiences he has already made.
- To Dr. Henry Eyring and to Patrick J. Ferguson for having the courage to become part of our joint research project in early 2017 and for allowing me to advance my academic skillset by demonstrating how high-quality research is to be conducted.

- To Dr. Johannes Tändl, Dr. Kathrin Santner, Roland Goriupp, Christof Sirk, Lucia Gjeltema and Dr. Michael Lang at Anton Paar SportsTec GmbH for providing expert knowledge from the domain of professional football and for making the randomized controlled trial possible despite the huge workload that has been associated with it.
- To the Laboratory for Innovation Science at Harvard (LISH), i.e. Prof. Dr. Karim R. Lakhani and Jin H. Paik, for welcoming me to the Lab community as an affiliated researcher and for providing guidance and recommendations during my visits.
- To my fellow research assistants at CSM and WHU – Daniel Beiderbeck, Florian Bünning, Dr. Klaus Eberhard, Nicolas Frevel, Dr. Florian Holzmayer, Dr. Mark Kassis, Felix Krause, Harry Krüger, Dr. Marc Lenz, Dr. Caroline Päßgen, Dr. Tobias Streicher, Lukas Lanz, Vera Schweitzer and Alexandra von Bernstorff – for becoming true friends, for always making time when I was in need of something and for keeping my spirits up at all times throughout this journey.
- To my colleagues at CSM, Maria Dahl, Kerstin Forword, Dennis Gottschlich, Sigrid Dethloff, and Renate Imoberdorf, for allowing me to focus on my research while taking care of many things.
- To my parents, sister, Philipp, and late grandmother, for their unconditional support throughout my whole life, for always having my back and for giving every ounce of energy left to support me even in the toughest of times – you are truly selfless role models that inspire me every day.

Finally, I will try to describe my gratitude to Sarah Och. You are always there for me while finding just the right words to keep me balanced and provide me with confidence to tackle every challenge. Having you by my side makes me truly happy and I cannot wait to see where we are heading.

Sebastian Koppers

Overview

Acknowledgements	II
Overview.....	IV
Table of Contents.....	V
List of Tables.....	VIII
List of Figures	IX
List of Appendices	X
List of Abbreviations.....	XI
List of Symbols.....	XIII
1 Introduction	1
2 Paper I: A Systematic Review of Performance Feedback Literature from Accounting, Management General and Organizational Behavior	17
3 Paper II: Less Information, More Comparison, and Better Performance: Evidence from a Field Experiment	83
4 Paper III: Providing Ranks as a Managerial Practice to Reduce Stress: How Performance Feedback Affects Individual Recipients	121
5 Conclusion	134
References	144
Appendix	162

Table of Contents

Acknowledgements	II
Overview.....	IV
Table of Contents.....	V
List of Tables.....	VIII
List of Figures	IX
List of Appendices	X
List of Abbreviations.....	XI
List of Symbols.....	XIII
1 Introduction	1
1.1 Background and Motivation	1
1.2 Research Questions and Theoretical Contribution	5
1.3 Research Methodologies and Datasets	9
1.4 Outline and Abstracts	13
1.4.1 Outline.....	13
1.4.2 Paper I: A Systematic Review of Performance Feedback Literature from Accounting, Management General and Organizational Behavior	14
1.4.3 Paper II: Less Information, More Comparison, and Better Performance: Evidence from a Field Experiment	15
1.4.4 Paper III: Providing Ranks as a Managerial Practice to Reduce Stress: How Performance Feedback Affects Individual Recipients.....	16
2 Paper I: A Systematic Review of Performance Feedback Literature from Accounting, Management General and Organizational Behavior	17
2.1 Introduction.....	17
2.2 Systematic Review Approach.....	28

2.2.1 Planning	28
2.2.2 Conducting.....	30
2.2.3 Reporting.....	35
2.3 Synthesis of the Empirical Performance Feedback Literature.	37
2.3.1 Definitions.....	37
2.3.2 Theories.....	42
2.3.3 Main Findings and Empirical Designs.....	47
2.4 Conclusion	78
3 Paper II: Less Information, More Comparison, and Better Performance: Evidence from a Field Experiment	83
3.1 Introduction.....	83
3.2 Theory and Hypotheses	87
3.3 Setting.....	94
3.3.1 Field Site	94
3.3.2 Industry and Incentives	96
3.3.3 Performance Constructs: Handling and Passing.....	97
3.4 Experiment Design	97
3.4.1 Subjects	97
3.4.2 Real-Effort Task.....	98
3.4.3 Treatment Conditions.....	99
3.4.4 Measures	101
3.5 Analysis	103
3.6 Conclusion	120
4 Paper III: Providing Ranks as a Managerial Practice to Reduce Stress: How Performance Feedback Affects Individual Recipients	121
4.1 Introduction.....	121

Table of Contents

4.2 Background.....	123
4.3 Experimental Design and Data	124
4.4 Results.....	128
4.5 Conclusion	131
5 Conclusion	134
5.1 Summary of Findings	134
5.2 Theoretical and Practical Implications	138
5.2.1 Theoretical Implications	138
5.2.2 Practical Implications.....	139
5.3 Future Research Directions.....	140
References	144
Appendix	162

List of Tables

Table 1. Overview and Positioning of Performance Feedback Reviews	25
Table 2. Selection Criteria and Article Base	34
Table 3. Performance Feedback Definitions in the Articles Most Cited Per Year.....	39
Table 4. Overview of Theories Investigated by Discipline	43
Table 5. Summary of Accounting Articles.....	56
Table 6. Summary of Management General Articles	59
Table 7. Summary of Organizational Behavior Articles	61
Table 8. Distribution of Environment Variables	66
Table 9. Distribution of Characteristics Variables	72
Table 10. Distribution of Person Variables	75
Table 11. Descriptive Statistics	103
Table 12. Effect of Relative Information and Detail in Performance Feedback	106
Table 13. Effect of Providing Both Absolute and Relative Performance Information on Reports of Number of Measures Being too Many.....	112
Table 14. Effect of Relative Performance Feedback on Social Comparison	114
Table 15. Relationship Between Social Comparison and Performance	115
Table 16. Effect of Relative Information and Detail in Performance Feedback, Partitioned by Conscientiousness	117
Table 17. Evidence of Attempt to Reduce <i>Fast_Time</i> by Reducing <i>Hitrates</i>	118
Table 18. Effect of Relative Information and Detail in Performance Feedback, Partitioned by Interest in Identifying Areas for Improvement.....	119
Table 19. Variable Definition, Measurement and Rationale for use.....	127
Table 20. Descriptive Statistics	128
Table 21. Determinants of MHRD	129

List of Figures

Figure 1. Structure of the Dissertation	13
Figure 2. Systematic Review Approach	28
Figure 3. Distribution of Articles Over Time per Discipline in our Article Base	36
Figure 4. <i>Hitrate</i> by Type of Measures in Performance Feedback	107
Figure 5. <i>Distance</i> by Type of Measures in Performance Feedback.....	107
Figure 6. <i>Avg_Time</i> by Type of Measures in Performance Feedback.....	108
Figure 7. <i>Fast_Time</i> by Type of Measures in Performance Feedback.....	108
Figure 8. <i>Hitrate</i> by Type and Number of Measures in Performance Feedback	110
Figure 9. <i>Distance</i> by Type and Number of Measures in Performance Feedback	110
Figure 10. <i>Avg_Time</i> by Type and Number of Measures in Performance Feedback.....	111
Figure 11. <i>Fast_Time</i> by Type and Number of Measures in Performance Feedback.....	111
Figure 12. Mean and Confidence Interval of <i>MHRD</i>	128

List of Appendices

Appendix 1. [Introduction] Transparency Table for Data Collected Based on the Pre-Registered Report (Koppers, 2019).....	162
Appendix 2. [Paper II] Layout of the Football Training Simulator skills.lab	163
Appendix 3. [Paper II] Real-Effort Task.....	163
Appendix 4. [Paper II] Target Visualization in the Real-Effort Task.....	164
Appendix 5. [Paper II] 3x2 Experiment Design.....	164
Appendix 6. [Paper II] On-Site Procedure of the Field Experiment	165
Appendix 7. [Paper II] Participant Flow During the Field Experiment	165
Appendix 8. [Paper II] Exemplary Performance Feedback Treatments.....	166
Appendix 9. [Paper II] Variable Definitions	167

List of Abbreviations

AMJ	Academy of Management Journal
APF	absolute performance feedback
bpm	beats per minute
CCL	Cahlíková et al., 2019
cm	centimeter
Comp.	compensation scheme
CRT	cognitive reflection test
CSM	Center for Sports and Management
e.g.	for example (Latin: ‘exempli gratia’)
EFK	Eyring et al., 2019 (Unpublished working paper)
et al.	and others (Latin: ‘et alii/aliae’)
ERIM	Erasmus Research Institute of Management
FE	fixed effects
H	hypothesis
i.e.	that is (Latin: ‘id est’)
JABA	Journal of Applied Behavior Analysis
JAP	Journal of Applied Psychology
JOBM	Journal of Organizational Behavior
kg	kilogram
LISH	Laboratory for Innovation Science at Harvard
M	mean
m	meter
Method.	methodology
MHRD	mean heart rate difference

List of Abbreviations

N	sample size
N/A	not applicable
OLS	ordinary least squares
p.	page
Partici.	participants
pct.	percentage
PF	performance feedback
RPE	relative performance evaluation
RPF	relative performance feedback
RPI	relative performance information
RQ	research question
RQs	research questions
s	seconds
SD	standard deviation
w	with
WHU	Wissenschaftliche Hochschule für Unternehmensführung

List of Symbols

*	statistical significance at the .1 level
**	statistical significance at the .05 level
***	statistical significance at the .01 level
◦	statistical significance at least at the .1 level
%	percent
R^2	coefficient of determination

1 Introduction

1.1 Background and Motivation

Performance Feedback (PF), which can be characterized as “a subset of information available to employees in their work environment” (Ashford & Cummings, 1983, p. 372), is prevalent in almost all organizations (Andiola, 2014). Organizations rely on PF as a managerial practice since it offers a potential solution for the challenging task to incentivize individuals to improve their work behavior – with effects at least equivalent to the effects of monetary incentive structures (Gosnell et al., 2020; Kramer et al., 2016; Luft, 2016; Song et al., 2018). Individuals have an intuitive curiosity to understand “how they are doing” (Sully De Luque & Sommer, 2000, p. 829) compared to their own standards or to their co-workers’ performance, particularly with regards to their career success (Jung et al., 2010). Therefore, the logic of providing individuals with PF is to address this pre-existing curiosity and to inform them about the status quo of their current level of performance (Bezuijen et al., 2009).

Organizations provide PF as a management practice to address this curiosity and try to close the productivity gap between individuals, thereby raising the average level of performance by facilitating learning and heightening motivation (Campbell, 2008; Gosnell et al., 2020; Lourenço et al., 2018). With regards to facilitating learning, PF can help individuals to identify superior ways to execute tasks at work (Song et al., 2018). In order to heighten motivation, managers use PF to create a visible gap between the current and preferred level of performance (Bezuijen et al., 2009). Consequently, individuals and teams within organizations rely on accurate and well-communicated PF to know which behaviors in the workplace are successful and therefore rewarded by employers (King et al., 2012). Without effective PF, these individuals and teams have no guidance for what they need to focus on to improve (Gonzalez-Mulé et al., 2016). By perceiving PF, individuals are forced to decide about either continuing the status-quo or redirecting the effort allocation in their work behavior (Dai et al., 2018).

In this dissertation, I aim to contribute to the guiding question of how organizations can apply PF as a management practice to motivate individuals and guide performance improvements by providing empirical evidence for the decisions about *what* information to choose and *how* to display them effectively. This guiding question arises for every executive within an organization that has implemented PF as a management practice or plans to do so, especially in human-capital intensive industries. Every PF system is defined by the decision about which type of information about past performance is presented in which way (Hannan et al., 2019; Lourenço et al., 2018). Since “information equals data plus meaning” (Checkland & Scholes, 1990, p. 303), providing information with a specific intent must not be confused with the pure presentation of raw data and organizations need to understand how PF recipients respond to the chosen display (Gill et al., 2019; Leiblein et al., 2017). Managers in organizations are forced to make return on investment decisions based on a thorough understanding of the specific PF characteristics (Tafkov, 2013), but also take into account the organizational structures, promotion schemes and individual incentives (Gill et al., 2019).

In an effort to inform the optimal design of PF systems, research on PF in organizations has a long-lasting tradition in the social sciences, particularly in the psychology and management-related literature. While experimental research on PF as a determinant of behavior has been conducted in psychology since the early 1900s (e.g. Arps, 1920; Brand, 1905), the management-related literature has been investigating the use of PF since the 1970s (Alvero et al., 2001; Sully De Luque & Sommer, 2000). Within the management-related literature, research on PF is especially prevalent in the accounting, management general¹ and organizational behavior literature since all three disciplines contribute to the organizational function of providing PF systems to affect employee behavior (S. E. Bonner & Sprinkle, 2002; Lourenço, 2016). More recently, these three disciplines have provided field experimental evidence on the effects of PF

¹ The specific wording of the discipline is based on the terminology applied by the Erasmus Research Institute of Management (ERIM).

in multiple industries, for example in health care (Song et al., 2018), education (Azmat et al., 2019; Eyring & Narayanan, 2018) or in maintenance companies (Casas-Arce et al., 2017).

Still, despite the increasing amount of studies on how to improve individual performance through the application of PF, the current body of research has produced a large variety of mixed results without many verifiable answers and has therefore largely been ignored by practitioners (S. S. K. Lam et al., 2002; Schnieder, 2018; Vancouver & Casey Tischner, 2004). While studies have reported that providing PF can lead to a large and long-lasting productivity increase (Blanes i Vidal & Nossol, 2011), not all PF is perceived as equal and review articles have shown that the application of particular PF characteristics can lead to either performance gains or losses (Schnieder, 2018). The meta-analysis by Kluger and DeNisi (1996) concluded that although PF generally has a positive effect on performance, “over 38% of the effects were negative” (Kluger & DeNisi, 1996, p. 258). The variety of results, however, is no surprise, given that PF is a heterogenous construct in itself (Alvero et al., 2001). The effects of PF on performance are dependent on many potentially confounding influences in the research design (Lourenço et al., 2018), for example the type (*what*) and format (*how*) of information, the source of PF or the provision of incentives (Alvero et al., 2001; Kluger & DeNisi, 1996).

The current ambiguity of mixed research results on how to apply PF as a management practice is especially relevant for managers today since the 21st century is characterized by “the changing nature of work” (Wegman et al., 2018, p. 353). Because “the world of work is now different than it was then, perhaps fundamentally so”² (Oldham & Hackman, 2010, p. 465), organizations are trying to adapt their PF systems to these conditions. Hence, improving PF design receives a high amount of managerial attention (J. L. Brown et al., 2016), but still, only two percent of organizations think that their current PF systems deliver the exceptional value they are expected to provide (Mercer, 2019).

² Oldham and Hackman (2010) refer to the 1970s as a reference point with “then”.

Within the new nature of work, three interdependent facets are particularly affecting the design of PF systems. First, rapid technological advances in recent years allow to track, record, analyze and display PF in real-time while managers have previously been limited to design PF based on specific, direct observations at limited time intervals (Bernstein & Li, 2017; Bhawe, 2014). The data sources for PF can now measure performance indicators in a reliable and objective way due to the rapid increase of available sensors that track data across all parts of an organization's value chain (Meeker, 2018), contrary to the previously used self-reported data from employees biased by subjectivity and misreports of results (Ewenstein et al., 2016). Second, due to the increasing complexity of jobs and the scarcity of talent, organizations focus on continuous development and learning to improve and engage employees as well as retaining them in their organization (Linderbaum & Levy, 2010). Therefore, effective PF systems are especially relevant to manage a modern workforce which wants to be actively involved in the process of achieving challenging and complex work results with a meaning to it (Oldham & Hackman, 2010; Wegman et al., 2018). Employees request PF systems that cater to individual differences in this process "to move away from one-size-fits-all approach[es]" (Linderbaum & Levy, 2010, p. 1389). Third, organizations are open to change previously established types of PF and start experimenting with new PF designs to create both accountability and performance improvements for PF recipients (Cappelli & Tavis, 2016). Relative ratings of PF, oftentimes expressed through ranking employees from worst to best, have been the most popular way to display PF across industries for multiple decades – even while withholding other types of available, e.g. absolute, information (Gill et al., 2019; Pope, 2009; Ray, 2007). Supported by the new technological possibilities, even the companies that traditionally relied exclusively on providing relative PF (RPF) to infer accountability for employee promotion or salary increases, for example GE or Microsoft, are implementing new ideas into their PF designs (Ewenstein et al., 2016). As a result, 15 percent of all global companies eliminated relative PF ranks from their PF systems and managers are trying to add PF design elements to the pre-existing systems

(Mercer, 2019). To face these three facets in designing PF systems in organizational practice, technological advancements, employee development and managers openness for change, “employee reviews are being modified, not abolished, and not necessarily for the better” (The Economist, 2016).

To contribute to the state of knowledge about the guiding question (*what* PF information to display *how*) and derive managerial implications for PF systems within the new nature of work, I will focus on synthesizing and providing empirical evidence on the effects of PF characteristics. In line with the increasing importance in both management and economics research on advancing the current state of knowledge through conducting clean field experimental designs with high levels of generalizability (Bloomfield et al., 2016; Feldman & Orlikowski, 2011; Floyd & List, 2016; Gneezy & Imas, 2017; Levitt & List, 2009), scholars have especially emphasized the lack of causal evidence in PF research for high-skilled individuals in industrialized countries (Gosnell et al., 2020; Schnieder, 2018).

1.2 Research Questions and Theoretical Contribution

The guiding question for this dissertation, *what* type of PF information to choose and *how* to display them effectively, will be further segmented into three research questions (RQs) that address separate research gaps in PF research. First, I will synthesize and analyze the empirical evidence on the effects of PF as well as the underlying research methodologies and theories across three management-related disciplines – namely accounting, management general and organizational behavior – to identify differences, similarities, and research gaps. Second, I will test the individual performance effects of the type of performance information in interaction with the number of measures included in the display of PF through the lens of social comparison theory as well as considering theories of information overload and salience. Third, I will build on this analysis and illuminate how the type of performance information in the PF display and

individual person variables affect recipients' stress levels. Each research paper will investigate one research question (RQ), these are:

RQ I: What is the state of knowledge on the effects of PF based on empirical evidence in the three management-related disciplines of accounting, management general and organizational behavior and which definitions, theories and research designs are applied?

RQ II: What type of display of PF information – absolute, relative, or absolute and relative – and which quantity of information works best to motivate individuals and improve performance?

RQ III: How does the type of display of PF information – absolute, relative, or absolute and relative – as well as individual person variables affect stress levels of individual PF recipients?

RQ I contributes to the literature since no prior literature review has synthesized and analyzed the empirical evidence on the performance effects of PF designs with a distinct focus on identifying similarities and differences across the different streams of management-related literature through a systematic approach. Recent reviews provided important advancements with a specific focus on certain sub-types of PF (e.g. Schnieder, 2018) or certain predetermined feedback characteristics in a theoretical model (e.g. Lechermeier & Fassnacht, 2018) to address the multidimensional nature of PF effects (Ilgen et al., 1979; Kluger & DeNisi, 1996). But since “few concepts in psychology have been written about more uncritically and incorrectly than that of feedback” (Latham & Locke, 1991, p. 224) and the term “has simply become professional slang” (Peterson, 1982, p. 102), the heterogeneity of PF definitions, synonyms and research applications up until today has considerably slowed down the progress in advancing the understanding of PF effects and generalizability of findings within and across disciplines (Anseel et al., 2015; Schnieder, 2018). Through a systematic review with the aim to identify

and compare the relevant empirical evidence on the effects of PF in the management-related disciplines of accounting, management general and organizational behavior, also considering the PF definitions, theories and research designs applied, I aim to inform the progress of the research stream based on a holistic understanding of the current state of evidence on PF interventions (Eppler & Mengis, 2004; Snyder, 2019).

RQ II contributes to the stream of research examining the individual performance effects of specific PF design choices through the lens of social comparison theory (e.g. Casas-Arce et al., 2017; Eyring & Narayanan, 2018; Gill et al., 2019; Hannan et al., 2019; Song et al., 2018). Social comparison theory predicts that RPF affects performance because individuals have “a drive to evaluate [his] opinions and [his] abilities” (Festinger, 1954, p. 117). Within this stream of research, RPF is most often expressed through ranking past performance in a descending order, comparable to the application of RPF in business and additional settings to determine salary increases, promotions or awards (Gill et al., 2019). In addition, prior to new technological advancements, RPF has already been simple to apply for organizations – and in research designs – because RPF only requires to assign ranks and not evaluate and determine the absolute performance level for an individual on a meaningful metric (Prendergast, 1999). Individual performance effects of RPF differ on a multitude of PF design choices and result in positive, negative or no effects on performance (Schnieder, 2018). Studies find that perceiving detailed RPF rather infrequently leads to better performance (Casas-Arce et al., 2017), that publicly disclosed RPF displays improves performance through best-practice learning over private ones (Song et al., 2018) and that individuals ranked first or last improve the most after receiving it (Gill et al., 2019). Because “the literature to date has tended to overemphasize the ‘R’” (Kachelmeier, 2019, p. 3) in RPF, research still has limited knowledge about how RPF affects individual performance (Kramer et al., 2016) in comparison to absolute PF (APF) even though providing only APF has also been found to affect performance positively (Bandiera et al., 2015). To the best of my knowledge, no prior study has specifically compared the performance

effects of providing either relative, absolute or both absolute and relative PF (Moore & Klein, 2008; Prendergast, 1999; Schnieder, 2018). In addition, I consider the performance effects of the quantity of PF information to examine additional theories of salience and information overload that lack an empirical investigation in PF research so far. Standard economic theory predicts that among Bayesian rational agents, more available information is superior than less to aid decision-making (Savage, 1954), but theory about salience predicts that individuals do not pay equal levels of emphasis to all information (Bordalo et al., 2012). Information overload theory predicts that receiving more information than an individual can process may interfere with the ability to execute a task (Eppler & Mengis, 2004). I aim to address this specific research gap by comparing these alternative approaches without interfering monetary incentives and investigate the distinct effect of information on individual performance improvements.

RQ III contributes to the PF literature by investigating how individual recipients of PF perceive the stress levels caused by absolute, relative, or both absolute and relative PF displays. Understanding the holistic effects of PF on recipients is important since organizations decide about which type of PF information to provide, but recipients decide if and how to act based on the PF display and how it is perceived (Dai et al., 2018; Schnieder, 2018). In order to understand how recipients react to PF, organizations need to be informed about how the display of PF affects the recipients' stress levels because higher individual stress levels affect individual decision making in various ways, for example through increasing risk aversion (Cahlíková & Cingl, 2017) or lowering competitiveness (Cahlíková et al., 2019). While stress can impact individual PF recipients, consequences on the aggregated level are even more severe for organizations and societies since five to eight percent of the annual healthcare costs in the U.S. alone are attributed to workplace stressors (Goh et al., 2016). Prior literature in ergonomics and consumer decision making has found that the amount of information provided to make decisions influences individual physiological conditions through anxiety and stress (Ettema & Zielhuis, 1971; Malhotra, 1982), but to the best of my knowledge, no prior study has

investigated the effects of types of PF displays on stress levels. In addition, I will also consider the effect of cognitive mechanisms and additional person variables, that is the perceived value of PF and the trust in the PF source, on the perceived level of stress.

1.3 Research Methodologies and Datasets

While all three RQs are derived from the guiding question of *what* type of PF information to choose and *how* to display them effectively, the corresponding research gaps are distinct and contribute to different streams of the PF literature. Therefore, I conduct three stand-alone, yet interlinked research papers, to investigate each RQ. All three papers and the applied research approaches specifically address the current lack of causal evidence to advance the understanding of using PF as a management practice in the conducted research designs and methodologies (Gosnell et al., 2020; Schnieder, 2018).

In Paper I, I conduct a systematic review that builds upon the guidelines by Snyder (2019) in general and by Tranfield et al. (2003) for management-related research. These guidelines emphasize the necessity of investigating specific RQs in a transparent way to synthesize and compare empirical, quantitative evidence in “areas in which the research is disparate and interdisciplinary” (Snyder, 2019, p. 333) and where methodologies are too heterogenous to conduct a meta-analysis. This systematic type of literature review is “as a research method more relevant than ever” (Snyder, 2019, p. 333), because the depth, rigor, replicability and usefulness differentiates this method from traditional narrative reviews, which are potentially flawed due to non-specific RQs and a lack of details on how the analysis was conducted (Palmatier et al., 2018).

The systematic review approach in Paper I is based on multiple discussions with researchers who have published peer-reviewed articles on PF. Based on these discussions, the scope of the review is to analyze quantitative evidence of PF effects across the three management disciplines of accounting, management general and organizational behavior. Therefore, a systematic

review protocol is realized that identifies 521 individual articles published between 1999 and 2018 in the primary set of 17 (STAR) journals for the three management disciplines from the ERIM Journal list (EJL) 2016-2021³. All 521 articles were read up until an individual assessment of three inclusion criteria was possible, which specify the scope in accordance with Paper I's RQs, i.e. that each article must (i) use quantitative methods to analyze experimental evidence, (ii) establish a relationship between PF characteristics and its effects and (iii) measure the effect of PF on the PF recipient. As a result, 87 articles fulfill all three criteria and five types of information are stored from each article: (1) *environment variables*, (2) *characteristics variables*, (3) *person variables*, (4) *theory* and (5) *main finding(s)*. Lastly, the final 87 articles are assessed individually and analyzed in an interdisciplinary and visual way, with the aim to highlight key challenges in PF research and derive recommendations for future research providing quantitative evidence.

For Paper II and Paper III, professional sports – and football⁴ in particular – is identified as an ideal setting to gather causal, empirical evidence. The setting of professional sports provides a unique labor market laboratory to conduct empirical analysis (Kahn, 2000), especially in the face of strong incentives (e.g. Allen et al., 2017; Massey & Thaler, 2013). Sports data in general offers five methodological advantages to study human behavior and economic decision making by individuals contrary to other field settings: (I) the observability of real behavior, (II) the existence of well-defined rules of the sport, (III) the presence of high individual incentives, (IV) the access to experienced professionals in their actual environment and (V) the possible versatility in the analysis of the data (Balafoutas et al., 2019).

All five methodological advantages are present in the field site *skills.lab* for Paper II and for Paper III, a state-of-the-art football training simulator. The sample consists of professional and

³ The EJL is compiled by researchers from its two founding institutions: Rotterdam School of Management at Erasmus University and Erasmus School of Economics.

⁴ Throughout this dissertation, the term “football” refers to European football, alternatively known as “soccer” in some countries.

semi-professional players from football clubs in Austria. First, I can observe real behavior since these athletes play drills in the simulator that mimic real training and game scenarios and measure real football performance in metrics that players and coaches are used to. Second, the training drills within the simulator adhere to the same basic rules of football that are governed globally. Third, high individual economics are present in our setting since players have naturally occurring incentives to performance based on individual career concerns, due to limited career spans with permanent competition for roster spots and starting positions, as well as through self-image concerns, since players derive parts of their identity from being perceived as elite performers by larger crowds, either through live attendance in stadiums or fans and followers on television or social media. Fourth, the simulator allows to observe players in their real environment since these athletes use the simulator regularly in their training routines. Fifth, the simulator allows for versatile data collection and it offers multiple opportunities for seamless experimental interventions since players regularly make entries on computer interfaces prior to or after the training where surveys can be incorporated to gather scarce psychological or cognitive evidence.

The combination of these five advantages in this field setting establishes high external validity and complements disadvantages of standard laboratory experiments, which typically rely on university students performing on tasks invented for the scope of the study, that translate into results that may not generalize to other settings and people (Gneezy & Imas, 2017). This “issue of generalizability” (Gneezy & Imas, 2017, p. 5) is especially relevant when experimental evidence is applied to inform researchers and managers about the effectiveness of theories. In a similar vein, this field setting provides the level of internal validity established through standard laboratory experiments since the fully automated nature of the simulators allows to introduce clean and randomized manipulations to measure causal effects.

Within this field setting, intricate data at the individual player level is collected through a pre-registered randomized controlled trial with the AEA RCT Registry⁵ (Koppers, 2019) for Paper II and Paper III. Based on a cooperation agreement for the execution of this randomized controlled trial with the producers of skills.lab, Anton Paar SportsTec, and in partnership with LISH, I design and conduct a field experiment in which the PF perceived by 117 professional and semi-professional players is manipulated during training. As part of the experiment, players receive PF on two key metrics which they are accustomed to – passing accuracy and speed – after completing a standardized passing drill. To implement the experimental treatments, a 3x2 between-subjects experimental design is employed in which players are randomly assigned into one of the six groups to perceive PF displayed as (1) either their absolute, relative, or absolute and relative performance (captured via rank); and (2) PF that contains either aggregate or detailed PF information (that is a smaller or larger number of performance measures). With the intent to address the current lack of complimentary evidence to explain underlying PF mechanisms (Schnieder, 2018), pre- and post-survey data is collected, which consists of demographic variables, multiple PF-related items as well as personality and cognitive traits. The design allows to identify the distinct, causal effects of PF without confounds (Schnieder, 2018) because the field setting and the intervention do not include monetary rewards or explicit performance incentives while strong implicit incentives, e.g. in the form of good training performances potentially translates into playing time, are present.

To test for causal effects, Paper II and Paper III follow guidance from econometric research to employ multiple ordinary least squares (OLS) regressions with heteroscedastic-robust standard errors (White, 1980) that control for covariates (Angrist & Pischke, 2008), as applied in prior field experimental research on the effects of PF (e.g. Eyring & Narayanan, 2018). In Paper II, I investigate the performance effects of the type of display of PF information on

⁵ The AEA RCT Registry is ‘The America Economic Association’s registry for randomized controlled trials’.

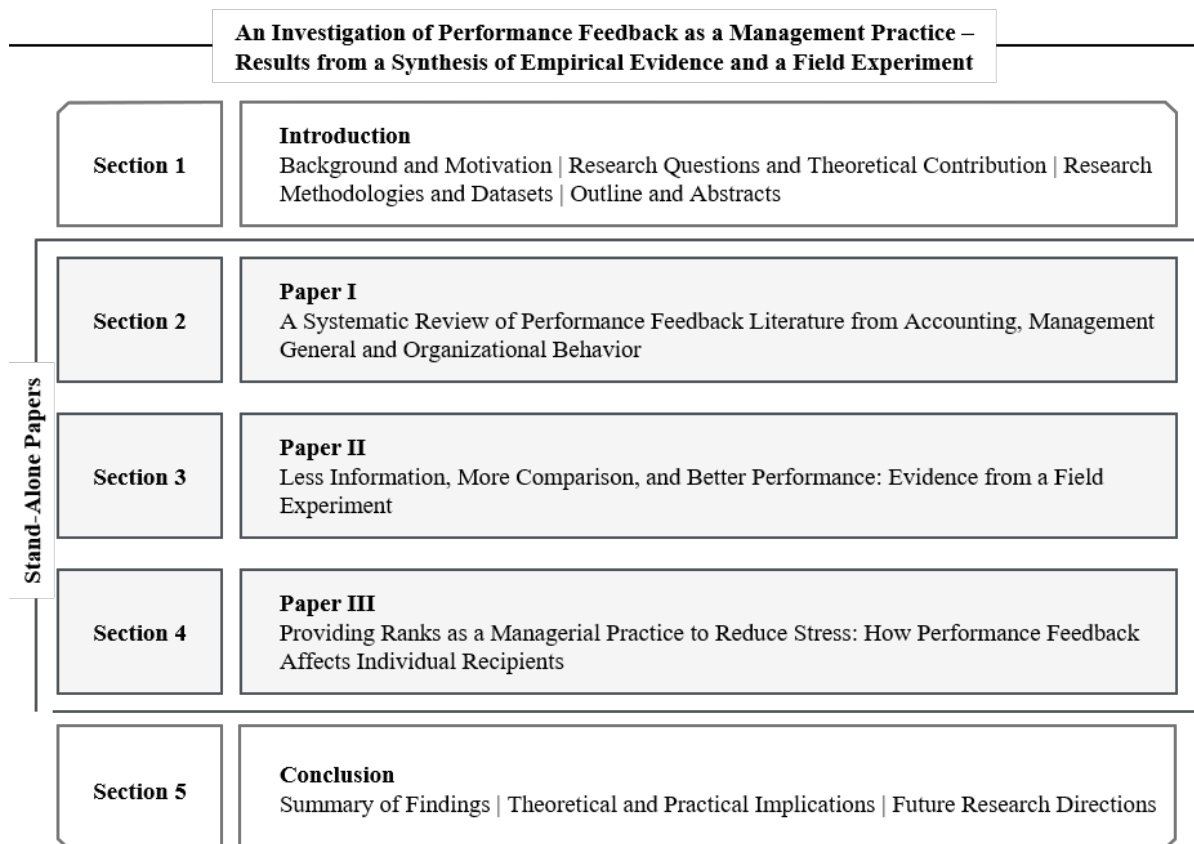
performance (RQ II) via four performance measures that are presented in the PF manipulations – two of the measures reflect accuracy and two of the measures reflect speed. In Paper III, I investigate the effects of the type of display of PF information on stress levels (RQ III) via unique heart rate data on an individual level for a sub-sample of players. Heart rate is an established proxy to capture individual stress in management research (e.g. Cahliková et al., 2019) since stress leads to an immediate physiological increase in heart rate (Kirschbaum et al., 1993). A further breakdown of which variables of the field experiment are analyzed in Paper II and III is provided in the data transparency table in Appendix 1.

1.4 Outline and Abstracts

1.4.1 Outline

This dissertation proceeds as visualized in Figure 1 with three stand-alone research papers.

Figure 1. Structure of the Dissertation



Section 1 concludes with the abstracts of the three papers. In sections 2 to 4, I present the three respective research papers that address the guiding question of this dissertation on *what* type of PF information to choose and *how* to display them effectively. Section 5 concludes this dissertation by summarizing the findings of all three papers, presenting the theoretical and practical implications as well as suggesting directions for future research.

1.4.2 Paper I: A Systematic Review of Performance Feedback Literature from Accounting, Management General and Organizational Behavior

PF refers to the management practice of presenting recipients with specific information to enable them to improve their performance. Due to the prevalence of PF across organizations and the consequences for an organization's performance associated with it, a review of PF is critical for researchers and practitioners alike. This paper presents a systematic review of research on the effects of PF in three management related disciplines – accounting, management general and organizational behavior. Our results indicate that PF research has advanced the understanding of specific variables. Still, research across disciplines shows a lack of cross-fertilization resulting in a current fuzziness about the understanding of the concept. Through this systematic review, we make three main contributions to provide a synthesis of the current state of management-oriented PF research and suggest opportunities for future scholarly attention for this literature. First, we develop an integrative definition to organize PF research. Second, we identify differences and similarities in the application of theories to advance the current understanding of their predictive power and allow considering theoretical insights culled from psychology or behavioral economics to explain the effects of PF on recipients. Third, we present current findings and research designs applied in PF research to provide transparency about the differences in the integration of key variables and to emphasize the future adaptation of field settings in combination with survey data.

This paper is co-authored by Prof. Dr. Sascha L. Schmidt. It has been submitted for publication to a peer-reviewed accounting journal and is currently under review. Minor adjustments have been made to maintain consistency within this dissertation.

1.4.3 Paper II: Less Information, More Comparison, and Better Performance: Evidence from a Field Experiment

We use a field experiment in professional sports to compare three approaches to PF – providing absolute, relative, or the combination of absolute *and* relative PF. Bayesian economic theory predicts that more information is better for performance, but behavioral economic forces that influence decision-making may make it optimal to draw attention to some information by omitting other information. We find that RPF alone yields the best performance effects – i.e., that a subset of information (relative) dominates the full information set (absolute and relative PF together) in boosting performance. To explain this, we consider theories of information overload and salience. We do not find evidence that restricting the number of measures per se in PF benefits performance, which suggests that our result is not driven by information overload. Rather, we find that restricting the scope of PF to relative measures makes social comparison more salient, benefiting performance.

This paper is co-authored by Dr. Henry Eyring (London School of Economics and Political Science) and Patrick J. Ferguson (Harvard Business School). This paper has been presented at the 55th annual *Journal of Accounting Research* conference (on September 18, 2020) and is currently under review at the associated journal, the *Journal of Accounting Research*. Minor adjustments have been made to maintain consistency within this dissertation.

1.4.4 Paper III: Providing Ranks as a Managerial Practice to Reduce Stress: How Performance Feedback Affects Individual Recipients

Most prior research on PF investigates the effects of RPF on performance with only minimal attention to how feedback affects individual recipients, even though the display of PF to individual recipients is an established management practice across a range of settings and industries. Because the effects of managerial practices on employees' health outcomes have important financial and performance implications for employers, we study how the provision of three different types of PF – absolute, relative, or both absolute and relative feedback – affects recipients' stress levels. Using data from a field experiment in professional football, we additionally investigate how the perceived value of feedback, trust by the recipient as well as cognitive traits influence stress measured via heart rate. Our results show that providing individuals with only RPF leads to a significant reduction of stress for the recipient. While individuals with high values for perceived value of feedback and for trust also experience lower levels of stress, more deliberate thinkers feel higher levels of stress. These results suggest that employers can potentially apply RPF to not only improve employee performance but also reduce individual stress levels in the workforce.

This paper is co-authored by Prof. Dr. Sascha L. Schmidt and Jun.-Prof. Dr. Dominik Schreyer and is currently being prepared for submission to a peer-reviewed journal.

2 Paper I: A Systematic Review of Performance Feedback Literature from Accounting, Management General and Organizational Behavior⁶

2.1 Introduction

In this paper, we present a systematic review of the literature on the effects of PF across three management disciplines, namely accounting, management general and organizational behavior. While the concept of feedback is context dependent and used in multiple ways, management related disciplines most commonly associate the term feedback with various forms of presenting recipients with specific information about their performance, for example on behaviors, strategies or outcomes (Atkins et al., 2002; Ilgen et al., 1979; Kluger & DeNisi, 1996).

PF is an established management practice across organizations to enable individuals to improve their performance (Gosnell et al., 2020; Lourenço et al., 2018). Organizations rely on individual performance improvements of their employees to raise the average performance level by implementing superior ways to execute tasks at work and PF can enable learning from other employees (Song et al., 2018). PF is also intended to be perceived as a motivational tool, for instance to close the productivity gap between individuals (Song et al., 2018). Individuals have a general interest in receiving feedback since it addresses their general interest in being informed about “how they are doing” (Sully De Luque & Sommer, 2000, p. 829). Disclosing PF also forces recipients to make decisions – either consciously or unconsciously – about their allocation of effort or other scarce resources by either continuing the status quo or redirecting these resources going forward (Dai et al., 2018; Sengul & Obloj, 2017). These individual trade-offs are especially pronounced in multi-task environments which represent the majority of

⁶ Koppers, S., & Schmidt, S. L. (2020). A Systematic Review of Performance Feedback Literature from Accounting, Management General and Organizational Behavior. Unpublished Working Paper.

modern workplaces (Hannan et al., 2013). Scholars have even concluded that the consequential effects of feedback on individual performance can be at least equivalent to the effects of monetary incentive structures (Kramer et al., 2016; Luft, 2016).

Therefore, organizations are presented with the task to determine which information are presented to recipients through which specific type of PF. Information must not be confused with the mere presentation of raw data (Leiblein et al., 2017), since “information equals data plus meaning” (Checkland & Scholes, 1990, p. 303). The task to determine which information is presented in in which way is inevitable after making the decision to provide PF. Consequently, an organization’s management is required to understand under which conditions PF can generate a positive effect to compare the advantages to the costs (Tafkov, 2013). These return on investment decisions not only have to take into account the content and timing of feedback to design effective policies (Tafkov, 2013), but also organizational structures, incentive schemes and individuals’ implicit incentives and preferences to grasp an understanding of how people respond to PF (Gill et al., 2019).

Several examples illustrate the prevalence of PF in human-capital intensive organizations across industries, with or without explicitly linking it to rewards or monetary incentives (Nordstrom et al., 1991; Schnieder, 2018). For instance, scholars have investigated situations in which PF is provided to employees in retail banking (Gino & Staats, 2011), in the public sector (Charness et al., 2014) or in health care (Song et al., 2018). Similarly, PF is also provided to non-employees in additional settings, for example in education (Azmat et al., 2019; Eyring & Narayanan, 2018), in energy consumption (Allcott, 2011) or on online recommendation web sites (Y. Chen et al., 2010).

The practical relevance of PF has further increased in recent years due to rapid technological progress (Bernstein & Li, 2017). This technological progress makes the design of PF within organizations even more challenging since the amount of collected data and the speed to distribute it has increased considerably (Casas-Arce et al., 2017). Digital performance

monitoring systems allow to continuously track, record, analyze and display PF. Based on these systems, organizations have new possibilities to implement PF contrary to traditional methods, which were based on direct observations by managers at specific time intervals (Bhave, 2014). New possibilities in designing PF systems will further emerge due to the predicted tenfold increase of available real-time data within organizations until 2025 (Reinsel et al., 2017). These data generating sources are also becoming independent from stationary workplaces across employee levels through the availability of sensors and the respective data (Meeker, 2018). The continuous availability of real-time data also facilitates the emergence of the managerial trend of gamification, in which feedback is combined with challenges to compete to spur performance within organizations (The Economist, 2012). Amazon, for instance, already launched multiple tools and processes to provide real-time PF across devices (E. Smith, 2018), while it was previously quite common to see charts or graphs of performance taped on boards in public work areas (LePine et al., 2002).

Multiple managerial research disciplines investigate the effects of PF to identify transferable findings for managers such that feedback can be used as a tool to improve performance in a variety of corporate and societal settings. Experimental research on behavioral effects, caused by the application of different types of PF in psychology, has been conducted since the early 1900s (Kluger & DeNisi, 1996). On the contrary, studies investigating the performance effects of feedback in the context of organizational behavior management date back to the 1970s (Alvero et al., 2001). Probably due to the increased prevalence of PF based on the technological progress, studies on the performance effects of feedback in management general and accounting have been predominantly executed within in the 2000s (e.g. Charness et al., 2014; Hannan et al., 2008; Tiefenbeck et al., 2018).

Still, there is no established consensus on the effects of PF across research disciplines due to the heterogeneity of definitions for PF, the variety of PF interventions and individual influences. PF can be defined in multiple ways and its implementation can differ on a multitude of

characteristics, e.g. the source of information, the medium and format of the information or its frequency (Alvero et al., 2001; Balcazar et al., 1985). Individual performance also has several antecedents, e.g. effort or risk taking, and research on PF oftentimes specifically focuses on one of them. Performance can be jointly influenced by several person characteristics (e.g. skill or personality traits) or environment characteristics (e.g. feedback setting and compensation scheme) (Schnieder, 2018). For instance, studies have shown that people “aspire to excel as compared to a standard of excellence” (Jung et al., 2010, p. 727), but at the same time two individuals do not have to reach the same conclusions based on the same set of information presented to them (Tallon et al., 2018). Contrary to economic theories, which assume that individuals make rational choices to maximize a utility function based on an appropriate and time-consistent processing of information (DellaVigna, 2009), individuals “do not use information optimally as traditional economic theory says they should” (Thaler, 2017a, p. 1803). Therefore, various theories are investigated in the scope of PF effects – within and across research disciplines – with the objective to advance the understanding of the underlying mechanisms.

Although the importance of PF is acknowledged in both theoretical and empirical work (Alder & Ambrose, 2005), investigating PF across managerial disciplines prompted a current fuzziness about the understanding of the concept, the antecedents, and the effects. Moreover, the increased amount of research over the last decades has produced a collection of different definitions (Alvero et al., 2001), various synonyms and different incentive schemes being investigated leading to conflicting results for specific types of PF (Schnieder, 2018). As a result, the constantly growing body of research on this topic is still fragmented and findings differ between disciplines.

With the intention of providing a more comprehensive – and less fragmented – understanding of the research conducted on PF, this paper provides a synthesis of the current literature based on a systematic review and thorough synthesis of prior research in management related

academic publications. The methodological choice of a systematic literature review employs specific criteria for inclusion and exclusion of studies from the review (Tranfield et al., 2003) combined with the application of a unique quality threshold (Bouncken et al., 2015). This approach allows us to analyze 87 studies from three managerial disciplines covering 17 top peer-reviewed journals by investigating three synonyms to create an exhaustive and replicable coverage. Each of the 87 studies was entered into a data repository specifically structured to synthesize the research characteristics.

Following Snyder (2019), our systematic review serves to build the foundation for advancing knowledge on the effects of PF and facilitating theory building by highlighting research areas in which an excess of studies exists and where further research is needed. To help researchers and managers alike to be informed about the current research on PF, three relevant RQs are addressed in our review covering the identified research gaps along the three management disciplines of accounting, management general and organizational behavior:

***RQ1.** How can the term PF be defined across the three disciplines?*

***RQ2.** Which theories are applied to investigate the effects of PF in each of the three disciplines?*

***RQ3a.** What are the main findings on the effects of PF in each of the three disciplines?*

***RQ3b.** Which research designs are operationalized to investigate the effects of PF in each of the three disciplines?*

While our paper is not the first review on the concept of PF, it is the first review to synthesize the research along the three RQs across three management disciplines. In line with the interdisciplinary review on the concept of information overload by Eppler and Mengis (2004), we aim to identify synergies and differences among specific management disciplines to enable further research. Each of the three management disciplines is included because of its relevance within the research on PF (Lourenço, 2016). Firstly, studies from the field of accounting are

included in this review because accountants play a key role in designing information and feedback systems and processes in organizations (Indjejikian, 1999; Schnieder, 2018; Tafkov, 2013). Secondly, studies from management general are in the scope of this review as well since researchers from the field of economics (e.g. Charness et al., 2014; Gill et al., 2019; Tiefenbeck et al., 2018) and further management related perspectives publish studies on PF with high relevance for managers in journals within this discipline. Thirdly, studies from organizational behavior are also part of the scope of this review because research on PF has – by tradition – been conducted in this discipline since it is a management tool in organizations (Balcazar et al., 1985; Ilgen et al., 1979; Lourenço, 2016). This set of disciplines allows for a holistic yet focused synthesis of the state-of-research.

While we complement prior reviews, our paper specifically addresses the current research gap of a systematic review of PF research across management-related disciplines. Table 1 provides an overview of the positioning of our literature review and emphasizes the key differences we identified and subsequently addressed with our systematic review approach. Summarizing prior reviews on PF since 1979, we find that 11 review articles have been published on the topic with no author publishing more than one review. Due to the heterogeneity in the understanding and application of PF research variables, only two of the reviews are meta-analytic reviews (Anseel et al., 2015; Kluger & DeNisi, 1996). The growing level of relevance of PF is reflected in the temporal distribution of these 11 articles as well. Most of these reviews, i.e. seven articles, are written after the year 2001 with four of the review articles being conducted in the last ten years (Andiola, 2014; Anseel et al., 2015; Lechermeier & Fassnacht, 2018; Schnieder, 2018). By comparing these 11 reviews concisely, our systematic review is different in its inherent scope and methodology along five criteria. In sum, these five criteria allow us to provide an understanding of the complex research area of PF contrary to prior approaches (Snyder, 2019).

First, no prior review of PF research has a dedicated scope of comparing three management disciplines. Most reviews have a specific focus on one discipline while a few identify and select

studies from all management disciplines without a filter or deliberate differentiation. While one of the most cited reviews with management relevance (Kluger & DeNisi, 1996) has been published in the field of psychology, three reviews are set in or emphasize the discipline of accounting, four reviews are set in or emphasize the discipline of organizational behavior and three reviews are set in or emphasize the discipline of management general. Recent reviews state to investigate “discipline-spanning literature” (Lechermeier & Fassnacht, 2018, p. 3) or “studies across different fields of research” (Schnieder, 2018, p. 15), but their review protocols only allow for a rather random identification of multidisciplinary studies based on very specific keywords and do not deliberately compare them.

Second, no prior review has been conducted with specific journal restrictions along a top peer-review focused quality threshold. Balcazar et al. (1985) and the subsequent update of the review by Alvero et al. (2001) specifically limit the set of journals, but there is no proxy or quality threshold being elaborated upon explaining why these four journals are selected.

Third, our review deliberately selects relevant articles based on three selection criteria to address the four RQs within the scope of our review. While other (narrative) reviews only briefly mention inclusion criteria, only Schnieder (2018) conducted a systematic review based on Tranfield et al.’s (2003) methodology with an emphasis on a list of selection criteria. But since Schnieder (2018) has a specific scope, i.e. RPF in accounting and is aligned around a conceptual framework within that scope, the methodology is not linked to transdisciplinary RQs.

Fourth, our systematic review enables us to investigate a larger-than-average amount of relevant articles while having a dedicated scope and associated RQs. Only six of the PF literature reviews report the number and details of the articles being analyzed. These six literature reviews report an average number of 54 articles (see Table 1) resulting from their processes, while our review synthesizes 87 articles from top peer-reviewed journals.

Fifth, the scope of our review is defined differently compared to all prior reviews due to its interdisciplinary scope and systematic, quality-focused process with the goal to identify synergies and differences across management disciplines. Prior reviews have a very specific or discipline-focused scope on, for example, feedback seeking (e.g. Anseel et al., 2015; Ashford et al., 2003), organizational behavior (Balcazar et al., 1985), auditing (e.g. Andiola, 2014) or focus on investigated pre-defined characteristics in a conceptual model (e.g. Lechermeier & Fassnacht, 2018).

The remainder of this paper is organized as follows: Section 2 describes the systematic methodology applied in this review; Section 3 synthesizes the main features of the final body of articles and provides the results for each RQ from an in-depth review of the articles; Section 4 concludes this review by discussing future directions for PF research.

Table 1. Overview and Positioning of Performance Feedback Reviews

Authors ^a	Discipline ^b	Type of review ^c	Journal restrictions ^d	Selection criteria	Articles resulting	Scope
<i>Ilgen et al. (1979)</i>	Organizational behavior ¹	Literature review	N/A	N/A	N/A	Objective and process-oriented review of the multidimensional nature of feedback organizing the recipients' perception, acceptance, and willingness to respond to feedback.
<i>Balcazar et al. (1985)</i>	Organizational behavior ¹	Literature review	AMJ, JABA, JAP, JOBM (1975-1985)	(1) Feedback as an independent variable (2) No analog and lab studies	69	Review experimental applications and cataloguing of feedback characteristics in organizations and distinguish the effects from goal setting and behavioral consequences.
<i>Nordstrom et al. (1991)</i>	Organizational behavior ²	Literature review	Psychology and business (1967-1987)	(1) Utility to the practicing manager (2) Measurable and observable (3) Cause-and-effect relationship (4) Detail for replication	21	Review of the effects of public posting of PF in work settings in contrast to prior general reviews of PF.
<i>Kluger & DeNisi (1996)</i>	Psychology ²	Meta-analytic review	None, mostly published papers and technical reports in English	(1) Feedback treatment group without other manipulations (2) Control group without Feedback (3) Measuring of performance (4) Ten participants or more sampled (5) Statistics to calculate <i>d</i> statistic	131	Historical review of feedback interventions on performance and providing a meta-analysis to investigate the large variability of effects on performance by different characteristics.
<i>Alvero et al. (2001)</i>	Organizational behavior ²	Literature review	AMJ, JABA, JAP, JOBM (1985-1998)	(1) Feedback as an independent variable (2) No analog and lab studies (3) Specific term feedback, no other synonyms	43	Update of the literature review by Balcazar et al. (1985) to compare their findings with the more recent literature (1985-1998) and investigate the impact the prior literature review. Additional mechanisms were also investigated.
<i>Bonner & Sprinkle (2002)</i>	Accounting ¹	Literature review	None	(1) Lab or field experiments (2) Effects of monetary incentives on individual effort/performance (3) No tasks without normative performance criterion	N/A	Focus on the effects of performance-contingent incentives and how they interact with environment variables, i.e. feedback.
<i>Ashford et al. (2003)</i>	Management general ¹	Literature review	N/A (1983-2003)	(1) Focus on outcomes from feedback seeking, not of feedback information itself	N/A	Organizing research on feedback-seeking behavior along three motives to discuss the respective impact of different feedback seeking characteristics.
<i>Andiola (2014)</i>	Accounting ¹	Literature review	Psychology and management	N/A	N/A	Synthesizing and reviewing (non-comprehensively) the feedback literature in management and psychology focusing on the effects of feedback on learning, performance, and motivation in auditing.

Table 1. Overview and Positioning of Performance Feedback Reviews (Continued)

Authors ^a	Discipline ^b	Type of review ^c	Journal restrictions ^d	Selection criteria	Articles resulting	Scope
<i>Anseel et al. (2015)</i>	Management general ¹	Meta-analytic review	None, published and unpublished research (1983-2011)	(1) Measure feedback-seeking behavior based on the definition (2) Only information seeking about performance appraisal information (3) Studies in actual or closely simulated organizational context	69	Presenting a meta-analytic investigation of the relationship between feedback seeking behavior and its antecedents and outcomes as well as the relationship of feedback seeking behavior with inquiry and monitoring as separate dimensions.
<i>Lechermeier & Fassnacht (2018)</i>	Management general ²	Literature review	None	(1) Empirical studies on effects of feedback source, timing, or valence on recipients' reaction (2) Studies investigating moderating effects of source, message, task, or individual characteristics on the main effect (3) Studies with only one feedback source or one timing of feedback (4-8) Specific exclusion criteria to streamline the analysis	64	Organizing and summarizing research findings on the main effects of feedback source, feedback timing and feedback valence and the respective interaction effects with the feedback source, message, task, and recipients' individual characteristics.
<i>Schnieder (2018)</i>	Accounting ²	Literature review	None, peer-reviewed and non-peer-reviewed journals from multiple disciplines	(1) English full-text available (2) Direct or indirect effects of RPF on employee performance or theoretically linked (3) Allow for clean inferences about effects of RPF and adhere to discipline's research approaches (4) Contrasting effects with a control group or with different manifestation levels (5-8) Specific exclusion criteria to streamline the analysis	64	Providing an overview of empirical findings on how management accountants can use RPF to affect employee performance through multiple forms of behavior as well as the interaction effects with characteristics, task, environment, and person variables.

Table 1. Overview and Positioning of Performance Feedback Reviews (Continued)

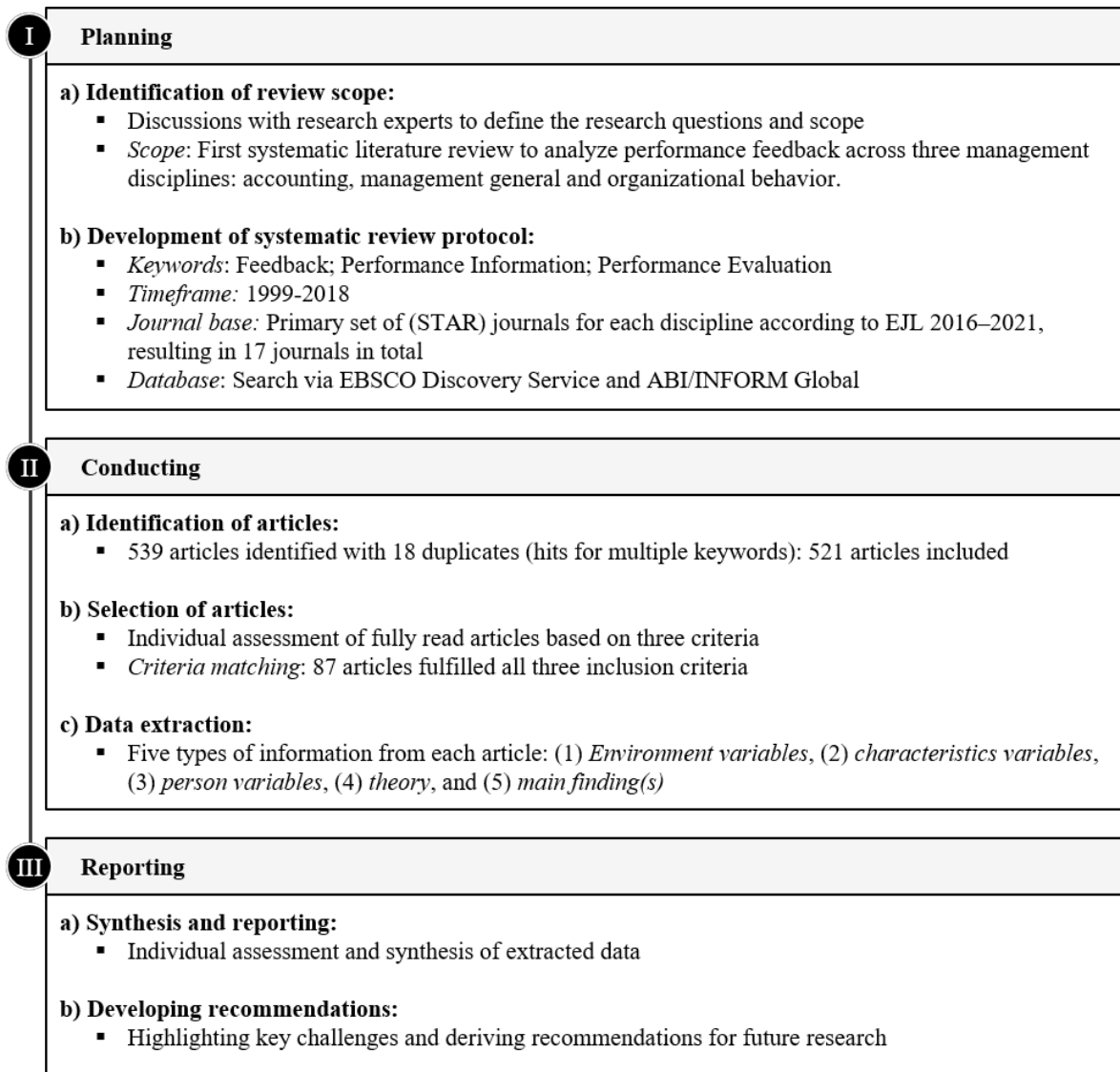
Authors ^a	Discipline ^b	Type of review ^c	Journal restrictions ^d	Selection criteria	Articles resulting	Scope
<i>Our paper</i>	Accounting, management general & organizational behavior	Literature review	Primary Set (STAR) of journals per discipline EIJ (1999-2018)	(1) Article must use experimental settings and empirical quantitative methods (2) The direct or indirect relationship is established between PF characteristics and PF effects. (3) The effect must be measured on the level of individuals or groups of individuals. Hence, the unit of analysis is the PF recipient rather than the sender or the organization.	87	Systematically review and analyze research on the effects of PF across three management disciplines to identify similarities and differences among the different perspectives. The review concentrates on current PF definitions, theories and research designs applied as well as the main findings to establish synergies between the different streams of research and highlight future research directions.

Notes: ^a Authors listed in chronological order. ^b Categorized ¹based on EIJ 2016-2021 ²based on journal homepage or self-description within the review, not categorized in EIJ 2016-2021. ^c Categorized as either literature review or meta-analytic review based on the methodology of the review. ^d Categorized based on self-description within the review; AMJ = Academy of Management Journal, JABA = Journal of Applied Behavior Analysis, JAP = Journal of Applied Psychology, JOBM = Journal of Organizational Behavior.

2.2 Systematic Review Approach

The methodology of our systematic literature review is based upon the recommendations by Snyder (2019) and Tranfield et al. (2003) to synthesize the current fragmented evidence base of PF research in a transparent and reproducible way. Figure 2 presents a detailed overview of the three main stages.

Figure 2. Systematic Review Approach



Notes: Methodology based on Snyder (2019) and Tranfield et al. (2003).

2.2.1 Planning

The first step in planning the systematic review was to discuss the current state of the PF research with research experts from management-related disciplines, which had authors among them, who previously published studies on the topic of PF in a journal of our journal base.

These discussions resulted in (i) the specific research scope and management disciplines to be investigated in our review, (ii) the four RQs to be addressed and (iii) the definition of both the search criteria and PF synonyms to conduct the review.

Within the step of planning the systematic review, we also specified the systematic review protocol and the respective search strategy. First, we defined three keywords to gather a holistic yet concise picture of the PF research: “feedback”, “performance information” and “performance evaluation”. In line with Schnieder (2018), we identified “performance information” and “feedback” as relevant synonyms and therefore, included both in our search strategy. We also included “performance evaluation” as a third keyword since the keyword generally ties PF to monetary incentives (Frederickson, 1992). Since we evaluate the compensation scheme used in each article, this combination is also in the scope of our review. For our first keyword, “feedback”, we refrained from including the specification of “performance” to collect an exhaustive collection of articles in the identification phase and used our subsequent search criteria to only select the applications relevant for performance. For the two other keywords, which are otherwise rather generic in nature, the specification of “performance” was necessary to only search within the scope of this review. Based on searching for these three keywords, we extend prior literature reviews which did not include synonyms for feedback in their search strategy (e.g. Alvero et al., 2001; Lechermeier & Fassnacht, 2018; Schnieder, 2018).

Second, we only searched in academic journals within the primary subset of ‘STAR’ journals according to the EJM 2016-2021⁷ to implement a quality threshold in our systematic review protocol. ERIM refers to this set of journals as “the absolute top ones among the best journals in the field” and “considered as truly distinctive” (2018). The EJM 2016-2021 includes six management disciplines and we limited our search strategy to the 17 primary journals in this

⁷ For this article, the EJM 2016-2021 version updated in September 2018 was applied (ERIM, 2018).

list for ‘Finance & Accounting’, ‘Management General’ and ‘Organization’.⁸ To the best of our knowledge, no prior review on PF has applied a quality threshold in its search strategy.

Third, we selected the electronic databases EBSCO Discovery Service and ABI/INFORM Global for our search strategy. These two databases allowed us to search for our three keywords in the titles and abstracts of published articles within these 17 journals.

Fourth, we limited our search strategy to the more recent PF literature from 1999 to 2018. This period of twenty years allowed us to investigate and compare the publication trends in each of the three management disciplines. In addition, we avoid a redundant synthesis since Alvero et al. (2001) limited their search to the timeframe from 1985 to 1998 and two out of their four journals in organizational behavior are also included in our journal base. We included articles within this period that were published online in advance and were published in print issues of the journals later (e.g. Gill et al., 2019).

2.2.2 Conducting

Based on the systematic review protocol (see Figure 2), we conducted the search and identified 539 articles across all 17 journals: 380 articles for the keyword “feedback”, 32 articles for the keyword “performance information” and 127 articles for the keyword “performance evaluation”. These results – that each keyword resulted in a distinct body of articles – gave proof for the validity of the search terms to collect a holistic overview. In addition, these results confirmed our decision that all three keywords are within the scope of our review, since 18 articles could be found under a combination of two of the keywords. After deleting the duplicates, the final article base included 521 articles.

Then, all articles were imported into the reference management software MendeleyTM and all 521 articles were reviewed to select the articles which were relevant to answer the four RQs in the scope of our paper. Thus, three criteria were derived to find relevant studies: to review

⁸ We further focus on the three disciplines that provide empirical research on PF as a management practice, e.g. only accounting journals and no finance journals out of the ERIM category ‘Finance & Accounting’.

experimental studies which establish *PF effects* on the *PF recipient*. The three inclusion criteria were defined as follows:

1. The article must use empirical quantitative methods on data gathered in an experimental setting.
2. The article must establish a direct or indirect relationship between PF characteristics and PF effects.
3. The article must measure the effect of PF on the level of individuals or groups of individuals. Hence, the unit of analysis is the PF recipient rather than the sender of the PF or an organization.

For inclusion, articles first had to investigate PF in an empirical setting since we are interested in exploring causal behavioral effects of PF (Levitt & List, 2009). This criterion is in line with recent findings by Schnieder (2018), who reported that most studies (94 percent) of the article base for his review (focused on one subcategory of PF, i.e. RPF) were experimental studies. Still, this criterion makes our study different from prior reviews which either focused only on field experiments (e.g. Alvero et al., 2001) or did also include further empirical methods, for instance surveys (e.g. Lechermeier & Fassnacht, 2018). The second criterion for inclusion was that articles had to establish a direct or indirect relationship between PF characteristics and effects. Through this criterion, our systematic review protocol accounted for both the heterogeneity in feedback characteristics as well as their respective effects in the current body of research. Therefore, studies that do not investigate a relationship between a PF characteristic and effect needed to establish an indirect relationship to be informative for our research scope. We consciously do not limit our review to feedback effects on performance, since performance itself is a term with a heterogeneous understanding and is highly context-dependent (Jung et al., 2010). In addition, performance includes many dependent variables, e.g. effort, human capital, employee interaction or budgeting (Schnieder, 2018), which could be linked to the main category of performance or which are used as a synonym or substitute for it (e.g. Gill et al.,

2019). Another reason why we refrain from limiting our scope to the effects of feedback on performance is that prior reviews found that the effect of this relationship is “quite variable” (Kluger & DeNisi, 1996, p. 254). Even though the literature traditionally assumes that feedback leads to subsequent performance improvements (Casas-Arce et al., 2017), prior reviews reported positive effects as well as negative or null effects (Alvero et al., 2001; Kluger & DeNisi, 1996; Schnieder, 2018). Hence, as our review focuses on a synthesis of the current fragmented state of research across three management disciplines, our primary interest is to present a holistic overview which allows researchers to further enhance our current understanding. At last, the third criterion determined that only studies were included that investigate effects of PF on recipients of this feedback, either as an individual or in groups.

All 521 identified articles from the review were read and checked for each of the three inclusion criteria. Hence, 368 articles (70.6 percent) fulfilled the first criteria, 118 articles (22.6 percent) met the second criteria and 196 articles (37.6 percent) satisfied the third criteria. If articles did not match each of the three criteria, they were excluded from our review. The main motives for exclusion were:

1. The use of the term feedback was not related to PF, for example by utilizing “feedback loop” or “feedback mechanism” within a different research topic.
2. The article did conduct a review or relied on non-experimental data (such as surveys) or the article did not conduct an empirical analysis within an experimental setting but built a theoretical model.
3. The unit of analysis was not the recipient of PF, but for example the feedback sender (e.g. in the context of leader-member exchange studies), an individual seeking feedback or an organization.
4. The research design did not allow to disentangle the effects of PF characteristics on recipients’ effects – either direct or indirect.

As a result, a total of 87 articles (17 percent) fulfilled all three inclusion criteria and remained in the final article base for this review. Table 2 provides a condensed overview about the distribution of all identified and selected articles from 17 journals across three disciplines within our systematic review protocol. Three conclusions can be derived based on the distribution characteristics of the final article base. First, all three management disciplines are relevant for the scope of our review. In line with the longest research history on PF, the 46 articles from the discipline of organizational behavior represent the majority in our article base. Still, the 25 articles from the discipline of accounting and the 16 articles from the discipline of management general add validity to the selection of the three management related disciplines. Second, specific journals within each discipline are more prone to publish research on the effects of PF. Eight out of the 17 journals do not include an article that has been identified for our review. From the remaining nine journals, each published an average of 9.7 identified articles within the timeframe (standard deviation (SD) 8.6). Third, “feedback” is the most relevant keyword (with 78 selected articles) compared to the two synonyms “performance information” and “performance evaluation” (with a combined nine selected articles). While the inclusion of all three keywords is confirmed to be relevant, a strong consensus within the body of research to rely on the term “feedback” can be established.

Table 2. Selection Criteria and Article Base

	<i>"Feedback"</i>			<i>"Performance information"</i>			<i>"Performance evaluation"</i>			<i>Total</i>			
	Identified	Selected	Share	Identified	Selected	Share	Identified	Selected	Share	Identified	Selected	Share	
Accounting	<i>The Accounting Review</i>	11	5	45%	7	2	29%	27	1	4%	45	8	18%
	<i>Accounting, Organizations and Society</i>	15	8	53%	4	1	25%	20	0	0%	39	9	23%
	<i>Contemporary Accounting Research</i>	26	4	15%	3	1	33%	9	0	0%	38	5	13%
	<i>Journal of Accounting Research</i>	14	3	21%	2	0	0%	8	0	0%	24	3	13%
	<i>Journal of Accounting and Economics</i>	2	0	0%	2	0	0%	4	0	0%	8	0	0%
	<i>Review of Accounting Studies</i>	3	0	0%	0	0		12	0	0%	15	0	0%
Management general	<i>The Academy of Management Journal</i>	21	2	10%	0	0		3	0	0%	24	2	8%
	<i>The Academy of Management Review</i>	13	0	0%	0	0		1	0	0%	14	0	
	<i>Administrative Science Quarterly</i>	8	0	0%	0	0		0	0		8	0	
	<i>Journal of Management</i>	34	1	3%	1	0		2	0	0%	37	1	3%
	<i>Management Science</i>	63	12	19%	3	1	33%	9	0	0%	75	13	17%
Organizational behavior	<i>Journal of Applied Psychology</i>	70	15	21%	2	0	0%	9	1	11%	81	16	20%
	<i>Journal of Business Venturing</i>	12	0	0%	0	0		1	0	0%	13	0	
	<i>Journal of Management Studies</i>	12	0	0%	1	0	0%	2	0	0%	15	0	
	<i>Organization Studies</i>	10	0	0%	0	0		0	0		10	0	
	<i>Organizational Behavior and Human Decision Processes</i>	43	28	65%	3	1	33%	4	1	25%	50	30	60%
	<i>Research Policy</i>	23	0	0%	0	0		2	0	0%	25	0	0%
Total	380	78	21%	28	6	21%	113	3	3%	521	87	17%	

Notes: Overview of selected articles from each journal based on the search for three keywords.

We extracted information from each of the 87 identified articles by qualitative means to ensure a thorough analysis of each respective article. From each article, we retrieved and stored (I) general information from the articles (title, authors, year, journal, discipline, etc.) as well as additional information relevant to answer our four RQs. For these additional information categories for PF, we relied upon our PF definition and stored (II) *environment variables* (for instance country, research method, setting, sample, occupation of participants or compensation scheme), (III) *characteristics variables* (for instance research design, definition of PF type, independent variables, dependent variables, source, privacy, recipient, realism, medium or frequency) and (IV) *person variables* (for instance self-efficacy, self-consciousness, ability, personality traits, cognitive abilities/traits or performance segments). In addition, we also retrieved (V) the *theoretical background* as well as (VI) the *main finding(s)* for each article.

2.2.3 Reporting

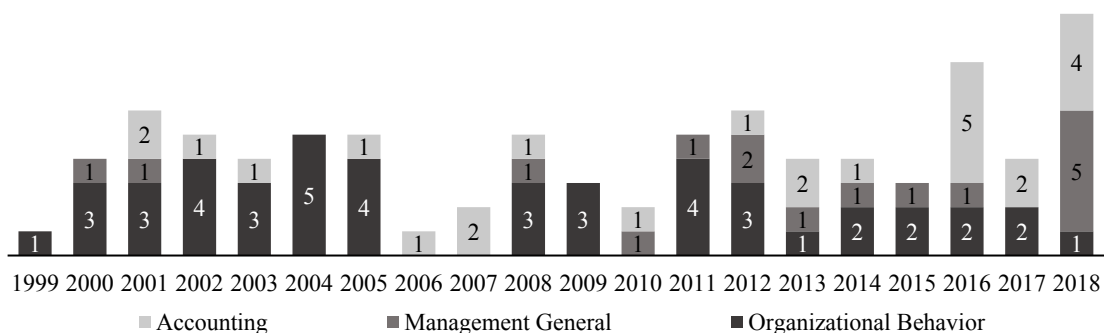
To conclude the systematic review, the aggregated information for all articles were then compared and analyzed in line with the recommended process by Snyder (2019), leading to the synthesis as presented in this review. Following the lines of the interdisciplinary review by Eppler and Mengis (2004), we provide the results of our systematic review in a visual format to make our synthesis and summary statistics visible in a highly applicable way.

Before presenting the synthesis of the extracted data to addresses each of the four guiding RQs accordingly, we first provide an overview over the year distribution in our article base – in total and by discipline. In total, the number of articles within the scope of this review has been growing steadily since 1999. In 1999, only one article (in the discipline of organizational behavior) was published. Later in the year 2018, ten articles (across all three management disciplines) were published. Both numbers accordingly represent the lowest and highest value. An average of 4.4 articles (SD 2.2) has been published per year that satisfied all three inclusion criteria. By splitting the twenty-year period in two halves, the chronological development within each of the three disciplines can be illustrated. In total, 38 articles on the effects of PF

within the scope of this review have been published from 1999 to 2008. In the second half of the sample, from 2009 to 2018, an additional 29 percent of articles (49) have been published. Within each of the three management related disciplines, the trend in percentages is more heterogeneous. In accounting, the number of articles selected for this review has increased from nine to 16 by 78 percent. In management general, the increase in percentage is larger, from three to 13 articles (333 percent), but the total amount is still lower as compared to the two other fields. While organizational behavior still accounts for the largest number of articles selected, the number of articles published in the respective time period has decreased from 26 in the first to 20 in the second half of the time period (minus 23 percent). This development is also reflected within the average age of the articles. Hence, the average age of articles in the discipline of accounting is 6.5 years (SD 6.0), the average age of articles in the discipline of management general is 5.4 years (SD 5.7) and the average age of articles in the discipline of organizational behavior is 10.5 years (SD 5.6).

Based on this distribution, recent claims from multiple studies can be confirmed. In line with the large research history, investigations of the effects of PF are still predominantly being conducted in the discipline of organizational behavior, but the two additional disciplines of accounting and management general increasingly contribute to the research debate. Figure 3 presents the development of the numbers of articles within our review.

Figure 3. Distribution of Articles Over Time per Discipline in our Article Base



Notes: Distribution from the year 1999 to 2018 for the 87 selected articles.

2.3 Synthesis of the Empirical Performance Feedback Literature.

Having described the methodology we apply to review the PF literature in three management disciplines, we next present the synthesis of PF research to address the four RQs of interest.

2.3.1 Definitions

To map the existing intellectual territory of PF, we set out to answer the first RQ: *How can the term PF be defined across the three disciplines (RQ1)?* This RQ is of relevance to advance the future understanding of the research topic, because feedback is an umbrella term which can be applied in a variety of contexts, for instance products, performances or also attributes (Fong et al., 2019). Interdisciplinary literature reviews on other concepts, for example Eppler and Mengis (2004) on ‘information overload’, stated that “the everyday use of the term [information overload] has led to various constructs, synonyms, and related terms” (p.326) which led to a fragmented stream of research – the same holds true for PF. While Anseel et al. (2015) identified that “the diversity of conceptualizations and measurements has slowed progress in the field” (p.319) of feedback-seeking behavior, we conclude for PF as well, that the current research – within and across management disciplines – lacks a consensual and integrative definition for the term. Although there are articles dedicated to address the lack of a definition of feedback and suggest their own definitions in management theory – “Feedback is information about the gap between the actual level and the reference level of a system parameter which is used to alter the gap in some way” (Ramaprasad, 1983, p. 4) – these definitions have not been referenced once in our selected articles within the systematic review and are therefore not facilitating an aligned stream of research across disciplines as intended.

In prior reviews on PF, the scope has been traditionally on investigating the effect sign and strengths of PF on performance, how these relationships interact with additional variables and what functions they can serve (e.g. Alvero et al., 2001; Kluger & DeNisi, 1996; Lechermeier & Fassnacht, 2018). First and foremost, the majority of literature reviews (e.g. Andiola, 2014) still

relies upon the general definition of PF by Kluger and DeNisi (1996): “actions [sic] taken by (an) external agent(s) to provide information regarding some aspect(s) of one’s task performance” (p.255). Recent reviews on the topic either define the specific type of PF they focus on and name synonyms (e.g. Schnieder, 2018) or restrain from discussing or elaborate upon definitions but rather focus on defining certain feedback characteristics within their guiding framework (e.g. Lechermeier & Fassnacht, 2018).

Overall, there has not been a research debate in recent reviews on how to precisely phrase the key term of the research topic, even though Alvero et al. (2001) specifically point to a commentary by Peterson (1982) who advised to not use the term ‘feedback’ anymore due to the implicit ambiguity as a result of a long history of heterogeneous applications. Since the term ‘feedback’ is still used in the large majority (89.7 percent) of the selected studies in our review, defining PF precisely is still of high relevance to allow interdisciplinary research to advance a common understanding and distinguish it from the status of a “generic label” (Nordstrom et al., 1991, p. 102)

For an overview of how various researchers across the three management disciplines define PF in their respective articles, we present exemplary definitions per discipline in Table 3. We chose these exemplary definitions by ranking the articles within each discipline by citations as a proxy for the most relied upon articles that facilitate the current state of knowledge to the next cohort of research articles. Following a thorough analysis, we conclude that a similar pattern emerges within the individual articles to the one that we identified in prior reviews on PF.

Table 3. Performance Feedback Definitions in the Articles Most Cited Per Year

	Article	Cited per year ^a	Definition of performance feedback or type of feedback investigated
Accounting	<i>Hannan et al. (2013)</i>	23	“Private RPI ^b involves employees knowing only their own relative standing compared to peers, and public RPI involves employees knowing not only their own relative standing, but also the relative standings of each of their peers.” (p.557)
	<i>Hannan et al. (2008)</i>	22	“Because performance feedback systems can differ in terms of the quality of relative performance information that is provided, we explore two levels of feedback by varying the precision of the feedback’s content. We vary the precision of the content by varying the number of partitions into which we divide the relative performance feedback (halves and deciles)” (p. 895)
	<i>Tafkov (2013)</i>	21	“In the private RPI ^b condition, participants receive the rank of their own performance within the group. In the public RPI condition, participants receive their own performance rank and the ranks of each of the other four participants in their group.” (p.334)
	<i>Casas-Arce et al. (2017)</i>	18	“We vary the frequency of feedback information (weekly vs. monthly) and the level of detail included in the report (the average score of all jobs performed during the period by a professional vs. the scores for each of the individual jobs)” (p.1052)
	<i>Fisher et al. (2002)</i>	16	“Under such relative performance evaluation (e.g., benchmarking), subordinates receive information regarding the activities and performance of comparable referents.” (p.853)
Management general	<i>Charness et al. (2014)</i>	70	“The ranking treatment is similar to the baseline treatment except that each participant is now informed about her relative performance.” (p.40)
	<i>Tiefenbeck et al. (2018)</i>	65	“By contrast, our intervention provided individuals with real-time feedback on a specific behavior while and where they engaged in it.” (p.1460)
	<i>Gill et al. (2018)</i>	59	“Rank-order RPE ^b , in which pay, promotion, employee appraisals, and nonpecuniary awards depend on the rank of individuals in the distribution of performance.” (p.1)
	<i>Song et al. (2018)</i>	56	“RPF ^b refers to the practice in which an organization provides information to each worker about her performance on a specific metric relative to her coworkers (Blanes i Vidal and Nossol 2011). For example, this may be in the form of a ranked list that shows a worker’s position on the distribution of productivity in the organization.” (p. 2628)
	<i>Shunko et al. (2018)</i>	37	“Workers use feedback about their actual performance relative to a goal to self-regulate behavior (Donovan 2001, Bendoly et al. 2010).” (p. 455-456)
Organizational behavior	<i>VandeWalle et al. (2001)</i>	47	“Participants were provided knowledge of their performance by the return of their graded exam in the next class session after the first exam. The feedback consisted of comments on essay exam answers and the number of points out of 100 that they had received on the exam. The participants were also provided with the class mean and standard deviation statistics displayed on the class whiteboard.” (p. 635)
	<i>DeShon et al. (2004)</i>	46	“The first form of feedback provided participants only with information about their own performance (e.g., individual score). Twenty-six teams (78 participants) were assigned to this individual feedback condition. The second form of feedback provided trainees only with information about their team’s overall performance (e.g., team score). Twenty-seven teams (81 participants) were assigned to this team feedback condition. Finally, the third form of feedback provided trainees with information about both their own performance and their team’s overall performance (e.g., individual score, team score).” (p. 1041)
	<i>Moreland & Myaskovsky (2000)</i>	44	“Providing them with feedback about one another’s performance before they worked together as a group.” (p.121)
	<i>Ilies & Judge (2005)</i>	33	“In this respect, performance feedback is important because it allows individuals to evaluate their previous performance relative to a specific goal or standard.” (p. 454)
	<i>Sheldon et al. (2014)</i>	28	“Participants received feedback about their scores.” (p. 127)

Notes: ^a Citation counts are based on results of Google Scholar on October 22nd, 2019. ‘Cited per year’ is the average number of citations per year, calculated by dividing the total citation count of an article by the number of years since it has been published until the end of the time frame of this literature review in 2018. ^b Abbreviations from the articles are the following: Relative Performance Information (RPI), Relative Performance Evaluation (RPE).

Within these 15 articles, authors most commonly define the particular type of PF they investigate within their research designs, but only seldomly link these types to the overarching

research topic. In the few cases where a general definition of PF was stated (e.g. Shunko et al., 2018), the definition was targeted to incorporate the relationship of PF with the specific theoretical background of the study. Authors in the discipline of accounting define key parts of, for instance, the PF environment (e.g. Hannan et al., 2008), or define the PF characteristics (e.g. Tafkov, 2013), but do not link it back to a general definition of PF. Similarly, authors in the discipline of management general refer to the specific PF characteristics under investigation (e.g. Song et al., 2018). Only Shunko et al. (2018) emphasize a feedback definition that incorporates not only PF characteristics but also emphasizes the within-person process by the recipient. In line with accounting and management general, authors in the discipline of organizational behavior either provide a detailed or high-level definition of the PF characteristics as their experimental interventions (e.g. DeShon et al., 2004; Sheldon et al., 2014).

Based on an analysis of current definitions and their applications in prior literature reviews and within the selected articles in each of the three disciplines of accounting, management general and organizational behavior, we synthesize overarching patterns and dimensions in the literature to propose a PF definition. This definition of PF should act as a focal point for future research and has the intent to align current synonyms: *PF is the deliberate presentation of information, with varying characteristics, on the task performance within a specific environment to recipients (individuals or groups), to allow them to assess and improve the effectiveness of specific aspects of their future performance.* This definition summarizes prior research and especially leans on the conceptual framework that Schnieder (2018) investigated in the scope of RPF. In this definition, we include multiple key parts to cover the multitude of dimensions to make PF and subsequent employee behavior manageable.

Since prior articles and reviews have shown that not all PF is equal, a large body of research has implied that particular characteristics of feedback can lead to either performance gains, stagnation, or losses. We therefore define PF as the *deliberate* presentation of information, to

differentiate it from random observations by employees at the workplace and emphasize the managerial responsibility of PF design. By highlighting the importance of *information*, we align PF research with the synonym of performance information and point out the distinct effect of information for PF. This effect can be diluted in research design based on the variety of interacting variables, but through including the term we provide a more thorough understanding to advance PF research. In addition, we point out the specific role that *characteristics*, *task*, *environment*, and *recipients* have in understanding the effects of PF (see the framework by Schnieder (2018) for more exhaustive definitions of these variable categories). We also accentuate the differentiation between *assess* and *improve* to account for in- or intra-person differences in transferring insights based on PF to actual performance. Since PF is a key driver of individual performance (Campbell, 2008), our definition focuses on *future performance* as the ends of all means on PF research.

This definition advances prior attempts and current research endeavors on PF in multiple way. First, this definition acts as an organizing frame for future research on the effects of PF across disciplines. Thereby, this definition is intended to be the starting point in introductory sections to accelerate progress in the fragmented literature, since it enables researchers to identify current findings across disciplines more easily and also enables frameworks on specific parts and types of PF effects to be built consecutively. Second, this definition allows to make specific components of PF comparable across articles by making variable definitions and ways of measuring them transparently. Thereby, specific parts of the causal relationships can be investigated thoroughly. Research on PF oftentimes focuses on understanding fundamental variables, e.g. frequency (Casas-Arce et al., 2017) or reference points (Eyring & Narayanan, 2018), and research on PF can only lead to highly applicable learnings for managers by connecting prior findings. Third, this definition summarizes multiple feedback definitions to keep the stream of research from becoming more fragmented. While further definitions for specific parts of the definition exist in isolation, for instance ‘task properties feedback’ covering

the environment or ‘cognitive feedback’ covering the individual perception (Leung & Trotman, 2005), these definitions are not part of the current research discussion and are referenced very infrequently by recent studies across disciplines. By allowing for the inclusion of these definitions, we allow researchers to grasp a holistic picture without having to dilute their specific research scope in future articles.

2.3.2 Theories

In the process of mapping the intellectual territory of PF, we next address the second RQ: *Which theories are applied to investigate the effects of PF in each of the three disciplines (RQ2)?* This RQ is of relevance as a starting point for studies that consider the variety of theories being investigated within the topic of PF. In the context of management research, a theory aims to explain a relationship between independent and dependent variables to predict the outcomes of interest (Colquitt & Zapata-Phelan, 2007) and “it is difficult to overstate the importance of theory to the scientific endeavor” (p.1281). Within this understanding, a theory can be defined as “a collection of assertions, both verbal and symbolic, that identifies what variables are important and for what reasons, specifies how they are interrelated and why, and identifies the conditions under which they should be related or not related” (Campbell, 1990, p. 65). We apply this definition of a theory in our management related disciplines to explain the effects of PF. In their review, Kluger and DeNisi (1996) already argued that a lack of consolidating theories makes the interpretation of findings from PF challenging, since “research is carried out by isolated pockets of researchers who share either a theoretical or paradigmatic orientation” (p. 254). While Kluger and DeNisi (1996) propose a new comprehensive theory, ‘feedback intervention theory’, to advance the understanding of PF research, we now – 24 years later – synthesize the current theories across the three disciplines of accounting, management general and organizational behavior which are being applied to enable potential consolidations or additional theoretical lenses in future research. Table 4 provides an overview of theories being investigated in the selected articles of this review.

Table 4. Overview of Theories Investigated by Discipline

Theory ^a	Total	Accounting	Management general	Organizational behavior
<i>Control theory</i>	4 (5%)	0 (0%)	1 (6%)	3 (7%)
<i>Decision making theory</i>	4 (5%)	1 (4%)	1 (6%)	2 (4%)
<i>Feedback intervention theory</i>	3 (3%)	0 (0%)	0 (0%)	3 (7%)
<i>Goal theory</i>	4 (5%)	2 (8%)	0 (0%)	2 (4%)
<i>Learning theory</i>	4 (5%)	2 (8%)	0 (0%)	2 (4%)
<i>Self-regulation theory</i>	3 (3%)	0 (0%)	0 (0%)	3 (7%)
<i>Social comparison theory</i>	14 (16%)	7 (28%)	4 (25%)	3 (7%)
<i>Other theories/ Exploratory^b</i>	51 (58%)	13 (52%)	10 (63%)	28 (61%)

Notes: ^a Based on an analysis of the dominant theory applied in each article. ^b Theories that were applied less than three times are summarized in ‘Other theories/Exploratory’.

Resulting from the process of synthesizing the theories being applied to explain the outcomes of PF, seven individual theories have been investigated more than two times while the collection of *other theories/ exploratory* contains the vast majority of the articles selected. The category of *other theories/ exploratory* is the summary for a variety of theories and includes two subcategories. First, it includes nine theories which were investigated less than three times as the main theory in our body of articles, for instance *implicit theories* (e.g. Zingoni & Byron, 2017) or the *dual-process model* (e.g. Anseel et al., 2009). Second, it includes exploratory theory sections, which combine multiple theories and make no predictions about directions of the hypothesized effects due to “the lack of definitive theory” (Loftus & Tanlu, 2018, p. 278). Even though the variety of theories can help to provide insights to not only explain positive results of PF (VandeWalle et al., 2001), the heterogeneity of theories also makes it difficult to subsequently advance the theory application, especially since the results of PF are highly dependent on the interaction effects of the several parts of the definition (Schnieder, 2018). Therefore, oftentimes theories from multiple research disciplines are combined, for example research draws on economic theory to explain the effects of compensation and psychology theory to explain individual responses to feedback (Tafkov, 2013).

Although our systematic review process only includes articles which use quantitative methods to analyze experimental data (inclusion criteria nr. 1), top peer-reviewed management journals

within the scope of our review expect theoretical contributions from empirical articles to advance management theory (Rynes, 2005). Nevertheless, Colquitt and Zapata-Phelan (2007) state that the majority of the most-investigated theories in management research stem from other outlets, such as book chapters or journals dedicated to theory building. While it is not within the scope of this review to investigate which factors might explain that trend and what recommendations are given based on this finding (see Colquitt and Zapata-Phelan (2007) for a thorough discussion), answering our RQ provides transparency for future research on PF for either testing specific theories or building theories to address current gaps.

Comparing similarities and differences regarding the theories tested across the three disciplines of accounting, management general and organizational behavior, the distribution of theories is heterogeneous. Authors in the discipline of accounting draw on *social comparison theory* the most often with no other theory being applied in a double-digit percentage. Social comparison theory is based on the work by Festinger (1954) and argues that individuals have a desire to compare their performance to the performance of similar others to derive inferences about their own abilities. Social comparison theory is oftentimes relied upon by researchers to explain why individuals engage based on PF in the absence of monetary incentives (Wang, 2017), which is especially relevant for management accountants to design incentive systems in combination with PF (Schnieder, 2018). In line with this practical relevance, recent articles in the discipline of accounting have tested social comparison theory to explain the PF effects of different incentive schemes (Tafkov, 2013), multi-task environments (Hannan et al., 2013) or multi-period competitions (Berger et al., 2018). *Goal theory* and *learning theory* are also tested more than once in the discipline of accounting. Goal theory predicts that more difficult performance goals lead to higher levels individual performance due to a higher motivation to achieve the lower probability of reaching this goal (Locke et al., 1986) and is tested in accounting research to explain the effects of PF on individual effort (Thomas, 2016). Theory

on learning is tested to investigate which types of training facilitate learning to improve performance (Moreno et al., 2007).

In line with the discipline of accounting, authors in the discipline of management general investigate social comparison theory the most as an individual theory (25 percent). Researchers in this discipline apply social comparison theory, for instance, to explain the PF effects of publicity (Song et al., 2018) or timing (Tiefenbeck et al., 2018). No other theory is tested more than once in the discipline of management general.

Contrary to the other two disciplines, authors in the discipline of organizational behavior test all seven individual theories quite homogeneously by testing each theory either two or three times. In addition to social comparison theory, three theories are also investigated three times: *control theory*, *feedback intervention theory* and *self-regulation theory*. Social comparison theory is applied in articles within the discipline of organizational behavior to test, for example, the PF effects of absolute or relative ability (Moore & Klein, 2008). Control theory, which argues that PF is perceived as a reference point relative to a performance standard (Hartwell & Campion, 2016), is tested, for instance, to explain the effects of the PF medium (Alder & Ambrose, 2005). Feedback intervention theory, as defined by Kluger and DeNisi (1996), is only applied in the discipline of organizational behavior, for instance, to understand if individuals rely on specific or global aspects when facing PF (Dimotakis et al., 2017). At last, self-regulation theory addresses an individual's ability to execute a desired change based on a habitual loop of behavior (Baumeister & Heatherton, 1996), and is investigated, for example, to test the effects of the PF condition being positive or negative (Ruttan & Nordgren, 2016).

Taken together, our synthesis confirms the currently fragmented landscape of theory being investigated across the three management related disciplines as Kluger and DeNisi (1996) stated more than two decades ago. Recent reviews include specific theories in their conceptual frameworks as “the most prominent mediating mechanisms” (Schnieder, 2018, p. 3), but our findings cannot confirm that these specific theories are in fact the most relied upon by recent

empirical studies. For example, Schnieder (2018) draws on social comparison theory and *impression management theory* in his conceptual framework to organize the effects of RPF on performance. With respect to social comparison theory we can confirm the inclusion due to its prominence, but we only find one brief mentioning of impression management theory within our body of articles (Hannan et al., 2013). By making the current landscape of theories being applied transparent, we hope to assist in the future development of conceptual frameworks for researchers across disciplines.

While a variety of theories is being investigated, we can confirm the observation by Colquitt and Zapata-Phelan (2007) – based on our selection of articles – that theory building and theory testing have not been emphasized equally in management research. Still, there are two findings that make the advancement of more aligned theory to explain the effects of PF promising. First, the number of articles with exploratory theory sections shows that authors are already considering previously untested theories – often in combination – to explain the effects of PF thoroughly. As a next step, PF research should propose theories, based on these exploratory sections and subsequent empirical findings, that add to the theories having been tested before. Even though the space in empirical articles to propose new theories is often too short (Barley, 2006), this theory building can take place in other research outlets, but should trickle down into the empirical body of research to test and iteratively improve it. Second, while social comparison theory is the only individual theory being investigated with an extended body of research across the three disciplines, this finding proves that individual theories can facilitate the empirical research and align disciplines on its results. The majority of theories being used to predict and explain the effects of PF stems from psychology theory to explain responses of recipients to PF, either individuals or groups. Still, future research on PF could incorporate recent theory building from psychology and behavioral economics to explain specific PF mechanisms, since social comparison theory (Festinger, 1954) or goal theory (Locke, 1968) are long-tenured theories.

2.3.3 Main Findings and Empirical Designs

Finally, we focus on answering the two sequential parts of our third RQ to assess the current state of empirical PF research: *What are the main findings on the effects of PF in each of the three disciplines (RQ3a)? Which research designs are operationalized to investigate the effects of PF in each of the three disciplines (RQ3b)?* Having suggested an integrative definition of PF and after reviewing the application of theory in PF research, the third RQ is relevant to build the foundation for future PF research by synthesizing which findings they produce (RQ3a) and which empirical research designs are currently applied (RQ3b) in each of the three management related disciplines. Based on this approach, we make the current academic discourse on PF transparent (Webster & Watson, 2002). Thereby, we differ from previous reviews which have focused on trying to answer the big question of whether and how PF has an influence on performance in general and on several components of performance particularly. Since Kluger and DeNisi (1996) identified a small overall effect of feedback on performance in their meta-analytic review, recent reviews have notably focused on breaking down this general effect into more specific conceptual frameworks for main and interaction effects of feedback characteristics (e.g. Lechermeier & Fassnacht, 2018) or the multiple dimensions of individual behavior that affect the construct of performance for specific types of PF (e.g. Schnieder, 2018). For this reason – and in line with our visual format – we organize our review of the along four elements of the integrative definition for PF: (A) the deliberate presentation of information with varying characteristics, (B) the specific environment, (C) the recipient (individuals or groups) and (D) future performance. We visually present our synthesis to provide evidence for each of the RQs across the three management disciplines.

First, we present a summary of all selected articles within each discipline to provide an overview about the information for each selected article (see Table 5, 6 and 7). These summaries contain the independent variable(s) investigated as part of the PF characteristics elements (A). In addition, we present information on the PF environment (B), namely the

experimental method, the number and type of participants and the compensation scheme applied. We present these three types of information to respond to recent calls across management and economics literature to take into consideration the specific experimental research design. All three environment variables differ on core dimensions which influence the possibility to abstract from the specific research setting, such as experimenter control and awareness of the study by the participants (Gneezy & Imas, 2017). Within these summary tables, we also specify the dependent variable(s) and a short description of the study findings to allow for inferences about future performance effects (D) based on the provision of PF.

Second, we present one table that accumulates the individual information on the PF environment (B) stored from each article (see Table 8). This table synthesizes the information for each of the three disciplines on an aggregated level by providing both absolute values and percentage shares within the discipline. This approach enables us to present a holistic picture of the current state of PF research while also enabling researchers to identify patterns or underrepresented elements to set up future PF studies.

Third, we present one table at an aggregated level by providing both absolute values and percentage shares for each of the three management related disciplines on the distribution of characteristics variables (A) (see Table 9). Concerning the characteristics variables covered in this table, we synthesize six variables that Balcazar et al. (1985) aggregated and Alvero et al. (2001) later updated for four journals in the discipline of organizational behavior: the PF source, the type of privacy, the type of recipient, the type of realism, the type of medium and the type of frequency. Even though we adjusted the respective subcategories for these variables, this approach makes it possible to compare our results for the discipline of organizational behavior to the findings of Alvero et al.'s (2001) review and extend the synthesis to the disciplines of accounting and management general. In addition, we also aggregate the information about the realism of the PF displayed, since PF studies only recently were able to provide realistic

feedback for performance instead of relying on manipulated feedback independent of the actual results of participants (Ilies & Judge, 2005).

Fourth, we provide one table that synthesizes the application of person variables (C) for each of the three disciplines on an aggregated level by providing both absolute values and percentage shares within the discipline (see Table 10). Therefore, we draw on recent reviews by Schnieder (2018) and Lechermeier and Fassnacht (2018) to provide an overview if the articles included self-efficacy, self-consciousness or the recipients' level of skill or ability for the task within the experimental design. While self-efficacy describes individual judgements about the own capacity to perform on a specific task (Gist & Mitchell, 1992), self-consciousness relates to individual judgments about how others perceive someone's own capacity to perform on a specific task (Lechermeier & Fassnacht, 2018). Skill or ability include further (objectifiable) person variables on specific tasks that can have an influence on the effects of PF (S. E. Bonner & Sprinkle, 2002). In addition, we also synthesize how the articles within the discipline included personality traits, cognitive ability/ traits, and performance segments in their research design. To the best of our knowledge, no prior review has aggregated the inclusion of these person variables in PF research. Research on personality traits concluded that personality traits are related to performance in general and that many personality traits especially have a high predictive power for performance in specific job contexts (Judge & Zapata, 2015). In addition, researchers have advocated to include the role of personality in performance research since the 1990s (VandeWalle et al., 2001). Aggregating cognitive abilities and traits in PF research is relevant since cognitive processes and limits can influence information processing (Narayanan et al., 2009). Furthermore, research in cognitive psychology has provided evidence that multiple fallacies and biases systematically affect individual performance (Budescu & Maciejovsky, 2005). We also synthesize the inclusion of performance segments in our review, for instance high- or low levels of initial task performance, based on recent articles which find that

performance segments have an influence on the relationship between PF and subsequent performance (e.g. Eyring & Narayanan, 2018).

In the following two subsections, we will address both RQs (RQ3a and RQ3b) respectively by first presenting major findings and then by discussing the current research designs applied in research on the effects of PF in all three disciplines. We conclude this section with a discussion of the current state of PF by assessing both the main findings and research designs conducted.

2.3.3.1 Main Findings

Research within the scope of our review of PF presents a broad range of findings to advance our knowledge about “understanding who, how, and when PF should be given to improve its effectiveness” (Andiola, 2014, p. 1). Two recent reviews on the topic addressed this guiding question with a specific scope on PF variables since Kluger and DeNisi (1996) have stated more than twenty years ago that findings about the effects of feedback on performance are heterogeneous due to the partial understanding of multiple variables of potential influence. Hence, Lechermeier and Fassnacht (2018) investigate the findings about the specific effects of the PF characteristics source, timing, and valence in several contexts. In addition, Schnieder (2018) reviews the effects of RPF on performance through multiple mediators, for instance effort, human capital, or employee interaction. Overall, Schnieder (2018) finds that RPF positively affects employee performance but that counterfactual findings still exist. To complement these reviews, we briefly synthesize the findings of PF research along the three management related disciplines of accounting, management general and organizational behavior to present the current body of findings in combination with the research designs that produced these findings. Therefore, we structure the discussion of the findings within each discipline based on the elements of our integrative definition of PF. In line with our visual approach, the findings for each selected article are summarized in the respective table for the discipline (see Table 5, 6 and 7).

Overall, research in the discipline of accounting predominantly investigates the effects of PF on performance and finds positive effects (Eyring & Narayanan, 2018) but also no effects (e.g. Lourenço, 2016) or negative effects (e.g. E. W. Chan, 2018). One of the most commonly investigated types of PF in accounting is RPF, which Schnieder (2018) defines as “information that allows individuals (or a group of individuals) to assess their performance in relation to peers (or peer groups)” (p.1). Therefore, many articles within the discipline of accounting strive to advance the understanding about how this type of feedback affects individual behavior depending on the characteristics, the environment, and the recipients. Specific PF characteristics have been shown to influence the relationship between RPF and its effects. Tafkov (2013) finds that the positive effect of RPF is greater when it is available in public and not communicated in private. But not all findings in the discipline of accounting are related to RPF.

Further studies in accounting find that the specific language and valence of the PF display have an influence on performance. Loftus and Tanlu (2018) show that causal language has a larger effect on performance when the feedback is negative while Jermias (2001) finds that negative PF also decreases the reluctance to change. Regarding the timing of PF, researchers in accounting find conflicting evidence. In his study, Thornock (2016) finds an inverted u-shaped relationship between the timing of PF and future performance in a laboratory experiment, meaning that the most effective time to present PF is immediately after a decision or performance. On the contrary, Casas-Arce et al. (2017) find in their field experiment that professionals overweigh the latest feedback report as most salient and that rather infrequent, delayed PF helps to improve performance. Investigations about the amount of information have produced mixed evidence as well. While Hannan et al. (2008) find in their laboratory experiment that the specific content – for instance coarse or fine – of PF has no impact on performance, Casas-Arce et al. (2017) find that detailed feedback with more information helps professionals to improve their performance more. These findings are an example for the

interplay between the multiple elements within our suggested definition, here characteristics and environment variables, and how defining the variables influences the research findings.

Due to the prevalence of agency theory and inherited positive expectations about effects of monetary incentives (Lourenço, 2016), many studies in accountings investigate the effects of incentives on PF effects within the study environment and produce mixed findings. In two laboratory experiments, Hannan et al. (2008) and Tafkov (2013) find that individual performance based on the provision of RPF is higher under an individual-based incentive scheme. As a result of a field experiment, Lourenço (2016) finds no change in performance when monetary incentives are introduced. The type of task also produces interaction effects. Christ et al. (2012) find that individual performance on dimensions of a task is dependent on the formal control and Hannan et al. (2013) find that RPF leads to decreases in performance when individuals can allocate effort in a multi-task environment within a laboratory experiment.

Especially performance segments have been considered as person variables in accounting research and lead to the conclusive finding, that they affect individual behavior based on PF. Eyring and Narayanan (2018) find in a field experiment that tailoring the reference point for peer-performance comparisons leads to larger performance improvements than providing a uniformly chosen one. When high performers are provided with PF about their current performance, Berger et al. (2018) find that these winners can sustain their efforts over multiple rounds. Regarding personality traits, Wang (2017) finds that RPF only results in performance effects for individuals that score high on three distinct personality traits of the Dark Triad (Machiavellianism, narcissism, and psychopathy).

Next, our synthesis of research on PF in the discipline of management general shows that this discipline also has its focus on the effects on performance but is more heterogenous than research in accounting, as evident through a larger variety of dependent variables, for instance including individual decisions and not primarily behavior (e.g. Engelbrecht-Wiggans & Katok, 2008). Still, the findings in the discipline of management general mostly present positive direct

effects (e.g. Shea & Howell, 2000; Song et al., 2018) or interaction effects (e.g. Adomavicius et al., 2012; Hoever et al., 2018) of PF. The types of PF characteristics that researchers in the discipline of management general have looked at are oftentimes comparable to the characteristics investigated in the discipline of accounting, also with a distinct emphasis of research on RPF. Blanes i Vidal and Nossol (2011) find that providing RPF in general increases the performance quantity without influencing the quality. Investigating the effects of reference points in RPF displays, Gill et al. (2019) conduct a laboratory experiment and find a u-shaped rank response function where individuals improve their performance the most when they are ranked first or last. In addition, Charness et al. (2014) find that RPF leads to greater average effort but also introduces unethical behavior by sabotaging others' ranks or inflating the own results. Also taking person variables into account, Kuhnen and Tymula (2012) find that recipients with overstated levels of self-efficacy decrease their performance.

Regarding timing variables as feedback characteristics, both Shunko et al. (2018) and Tiefenbeck et al. (2018) find effects of real-time feedback and salience. Shunko et al. (2018) conduct a laboratory experiment and find that performance under individual-performance based compensation is dependent on feedback availability. In a large-scale field experiment, Tiefenbeck et al. (2018) show that providing real-time feedback can be an effective way to overcome multiple biases and lead to a large impact on individual performance in the form of reduced resource consumption. In addition to these variables, Song et al. (2018) find – in line with laboratory experiments in accounting – that providing PF publicly in a natural experiment increases performance compared to communicating it privately.

Studies in management general also present findings that combine different PF characteristics, recipients on a group level and a diverse set of dependent variables in one research designs. Hoever et al. (2018) find that positive effects of feedback valence on creativity at a team level are contingent on the diversity of task-relevant knowledge within the team. Also, with regards to recipients on a group level, Jung et al. (2010) find that groups perform the best on idea

quantity and quality when they are provided with PF and pseudonyms to identify group members.

In the discipline of organizational behavior, 21 (of 46) selected articles investigate the effects of feedback on a dependent variable that includes the term “performance” with heterogeneous results. Research on PF effects within organizational behavior is diverse because articles have differentiated between investigating perceptions of feedback or behavioral outcomes until recently (Zingoni & Byron, 2017). Further studies present findings on multiple other effects, for instance investment decisions (Schultze, Pfeiffer, et al., 2012), confidence (S. S. K. Lam et al., 2004) or on goals (Ilies & Judge, 2005). Compared to research in the disciplines of accounting and management general, fewer studies provide evidence on the effects of RPF. Bonner et al. (2002) find in a laboratory experiment that the provision of ranking information has no positive effect on group performance. In addition, Moore and Klein (2008) are the only researchers within the selection of articles that investigate the effects of absolute versus RPF on bet placement in a laboratory setting. The findings of their two experiments state that absolute feedback leads to stronger and more consistent effects on recipients’ satisfaction and self-esteem. Building on the research by Moore and Klein (2008), Zingoni and Byron (2017) conduct further laboratory experiments and find that subsequent performance is contingent on individuals’ implicit theories of ability. Their findings conclude that if a recipient believes that individuals cannot change their abilities, receiving negative relative feedback is associated with worse performance. On the contrary, perceiving negative absolute feedback leads to worse performance outcomes for individuals who believe that abilities can be changed.

Furthermore, researchers in organizational behavior present findings on additional PF characteristics. Focusing on the medium, Gibson (2000) finds that providing a graph for performance development over time has a positive effect on performance. Atkins et al. (2002) find that different mediums have diverging effects, since presenting PF in a graphical format leads to an increase in performance but a tabular presentation format fertilizes learning. Other

studies provide evidence on the effects of valence. Kim et al. (2003) show that displaying negative feedback about ability results in the worst performance outcomes for individuals in negotiations. Within the same context, Dimotakis et al. (2017) find that positive feedback enhances the recipients' self-efficacy for improvement and S. S. K. Lam et al. (2004) further demonstrate that self-consistency effects in evaluating PF valence apply for groups as they do for individuals. Vancouver and Tischner (2004) deliver evidence that the task for which feedback is provided within the laboratory experiment has an influence on the performance as well, in a way that positive feedback is associated with performance improvements for highly cognitively challenging tasks.

Studies in organizational behavior also show that the feedback characteristics of timing and specificity influence performance. Based on the results of a computer-based class-scheduling tasks, Northcraft et al. (2011) find that specific, immediate feedback leads to the largest performance improvements. On the contrary, Lurie and Swaminathan (2009) find in multiple laboratory experiments that more frequent feedback may lead to worse performance due to the inability of participants to adequately compare information over time. With regards to the time horizon of PF, Hartwell and Campion (2016) find in a field setting that cumulative PF with a mean rating over multiple rounds has a larger effect on behavioral changes than the most recent feedback report.

Table 5. Summary of Accounting Articles

Article	Independent variable(s)	Dependent variable(s)	Method. ^a	Partici. ^b	Comp. ^c	Finding
<i>Berger et al. (2018)</i>	Tournament horizon/ Percentage of winners	Overall performance/ Social comparison	a	a (400)	a	Grand tournaments (w. cumulative PF) indirectly result in higher individual performance than repeated tournaments (w. current PF) by inducing higher levels of social comparisons.
<i>J. L. Brown et al. (2016)</i>	Task difficulty/ Feedback	Performance pay	a	a (148)	c	Higher-skilled individuals choose relative-performance-based payments appropriately more often when provided with RPF.
<i>Casas-Arce et al. (2017)</i>	Frequency/ Level of detail	Customer satisfaction/ Operational performance	b	b (800)	c	Individuals have a higher performance when they receive detailed but infrequent PF, suggesting professionals overweigh the latest PF they receive.
<i>Chan (2018)</i>	RPF	Performance	a	a (220)	c	Low and High performer respond different to a promotion only when they are provided with RPF.
<i>Christ et al. (2012)</i>	Formal control/ Timeliness of PF	Accuracy/ Speed	a	a (131)	e	Immediate PF increases performance for the controlled dimension (accuracy) but not for the compensated dimension (speed).
<i>Christ et al. (2016)</i>	Compensation controls/ Feedback control	Accuracy/ Speed	a	a (125/ 87)	e	PF is associated with improving performance on task dimensions that are not incentivized by compensation to increase overall performance.
<i>DeZoort et al. (2006)</i>	Accountability pressure	Effort	a	b (167)	i	Higher levels of accountability through PF lead to conservative judgements taking more time with less variability.
<i>Earley (2001)</i>	Explanatory feedback/ Elicited self-explanations	Performance	b	b (150)	i	Providing explanatory feedback improves novices' performance when feedback is given in a summary form (for rounds).
<i>Eyring & Narayanan (2018)</i>	RPF	Activity level/ Grade	b	d (15.171)	i	Providing the more effective reference point, depending on initial performance, has a greater performance effect than providing a reference point uniformly.
<i>Fisher et al. (2002)</i>	Information asymmetry/ Resource allocation/ Periods	Performance	a	a (174)	d	The degree of information asymmetry in PF between subordinates does not significantly affect subordinates' performance.
<i>Hannan et al. (2008)</i>	RPF/ Incentive scheme	Performance/ Change in Performance	a	a (134)	b	Providing RPF (independent of the level of detail) increases performance under an individual incentive scheme but decreases performance under a tournament incentive scheme.
<i>Hannan et al. (2013)</i>	Effort allocation/ RPF	Problems solved/ Time Difference	a	a (90)	d	RPF results in distorted effort and (public PF) reduces performances in a multi-task environment where individuals determine effort allocations.

Table 5. Summary of Accounting Articles (Continued)

Article	Independent variable(s)	Dependent variable(s)	Method. ^a	Partici. ^b	Comp. ^c	Finding
<i>Hodge et al. (2010)</i>	Proximity/ Feedback	Forecast error/ Dispersion	a	a (39)	e	Receiving extensive PF on a single page leads to greater learning in early rounds and more consistent accuracy in later rounds.
<i>Jermias (2001)</i>	Commitment/ Feedback	Inertia	a	a (82)	e	People are more reluctant to change and have high inertia when they receive positive PF instead of negative PF.
<i>Kelly (2007)</i>	Firm type/ Performance Evaluation System	Decision quality	d	a (122)	c	Providing both PF and incentives on nonfinancial measures improves decision quality more in intangible assets firms than in tangible assets firms.
<i>Leung & Trotman (2005)</i>	Feedback/ Configural demands	Judgement accuracy	a	b (224)	g	While task properties feedback increases performance independent of tasks, the effects of outcome feedback or cognitive feedback are dependent on the level of configural cue processing required.
<i>Loftus & Tanlu (2018)</i>	Causal language/ Valence	Change in scores	a	a (108)	a	Causal language in PF improves performance when feedback is negative and with a lower improvement when PF is positive.
<i>Lourenço (2016)</i>	PF/ Incentives/ Recognition	Average sales/ Goals per rep-week	b	b (352)	e	Providing PF leads to no change in performance and is independent of the provision of monetary incentives or recognition.
<i>Moreno et al. (2007)</i>	Outcome feedback	Accuracy	a	a (169)	i	Combining multiple training techniques (e.g. worked-out examples and self-explanation) improves decision making more compared to providing outcome feedback alone.
<i>Nelson et al. (2003)</i>	Feedback	Net Security Position	a	a (54)	c	Individuals rely more on disciplined strategies when they received PF that their unaided performance was unprofitable.
<i>Newman & Tafkov (2014)</i>	Tournament prize structure/ RPF	Performance	a	a (80)	b	RPF has a positive effect on performance in a reward and punish tournament, but a negative effect in a reward tournament.
<i>Tafkov (2013)</i>	RPF/ Type of RPF/ Incentive scheme	Performance	a	a (120)	c	RPF positively affects performance and has a greater positive effect under an individual performance-based contract compared to a flat-wage contract. Providing public RPF has a greater positive effect.
<i>Thomas (2016)</i>	Priming/ Feedback	Future effort	a	a (118)	g	PF of success leads to more effort in a high-level goal condition and PF of failure leads to more effort in the absence of a high-level goal.

Table 5. Summary of Accounting Articles (Continued)

Article	Independent variable(s)	Dependent variable(s)	Method. ^a	Partici. ^b	Comp. ^c	Finding
<i>Thornock (2016)</i>	Delay of decision quality PF	Performance	a	a (90)	e	The relation between timing of PF and future performance has an inverted U-shape where no-delay and long delay adversely affect performance.
<i>Wang (2017)</i>	Recognition	Correct entries	a	a (76)	c	Public and private PF increases performance for High Dark Triad individuals, but not for low Dark Triad individuals.

Notes: ^a Methodology (Method.): a = Laboratory experiment, b = Field Experiment, c = Natural Experiment, d = Online Experiment. ^b Participants (Partici.): a = Students, b = Professionals, c = Mix, d = N/A. ^c Compensation Scheme (Comp.): a = Tournament & individual performance-based compensation, b = Tournament & effort based compensation, c = Flat wage & individual performance based compensation, d = Flat wage & effort based compensation, e = Individual performance-based compensation, f = Effort based compensation, g = Flat wage, h = Other, i = N/A.

Table 6. Summary of Management General Articles

Article	Independent variable(s)	Dependent variable(s)	Method. ^a	Partici. ^b	Comp. ^c	Finding
<i>Adomavicius et al. (2012)</i>	Information feedback	Individual surplus	a	a (249)	c	PF has no impact on an individual's surplus as a direct effect, only in interaction with competition types.
<i>Blanes i Vidal & Nossol (2011)</i>	Average wage per hour/ Rank-order position	Number of goods dispatched p. hour	b	b (63)	a	Rank-order information increases productivity across performance segments in the long-run with no negative effect on quality or retention.
<i>Buser & Dreber (2016)</i>	PF	Individual contribution	d	d (1700)	a	Receiving PF makes individuals contribute less in a public goods game.
<i>Charness et al. (2014)</i>	PF	Initial performance	a	a (585)	g	RPF has a positive effect on individual performance through an increased level of effort, but also invites unethical behavior.
<i>Engelbrecht-Wiggans & Katok (2008)</i>	Winner's regret PF/ Loser's regret PF	Bid/ Value	a	a (80)	a	PF makes individuals adapt their bids in auctions based on regret; winners that overpay reduce their bids and losers that missed out increase their bids.
<i>Gill et al. (2019)</i>	Rank-order PF	Effort	a	a (306)	g	RPF increases performance via a U-shaped rank-response function independent of its distribution channel and publicity.
<i>Hoever et al. (2018)</i>	Informational diversity/ Feedback valence	Elaboration/ Team creativity/ Generative processing	a	a (234)	c	PF valence does not affect team creativity, only in interaction with informational diversity.
<i>Jung et al. (2010)</i>	Pseudonymity/ PF	Idea quantity/ Idea quality	a	a (260)	h	Groups provided with PF and pseudonyms outperform all other groups.
<i>Karagözoğlu & Riedl (2015)</i>	Performance outcome/ RPF	Performance-specific entitlement	a	a (348)	c	Performance-specific entitlements are established by providing PF.
<i>Kuhnen & Tymula (2012)</i>	PF	Output	a	a (54)	g	Individuals with higher levels of self-efficacy improve their performance the most when they are probable to receive RPF.
<i>Maciejovsky et al. (2013)</i>	Decision unit/ PF	Correct choices	a	a (240)	c	Team decision making is advantageous compared to individual decision making and not dependent on the type of PF given.
<i>Shea & Howell (2000)</i>	Task feedback	Task performance quality	a	a (148)	i	Individuals receiving task feedback outperform those that do not. External task feedback is more likely to align self-efficacy and performance perception.
<i>Shunko et al. (2018)</i>	Queue structure/ Queue visibility	Performance	a/d	a (248) d (481)	h	Less salient feedback makes individuals slowdown in their performance, but this effect might be mitigated or reversed by pay scheme.

Table 6. Summary of Management General Articles (Continued)

Article	Independent variable(s)	Dependent variable(s)	Method. ^a	Partici. ^b	Comp. ^c	Finding
<i>Song et al. (2018)</i>	PF Disclosure	Productivity	c	b (172)	g	Introducing public RPF results in a productivity increase that persists for 12 months attributed to sharing best practices.
<i>Stajkovic & Luthans (2001)</i>	PF	Performance improvement	b	b (182)	h	Providing PF improves performance, but less than either monetary incentives or social recognition.
<i>Tiefenbeck et al. (2018)</i>	PF	Energy used	b	c (636)	i	Providing real-time PF on a specific behavior can induce large behavioral changes and overcome salience bias.

Notes: ^a Methodology: a = Laboratory experiment, b = Field Experiment, c = Natural Experiment, d = Online Experiment. ^b Participants: a = Students, b = Professionals, c = Mix, d = N/A. ^c Compensation Scheme: a = Tournament & individual performance-based compensation, b = Tournament & effort based compensation, c = Flat wage & individual performance based compensation, d = Flat wage & effort based compensation, e = Individual performance-based compensation, f = Effort based compensation, g = Flat wage, h = Other, i = N/A.

Table 7. Summary of Organizational Behavior Articles

Article	Independent variable(s)	Dependent variable(s)	Method. ^a	Partici. ^b	Comp. ^c	Finding
<i>Alder & Ambrose (2005)</i>	Control over PF/ Constructiveness/ PF Medium	Task satisfaction/ Task performance	a	a (165)	c	Constructive PF delivered by supervisors instead of computers increases fairness judgements, which is associated with higher performance quantity and quality.
<i>Anseel et al. (2009)</i>	PF/ Reflection/ Task	Performance	b, a	b (640) a (488)	i	PF combined with reflection improves task performance more than providing only PF.
<i>Atkins et al. (2002)</i>	Levels of PF/ System delay/ System PF gain	Performance	a	a (18)	g	PF in a graphical format leads to better performance improvements but less learning effects than PF in a tabular format.
<i>Bachrach et al. (2001)</i>	Feedback	Citizenship behavior	a	a (412)	a	Citizenship behavior in groups might be a function of the nature of PF with negative PF being more important than positive PF.
<i>Barkan (2002)</i>	Feedback/ Incentive condition	Learning	a	a (60)	e	PF as outcome feedback or cognitive feedback decreases risk-taking and improves learning.
<i>Bartol et al. (2001)</i>	Performance rating segmentation/ Performance rating levels	Performance improvement	a	a (305)	h	Individuals receiving PF at the low and middle end set lower goals and have lower self-efficacy but do not differ regarding performance improvements.
<i>Bonner et al. (2002)</i>	Difficulty level/ Ranking information	Performance	a	a (360)	g	Providing Ranking PF does not affect group performance.
<i>Bragger et al. (2003)</i>	Uncertainty	Number of times invested/ Total amount invested	a	a (136)	c	Uncertain PF makes individuals invest more resources but not more investments in total.
<i>Chen & Matieu (2008)</i>	Goal frame/ Feedback referent	Performance	a	a (104)	g	PF based on a group score versus individual performance history has no effect on performance trajectories.
<i>Chun et al. (2018)</i>	Type of comparisons/ Evaluation valence	Procedural and interpersonal fairness	d	d (400)	i	PF with temporal comparisons leads to higher perceptions of fairness than PF with social comparison since participants feel it is individualized.
<i>DeShon et al. (2004)</i>	Feedback	Individual performance/ Team-oriented performance	a	a (237)	h	Individuals improve either on individual- or team-performance depending on which form of PF they receive.
<i>Dimotakis et al. (2017)</i>	Feedback received	Self-efficacy/ Career outcomes	b	b (126)	i	Positive feedback has a positive relationship with self-efficacy, but not directly with career outcomes.
<i>Fernbach et al. (2014)</i>	Explicit PF	Self-deception/ Effort denial	a	a (38)	i	High or low endurance benefits have only marginal effects on behavior when relative, negative PF is provided.

Table 7. Summary of Organizational Behavior Articles (Continued)

Article	Independent variable(s)	Dependent variable(s)	Method. ^a	Partici. ^b	Comp. ^c	Finding
<i>Fischer & Budescu (2005)</i>	Feedback types	Performance/ Confidence	a	a (90)	c	While complete and efficient PF results in performance and confidence improvements, reduced PF has the same effects after large amounts of repetition.
<i>Gibson (2000)</i>	Feedback delay/ Graphs	Performance	a	a (90)	g	PF with a graphical representation of past results improves performance.
<i>Gonzales (2005)</i>	Control/ Feedback/ Self-exemplar/ Feedback-exemplar/ Expert-exemplar	Performance	a	a (88)	g	Watching videos from expert performances increased individual performance while other types of PF make no difference.
<i>Goodman & Wood (2004)</i>	Feedback specificity	Learning (performance)	a	a (183)	g	Higher feedback specificity allows individuals to learn from good performance but is detrimental to the learning from poor performance.
<i>Goodman et al. (2011)</i>	Feedback specificity	Information processing	a	a (48)	g	PF with low specificity is negatively related to explicit information processing.
<i>Goodman et al. (2004)</i>	Feedback specificity	Practice performance/ Learning	a	a (149)	g	Feedback specificity increases individual performance but does not have a direct impact on learning.
<i>Hartwell & Campion (2016)</i>	Normative feedback	Interviewer rating	b	b (118)	i	Normative feedback increases within-interviewer rating variance and levels extreme individuals. Cumulative feedback is more important than the most recent report.
<i>Ilies & Judge (2005)</i>	PF	Goals	a	a (745/162)	g	PF predicts goal regulation and this relationship is affected by affective reactions.
<i>Johnson et al. (2013)</i>	Diagnostic List/ Structural alignment feedback	Structural change/ Personnel change/ Process change	a	a (312)	c	Providing teams with PF on their structure allows them to make more self-regulating changes.
<i>Kim et al. (2003)</i>	Valence/ Scope	Competitiveness/ Performance	a	a (160)	g	Negative PF about ability leads to the worst individual performance in negotiations.
<i>C. F. Lam et al. (2011)</i>	Feedback frequency	Task performance	a	a (86)	c	PF frequency has a curvilinear, inverted-U shaped relationship with performance mediated by effort.
<i>S. S. K. Lam et al. (2004)</i>	Valence	Confidence in improved performance	b	b (144)	i	After positive (negative) PF, ability (effort) is positively related to confidence in future performance for groups.
<i>Lejarraga & Gonzalez (2011)</i>	Type of information/ Complexity/ Orders	Choice behavior	a	a (152)	e	When PF is available, written descriptions of possible outcomes and probabilities are ignored in decision making.
<i>LePine et al. (2002)</i>	Mean bias feedback	Team performance	a	a (320)	e	Providing PF on the mean bias in groups reduces biases in subsequent decisions, especially for predominantly male teams

Table 7. Summary of Organizational Behavior Articles (Continued)

Article	Independent variable(s)	Dependent variable(s)	Method. ^a	Partici. ^b	Comp. ^c	Finding
<i>Lurie et al. (2009)</i>	Feedback frequency/ Profit environment	Performance/ Information selection	a	a (27-253)	g	More frequent PF leads to performance declines through more compensatory decision processes.
<i>Mak et al. (2015)</i>	Feedback/ Market information	Mean payment/ Mean no. of rounds	a	a (160)	c	Communication via chat alone can facilitate coordination to level the influence through the availability of feedback.
<i>Moore & Klein (2008)</i>	Absolute Feedback/ Comparative Feedback	Bets/ Affective Response/ Confidence/ Self-Esteem	a	a (415/ 182)	g	Individuals use either absolute or comparative PF to make betting decisions, but APF had a stronger influence on bets and psychological traits.
<i>Moreland & Myaskovsky (2000)</i>	Training conditions	Procedural recall/ Assembly errors	a	a (186)	c	Groups with PF and group training performed better in procedural recall and performance than individually training groups but did not differentiate between each other.
<i>Northcraft et al. (2011)</i>	Feedback timing/ Feedback specificity	Performance	a	a (55)	g	Timely and positive feedback is linked to positive performance across competing tasks.
<i>Oc et al. (2015)</i>	Feedback condition	Resource allocation	a	c (80/ 105)	c	Candid feedback by subordinates increases powerholders self-regulation in resource allocations.
<i>Pillutla & Chen (1999)</i>	Feedback/ Context	Investment	a	a (195)	c	PF affects behavior only when its unexpected and negative, there is no main effect for competitive vs cooperative PF.
<i>Rolison et al. (2012)</i>	Cognitive feedback/ Cue polarity	Judgement (performance)	a	a (78)	g	Individuals with higher working memory capacity were quicker to learn the relevant cues and maintain a learning advantage when provided with PF.
<i>Rose & Windschitl (2008)</i>	Type of feedback/ Difficulty/ Round	Likelihood judgements	a	a (56/ 44)	g	Comparative PF is not utilized to adapt likelihood estimates or knowledge judgements.
<i>Ruttan & Nordgren (2016)</i>	Feedback/ Self-control depletion	Validity ratings/ Correct answers	a	c (154/ 180)	i	Positive PF increases validity ratings of a test since depletion in a negative PF conditions is more likely to derogate the rating.
<i>Schultze, Mojzisch et al. (2012)</i>	Expertise feedback/ Trial	Learning/ Performance	a	a (228)	c	PF increases group performance and allows individuals to transfer learn in a group setting.
<i>Schultze, Pfeiffer et al. (2012)</i>	Responsibility/ Feedback	Amount of reinvestment	a	a (269/ 204)	g	Investment decisions based on feedback are not biased by selective exposure to information.
<i>Seifert et al. (2003)</i>	Feedback report	Influence behavior	b	b (59)	i	Feedback workshops with facilitators increase the perceived utility and results in managerial behavior change compared to only giving reports.

Table 7. Summary of Organizational Behavior Articles (Continued)

Article	Independent variable(s)	Dependent variable(s)	Method. ^a	Partici. ^b	Comp. ^c	Finding
<i>Seo & Ilies (2009)</i>	Past performance	Self-efficacy/ Affect	d	d (118)	e	The relationship between PF and performance is mediated by positive affect but disappears when self-efficacy is introduced.
<i>Sheldon et al. (2014)</i>	Feedback	Test accuracy/ Relevance	a	a (66-157)	i	Low performers are more likely to question test accuracy even when being provided with explicit PF.
<i>Stone and Opel (2000)</i>	Type of training/ Stimulus difficulty	Change in scores/ Judgements	a	a (84)	c	PF reduces overconfidence but leads to no other component of improvement.
<i>Vancouver & Tischner (2004)</i>	Task resource sensitivity/ Feedback sign/ Self-affirmation	Performance	a	a (160)	g	Positive feedback leads to performance improvements for high cognitively challenging tasks and negative feedback for low cognitively challenging tasks.
<i>VandeWalle et al. (2001)</i>	Feedback/ Goal level	Performance	a	a (95)	i	Performance improvement based on feedback is mediated by effort, self-efficacy, and goal-level.
<i>Zingoni & Byron (2017)</i>	Feedback valence/ Feedback standard/ Implicit theory	Effort/ Learning	a	a (188/ 158)	g	Individuals' theory of ability interacts with whether relative or absolute feedback is used for performance comparison. Implicit beliefs influence which type of feedback is perceived as valuable or threatening.

Notes: ^a Methodology: a = Laboratory experiment, b = Field Experiment, c = Natural Experiment, d = Online Experiment. ^b Participants: a = Students, b = Professionals, c = Mix, d = N/A. ^c Compensation Scheme: a = Tournament & individual performance-based compensation, b = Tournament & effort based compensation, c = Flat wage & individual performance based compensation, d = Flat wage & effort based compensation, e = Individual performance-based compensation, f = Effort based compensation, g = Flat wage, h = Other, i = N/A.

2.3.3.2 Research Designs

Environment Variables

With regards to the research designs the main findings are based on, we first turn towards the environment variables (B in the integrative definition) of PF by presenting the overall distribution across discipline before discussing the differences by discipline (see Table 8 for details). Overall, the selected articles within our review mostly conduct laboratory experiments, then field experiments as well as online experiments and only one natural experiment. In line with the choice of research methodology, most studies use student samples to recruit their participants, then professionals, followed by mix of participants and studies without specifying the choice of sample. Student samples are often categorized as standard samples due to the ease of access for researchers, but relying on them can potentially bias estimates of causal effects since the subjects are aware of their status of participants in an experiment and their potentially heterogeneity (Floyd & List, 2016). A variety of compensation schemes have been applied within the research designs with two compensation schemes representing the most common applications: Flat wage compensation schemes and flat wage plus individual performance-based compensation. For flat wage compensation schemes, participants usually are compensated by a standardized sum of money or by course credits for student samples. Interestingly, 16 studies do not specify the compensation scheme applied within the research design.

In total, the combination of a laboratory experiment with student participants represents the vast majority of empirical studies in our sample (74 percent). Most often, this combination is investigated with a flat wage (23 percent) or flat wage plus individual performance-based compensation (23 percent). Most of the other combination of environment variables are only applied infrequently with one or two applications (one or two percent).

Table 8. Distribution of Environment Variables

Environment variables^a				
Methodology	Total	Accounting	Management general	Organizational behavior
<i>Laboratory experiment</i>	70 (80%)	20 (80%)	11 (69%)	39 (85%)
<i>Field experiment</i>	12 (14%)	4 (16%)	3 (19%)	5 (11%)
<i>Natural experiment</i>	1 (1%)	0 (0%)	1 (6%)	0 (0%)
<i>Online experiment</i>	4 (5%)	1 (4%)	1 (6%)	2 (4%)
Type of participants	Total	Accounting	Management general	Organizational behavior
<i>Students</i>	64 (74%)	19 (76%)	10 (63%)	35 (76%)
<i>Professionals</i>	12 (14%)	5 (20%)	3 (19%)	4 (9%)
<i>Mix</i>	6 (7%)	0 (0%)	2 (13%)	4 (9%)
<i>N/A</i>	5 (6%)	1 (4%)	1 (6%)	3 (7%)
Compensation scheme	Total	Accounting	Management general	Organizational behavior
<i>Tournament & individual performance-based</i>	5 (6%)	1 (4%)	3 (19%)	1 (2%)
<i>Tournament & effort-based</i>	2 (2%)	2 (8%)	0 (0%)	0 (0%)
<i>Flat wage & individual performance-based</i>	22 (25%)	7 (28%)	4 (25%)	11 (24%)
<i>Flat wage & effort-based</i>	2 (2%)	2 (8%)	0 (0%)	0 (0%)
<i>Individual performance-based</i>	10 (11%)	6 (24%)	0 (0%)	4 (9%)
<i>Effort-based</i>	0 (0%)	0 (0%)	0 (0%)	0 (0%)
<i>Flat wage</i>	25 (29%)	3 (12%)	4 (25%)	18 (39%)
<i>Other</i>	5 (6%)	0 (0%)	3 (19%)	2 (4%)
<i>N/A</i>	16 (18%)	4 (16%)	2 (13%)	10 (22%)

Notes: ^a Based on an analysis of the research design in each article.

In the discipline of accounting, the application of research methods and the choice of participants is similar to the overall distribution across disciplines. The vast majority of empirical studies in accounting conducts laboratory experiments (e.g. E. W. Chan, 2018; Hannan et al., 2008; Tafkov, 2013). In addition to these studies, four articles test their PF hypotheses with field experiments in service firms or online education, especially in recent years (e.g. Casas-Arce et al., 2017; Eyring & Narayanan, 2018; Lourenço, 2016). Only one study of all accounting articles analyzes data from an online experiment (Kelly, 2007). This

distribution of research methods for PF research is in line with recent discussion within the discipline of accounting, in which researchers have increasingly called to use field experiments to explore the underlying mechanisms (Floyd & List, 2016) and move on from the plethora of laboratory experiments (Bloomfield et al., 2016). While Bloomfield et al. (2016) state that only one out of 277 articles in managerial accounting has analyzed field experimental data up until 2016, the recent increase in the topic of PF is an encouraging sign that researchers in accounting are expanding the research designs they investigate.

In line with the selection of research methods, most empirical studies in accounting rely on student samples to recruit their participants (e.g. Loftus & Tanlu, 2018; Newman & Tafkov, 2014; Wang, 2017). In empirical studies in accounting, five articles – mostly field experiments – report professionals as their subjects (e.g. Casas-Arce et al., 2017; Lourenço, 2016) and only one study has no further information on the participants' occupation in their field experiment (Eyring & Narayanan, 2018). Even though researchers in accounting have argued that relying on students to perform on tasks in laboratory experiments is similar to professionals at the beginning of a career with no prior experience (Moreno et al., 2007), the discipline trends towards exploring the effects of PF on experienced professionals. Hence, the number of articles with professional participants and/or field experiments is slightly higher than in the overall distribution.

Regarding compensation schemes, the proximity of accounting researchers to agency theory (Bol, 2011) leads to a different distribution of compensation schemes in experimental studies. According to agency theory, individuals are motivated to invest greater efforts in performance improvements when performance is linked to monetary payments (Holmstrom, 1979). Therefore, individual performance-based compensation without a flat wage component is more common in accounting research on PF than in the other two disciplines. Studies oftentimes rely on only individual performance-based compensation (e.g. Christ et al., 2016; Thornock, 2016) or

manipulate the type of compensation to investigate the interaction effects with PF types (e.g. Hannan et al., 2008; Tafkov, 2013).

The distribution of environment variables (B) in selected articles published in the discipline of management general is slightly different compared to the overall distribution across disciplines. The majority of empirical studies on the effects of PF are laboratory experiments (e.g. Charness et al., 2014; Gill et al., 2019), three studies are field experiments (e.g. Blanes i Vidal & Nossol, 2011) and one study is an online experiment (Buser & Dreber, 2016). Contrary to the other two disciplines, the only natural experiment within our selection of articles stems from the discipline of management general in a hospital setting (Song et al., 2018). Empirical studies in management general also rely less often on student samples compared to the overall distribution, but instead have a higher share of professionals as participants (e.g. Song et al., 2018; Stajkovic & Luthans, 2001) or a mix of participant backgrounds within their experiment (e.g. Shunko et al., 2018). Because researchers from economics departments (e.g. Buser & Dreber, 2016; Charness et al., 2014) publish their articles on the effects of PF in the discipline of management general, tournament-style compensation schemes are more prevalent than in the other two disciplines. Under tournament-based compensation schemes, individuals or groups compete for their final rank at the end of a competition which determines the share of a limited set of rewards that they get rewarded with (Hazels & Sasse, 2008). In contrast to the discipline of accounting, no experimental study in management general solely relies on individual performance-based compensation. The majority of research designs implements either a flat wage compensation scheme or a flat wage plus individual performance-based compensation scheme.

The discipline of organizational behavior presents a higher share of more traditional empirical studies compared to the overall distribution. In organizational behavior, a large majority are laboratory experiments (e.g. Moore & Klein, 2008; Zingoni & Byron, 2017), followed by only five field experiments (e.g. Anseel et al., 2009; Dimotakis et al., 2017) and two online

experiments (e.g. Chun et al., 2018). Consequently, 35 studies use a student sample (e.g. Goodman et al., 2004; Ilies & Judge, 2005) and only four articles – the lowest number out of all three management related disciplines – recruit professionals as participants (e.g. Dimotakis et al., 2017; Hartwell & Campion, 2016). Researchers conduct laboratory experiments to study the impact of specific PF variables in isolation (Bartol et al., 2001), but the experiments already try to mimic field settings and natural environments by introducing noise to the PF display (e.g. Rolison et al., 2012), adopting criteria for the experimental task (e.g. Schultze, Mojzisch, et al., 2012) or recruit older students with potentially first working experience (e.g. Zingoni & Byron, 2017). Articles on PF in organizational behavior also more often include multiple experiments, in which multiple rounds of experimental studies with slight adaptations are conducted, for instance to use a larger sample in an additional round of the experiment to increase the statistical test power (e.g. Schultze, Mojzisch, et al., 2012). These experimental iterations are easier to conduct in laboratory experiments with student participants, because the availability of these subjects is much higher and the coordination process with field sites is not necessary. Still, recent scholars in the discipline of organizational behavior have proposed to use field experiments to advance organizational theory based on causality in feasible ways (Hauser et al., 2017). From the current distribution of PF research designs, we cannot observe a trend towards more experimental designs taking place in the organizational context. It follows that flat wage compensation schemes are more prevalent in experiments on PF in organizational behavior than in the average distribution. Empirical studies in this discipline also report the compensation scheme less often than the other two disciplines, with ten studies not providing details.

Characteristics Variables

Being a key element of our integrative definition, we also analyzed the application of characteristics variables in research designs on the effects of PF. While Table 9 provides a detailed analysis of the aggregated level of characteristics variables per discipline, we will

discuss the overall results and highlight similarities and differences between the three disciplines.

In line with recent advances in technology that make the display of individual PF more efficient to implement, most experimental studies use computer screens as the source of PF for recipients (e.g. E. W. Chan, 2018; Eyring & Narayanan, 2018). The application of computer screens as the dominant source is consistent through all three disciplines, but for other sources the state of PF research is more heterogeneous. While the disciplines of accounting and management general have a larger share of articles using firm representatives or supervisors as the PF source (e.g. Dimotakis et al., 2017; Song et al., 2018), only two studies in the discipline of organizational behavior rely on this type of source. On the contrary, the discipline of organizational behavior utilizes researchers more frequently than average as sources for PF (e.g. Hoever et al., 2018; Vancouver & Casey Tischner, 2004), while accounting or management general almost decline to use this source. This distribution is in line with the prevalence of traditional laboratory experiments in the discipline of organizational behavior. However, an above average number of articles in accounting does not provide the specific source of the PF (e.g. J. L. Brown et al., 2016; Jermias, 2001), while only three articles of management general and organizational behavior combined follow this approach.

Both studies that investigate PF in a public context (Hoever et al., 2018; Song et al., 2018) are published in the discipline of management general. In total, most PF articles invest the effects in a private setting. Fewer studies include both, public and private settings (e.g. Tafkov, 2013) in their research designs, while this combination is more prevalent in the discipline of accounting than in organizational behavior.

Individuals are the most investigated type of recipient in PF research designs. In line with the observation by Schnieder (2018), especially the discipline of accounting is focused on understanding individual behavior with all experiments displaying PF to individuals. The disciplines of management general and organizational behavior include a small number of

studies that investigate the effects of PF on groups as well as on individuals and groups in one design (e.g. LePine et al., 2002; Maciejovsky et al., 2013).

With respect to the realism of the PF participants received, most selected articles display real PF based on actual performance to recipients (e.g. Casas-Arce et al., 2017; Tiefenbeck et al., 2018). Especially studies in accounting and management general avoid using bogus PF to investigate its effects. On the contrary, articles in the discipline of organizational behavior regularly apply bogus feedback with predetermined values without a real link to the recipients performance – also in recent years of publication (e.g. Ruttan & Nordgren, 2016; Zingoni & Byron, 2017). With the exception of one field study (S. S. K. Lam et al., 2004), all other research designs with bogus PF were implemented in a laboratory (e.g. Fernbach et al., 2014; Moore & Klein, 2008) or online experiment (Chun et al., 2018). Hence, the presentation of bogus PF in organizational behavior seems to be linked to handicaps of the types of experiments conducted. Research designs in laboratory experiments are criticized for inherent challenges of contextuality and operationalization due to the lack of a “real world-organizational context” (Hauser et al., 2017, p. 188). Since they often focus on causal relationships between specific variables (Hauser et al., 2017) and are predominantly implemented at computer-based lab in our selection of articles, researchers are either limited in the possibility to show real PF on real-effort tasks or specifically manipulate the level of PF, for instance valence of the feedback (e.g. Chun et al., 2018).

In many experimental studies on the effects of PF across disciplines, the medium to display PF was text. Displaying PF in a visual way with a graph was applied more often in the discipline of management general (e.g. Tiefenbeck et al., 2018) and almost not implemented in the other two disciplines. Further, the discipline of organizational behavior also presented PF in a verbal way or by combining multiple types of mediums at once (e.g. Alder & Ambrose, 2005; Rose & Windschitl, 2008). While the type of medium is mostly not described in sufficient detail,

especially the discipline of accounting includes multiple articles that refrain from specifying the PF medium.

Most experimental studies across disciplines display PF in-between two rounds of performance on the experimental task. Real-time PF displays or a mix of frequencies are slightly more commonly applied in the discipline of management general (e.g. Adomavicius et al., 2012; Jung et al., 2010). In contrast, the vast majority of research designs is built upon the procedure of first performing on a specific task, then receiving PF before performing on the task for a second time, like most PF reviews or displays are traditionally applied in the organizational context without implementing novel types of display due to technological advancements.

Table 9. Distribution of Characteristics Variables

Characteristics variables in accounting					
	Firm/ Supervisor	Researcher	Computer	Other/ Mixed	N/A
<i>Source</i>	3 (12%)	0 (0%)	16 (64 %)	2 (8 %)	4 (16%)
	Public	Private	Both	N/A	
<i>Privacy</i>	0 (0%)	21 (84%)	4 (16%)	0 (0%)	
	Group	Individual	Both	N/A	
<i>Recipient</i>	0 (0%)	25 (100%)	0 (0%)	0 (0%)	
	Real	Bogus	Both	N/A	
<i>Realism</i>	22 (88%)	2 (8%)	0 (0%)	1 (4%)	
	Graphs	Text	Verbal	Multiple	N/A
<i>Medium</i>	1 (4%)	14 (56%)	1 (4%)	1 (4%)	8 (32%)
	Real-time	In-between	Delayed	Mixed	N/A
<i>Frequency</i>	1 (4%)	20 (80%)	1 (4%)	2 (8 %)	1 (4%)
Characteristics variables in management general					
	Firm/ Supervisor	Researcher	Computer	Other/ Mixed	N/A
<i>Source</i>	3 (19%)	1 (6%)	9 (56%)	2 (13%)	1 (6%)
	Public	Private	Both	N/A	
<i>Privacy</i>	2 (13%)	12 (75%)	2 (13%)	0 (0%)	

	Group	Individual	Both	N/A	
<i>Recipient</i>	2 (13%)	13 (81%)	1 (6%)		
	Real	Bogus	Both	N/A	
<i>Realism</i>	15 (94%)	1 (6%)	0 (0%)	0 (0%)	
	Graphs	Text	Verbal	Multiple	N/A
<i>Medium</i>	3 (19%)	9 (56%)	0 (0%)	3 (19%)	1 (6%)
	Real-time	In-between	Delayed	Mixed	N/A
<i>Frequency</i>	3 (19%)	10 (63%)	1 (6%)	2 (13%)	0 (0%)

Characteristics variables in organizational behavior

	Firm/ Supervisor	Researcher	Computer	Other/ Mixed	N/A
<i>Source</i>	2 (4%)	11 (24%)	26 (57%)	5 (11%)	2 (4%)
	Public	Private	Both	N/A	
<i>Privacy</i>	0 (0%)	42 (91%)	3 (7%)	1 (2%)	
	Group	Individual	Both	N/A	
<i>Recipient</i>	4 (9%)	38 (83%)	4 (9%)	0 (0%)	
	Real	Bogus	Both	N/A	
<i>Realism</i>	31 (67%)	13 (28%)	2 (4%)	0 (0%)	
	Graphs	Text	Verbal	Multiple	N/A
<i>Medium</i>	1 (2%)	26 (53%)	4 (9%)	10 (22%)	5 (11%)
	Real-time	In-between	Delayed	Mixed	N/A
<i>Frequency</i>	2 (4%)	42 (91%)	0 (0%)	2 (4%)	0 (0%)

Notes: Categorization based on the authors' judgement, if no information is provided in the article 'N/A' was assigned to the variable category.

Person Variables

Regarding person variables (C) investigated in PF research designs, we can differentiate between two categories of variables: occasionally implemented and almost never implemented. Even though the large majority of PF articles does not provide insights into how person variables of the recipient (either as an individual or a group) might affect behavior, the three person variables of self-efficacy, skill/ability and performance segments in the first category get implemented occasionally, while the three person variables of self-consciousness,

personality traits and cognitive ability/traits are almost never included in PF experiments. The general scarcity of considering person variables in PF experiments across the three disciplines confirms Schnieder's (2018) finding for the subcategory of RPF in accounting that future experiments should be complemented with surveys to understand effects of person variables. Table 10 presents the aggregated results for the inclusion of person variables in our selected articles per discipline.

In the first category of person variables, which occasionally are implemented in PF research designs, performance segments and self-efficacy are used more often than skill/ability. Especially the application of performance segments further differs by disciplines. A group of articles in accounting and management general differentiates the effects of PF along specific performance segments, for instance high- or low-performer (e.g. Eyring & Narayanan, 2018) or for the best performers (e.g. Newman & Tafkov, 2014). Articles in the discipline of organizational behavior only rarely implement performance segments in the research design. Studies in accounting and management general also more often include skill or ability as a specific variable in the research design to differentiate between performance segments than organizational behavior. For self-efficacy, a person variable applied in PF research for decades (Lechermeier & Fassnacht, 2018), the implementation across discipline only shows small differences with management general analyzing its effects more often than accounting and organizational behavior.

The distribution of all three variables within the second category of person variables is quite homogeneous across disciplines: self-consciousness, personality traits and cognitive ability/traits are only applied in experimental research designs in single-digit percentages. The influence of self-consciousness on the effects of PF is only investigated in two articles in management general (e.g. Tiefenbeck et al., 2018). Specific survey items to account for the influence of individual personality traits are measured in a few articles across all three disciplines, for instance the Dark Triad of personality traits (Wang, 2017), the HEXACO

Personality Inventory (Tiefenbeck et al., 2018) or the Machiavelli personality test (Karagözoğlu & Riedl, 2015). Other established personality traits, for example the five-factor model as the “most ubiquitous and widely accepted trait framework in the history of personality psychology” (Judge & Zapata, 2015, p. 1150), are not investigated within our selected body of articles. Even though it seems intuitive to investigate the influence of cognitive ability/traits on the effects of PF, only three articles in organizational behavior introduce them in their research designs – for example by having subjects report ACT or SAT scores (DeShon et al., 2004), the results of standardized tests that are widely accepted for college admissions in the United States.

Table 10. Distribution of Person Variables

Person variables	Accounting		Management general		Organizational behavior	
	Yes	No	Yes	No	Yes	No
	<i>Self-efficacy</i>	6 (24%)	19 (76%)	5 (31%)	11 (69%)	10 (22%)
<i>Self-consciousness</i>	0 (0%)	25 (100%)	2 (13%)	14 (88%)	0 (0%)	46 (100%)
<i>Skill/Ability</i>	4 (16%)	21 (84%)	3 (19%)	13 (81%)	4 (9%)	42 (91%)
<i>Personality traits</i>	1 (4%)	24 (96%)	2 (13%)	14 (88%)	2 (4%)	44 (96%)
<i>Cognitive ability/ traits</i>	0 (0%)	25 (100%)	0 (0%)	16 (100%)	3 (7%)	43 (93%)
<i>Performance segments</i>	9 (36%)	16 (64%)	7 (44%)	9 (56%)	5 (11%)	41 (89%)

Notes: Categorization based on the definitions presented in this review.

2.3.3.3 Current State of Research Across Disciplines

In summary, our review of the main findings on the effects of PF and the underlying research designs across all three management related disciplines confirms the results of synthesizing the PF definition and theory. All three disciplines present findings on multiple variables with a shared interest, for instance the characteristics variables of timing, valence or ranks. But beyond these consistencies, findings are heterogeneous due to the diversity of additional characteristics and person variables investigated in different research designs. Especially the findings in organizational behavior demonstrate how research on the effects of RPF in this discipline tends

to apply a broader set of characteristics and person variables (in interaction) to investigate PF effects.

Considering the intersection of research findings pointed out, the degree of cross-fertilization across the three disciplines is heterogenous as well and shows that silos between the three disciplines exist on a variety of variables and findings. The recent study by Casas-Arce et al. (2017) is a good example of how more intense cross-fertilization can lead to a holistic set of findings. Published in the discipline of accounting, Casas-Arce et al. (2017) reference findings from the other two disciplines to explain the superiority of their research design and applicability of their findings within the interdisciplinary academic discourse on PF. For example, they discuss how their study differs from the study by Lurie and Swaminathan (2009) – which is set in the discipline of organizational behavior – due to their field setting, the nature of their task and the experimental treatments, even though both studies find that frequent feedback may lead to a decrease in performance. Casas-Arce et al. (2017) also include further studies from organizational behavior when reviewing the PF literature, for instance Goodman et al. (2004) or Northcraft et al. (2011) as well as studies from accounting, for example Tafkov (2013) or Hannan et al. (2013). While cross-fertilization between the disciplines of accounting and management general is more common, PF evidence derived from the consolidation of findings from all three disciplines is still the exemption, not the rule.

Reasons for the lack of cross-fertilization on PF findings across the three disciplines are expected to be twofold. First, the discipline of accounting has traditionally been “singularly concerned with the quantitative expression of economic phenomena” (Davidson, 1966, p. iii) and still is aligned on realizing Davidson’s tenet fifty years later (Bloomfield et al., 2016). Therefore, the proximity of the discipline of accounting with the discipline of management general can be assumed to be higher since management general strives for “an accomplishment of the economic and social objectives of an industrial society with increasing economy and effectiveness” (Dauten, 1958, pp. 5–6) and many findings within this discipline are based on

research by economists. In line with Zingoni and Byron's (2017) conclusion on the discipline's immanent lack of combining characteristics with actual individual behavior, the disciplines of accounting and management general prioritize findings on applicable variable levels that decision makers in organizations can actually manage (Casas-Arce et al., 2017).

Second, our synthesis shows that the lack of consistency in using terms for key variables might slow down cross-fertilization between studies and disciplines regarding the main findings from PF. To provide an example, we focus on the vast body of PF research that investigates the effects of RPF or rank distributions. A stream of literature across all three disciplines has investigated underlying mechanisms and a diverse set of confounding variables for this topic and almost all studies to date have the same focus on the comparisons recipients perform when their own performance is displayed relative to others (Kachelmeier, 2019). Analyzing recent studies on the topic of RPF, we find cross-citations by accounting studies that reference studies from management general and organizational behavior (e.g. Eyring & Narayanan, 2018; Loftus & Tanlu, 2018). We also find studies from management general that cite studies from accounting, but not from organizational behavior (e.g. Gill et al., 2019; Song et al., 2018). Still, we find organizational behavior to be the most self-centered discipline with Zingoni and Byron (2017) or Chun et al. (2018) not citing any study from accounting and only seldomly citing management general. We suggest that the variety of terms within RPF is heterogenous and contributes to the status quo of a lacking interdisciplinary discourse to advance findings. In the disciplines of accounting and management general, the term RPF is established, even though the exact definition of what type of relative comparison is included can vary dependent on the study. Still, Gill et al. (2019) use *relative performance feedback* as a keyword even though they use the synonym of *rank-order relative performance evaluation* throughout their study to describe the research topic and refer to it. In contrast, multiple terms are used within organizational behavior to describe RPF: Bonner et al. (2002) call it *ranking information condition*, Moore and Klein (2008) call it *comparative feedback* and Hartwell and Campion

(2016) call it *normative feedback*. This variety of terms also prevents the exemplary findings to be included in recent literature reviews on the topic since they are not covered by the keywords, for instance by Schnieder (2018). Multiple synonyms can be found for other variables as well, for instance timing or frequency and feedback valence or feedback sign.

Taken together, even though PF research in all three disciplines has advanced our understanding on specific underlying mechanisms, more iterating research across disciplines is needed to further explore the effects of PF based on specific elements of the integrative definition, so that future reviews are able to address the guiding question about the effectiveness of PF by reviewing synergetic findings from multiple management related disciplines.

2.4 Conclusion

While empirical research on PF effects dates back more than a century, the last two decades produced a consistently growing body of research in the disciplines of accounting, management general and organizational behavior while adopting new research designs with an increasing number of field experiments. Consequently, individual studies present findings on specific variables on the relationship between PF and its effects. However, the overall research area is still fragmented, and this state further perpetuates a pre-existing fuzziness of the concept of PF and interferes with a comprehensive understanding of underlying mechanisms. Even though literature reviews have provided several overviews with specific methodologies and scopes to try to respond to the guiding questions of how and when to provide PF, the narrow scoping of reviews and empirical studies within this research area has prevented a sequential and iterative stream of research. This review conducts a systematical review and synthesis of management related literature on the effects of PF in three disciplines in order to build the foundation for future empirical research designs and theory building by organizing, synthesizing and categorizing PF research from the last twenty years.

To organize the current state of the managerial research on PF effects, we present a systematic literature review of 87 selected articles published in the highest-ranked journals within the disciplines of accounting, management general and organizational behavior. We synthesize and categorize this body of articles in a visual way along four main RQs to make the current state of research transparent: (RQ1) the definition of the term PF, (RQ2) the theories investigated to explain PF effects, (RQ3a) the main findings in PF research and (RQ3b) the main research designs. As part of our synthesis of the current definitions of the term, we propose an integrative definition based on prior studies and reviews to consolidate future PF research and subsequently structure our review along this definition. With regards to the theories applied, we find a shared emphasis on social comparison theory in addition to heterogeneous applications of specific theories within disciplines and a consistently large number of exploratory theory sections. Synthesizing the main findings on the effects of PF, we present the main results for each article and discuss current findings for the most investigated variables. In addition, we identify a current lack of cross-fertilization, assumed to be due to the immanent nature of the three disciplines and differences in the definition and application of key terms between disciplines. To address the second part of the third question, our review organizes the application of characteristic and person variables in each discipline. While the disciplines differ in their focus on specific characteristics variables, we conclude that a general lack of integrating person variables in PF research designs is prevalent. We also find a common focus across all three disciplines on laboratory experiments with student samples. On the contrary, the type of compensation scheme applied is differentiated between disciplines.

Based on the reviewed literature in three management related disciplines, three promising opportunities for future research arise. First, future research should increase the amount of cross-fertilization and subsequent iterations on the findings for specific mechanisms of PF effects. At this point, we identify room for further research on all elements of our suggested definition, since the lack of transferability among empirical studies – within and across

disciplines – can partly be attributed to incoherent variable names and a lack of shared variable definitions. PF research has been under criticism for not providing cumulative findings and still produces contradicting results (Schnieder, 2018). Therefore, research on PF should map its findings to the elements of the integrative definition to allow fellow researchers to find results and develop a consistent terminology for specific terms to align findings across disciplines and research designs.

Second, future research should emphasize employing field settings in combination with additional (survey) data on PF recipients to holistically study the effects of PF. As we have shown in this paper, the majority of scholars has opted for laboratory experiments with student participants across all three disciplines. Even though laboratory experiments are predominantly used to investigate clean designs and establish causality, our review also identified a current lack of investigating person characteristics, which are typically measured on a survey basis within laboratory designs. In addition, recent field experiments have provided evidence for contradictory results in comparison to prior, laboratory-based studies on multiple PF variables. In line with recent calls in management related disciplines to conduct randomized experiments in field settings (e.g. Floyd & List, 2016; Hauser et al., 2017) and emerging methodologies to combine laboratory and field experiments applied in economics, for instance the lab-in-the-field design (Gneezy & Imas, 2017), we call to advance the applied research designs on this topic in the future.

Third, future research should refine the theoretical background to predict PF effects. Our review shows that only few theories are applied consistently across disciplines and many studies combine multiple theories in an explorative way. Therefore, we recommend that future research on PF effects either specifically focusses on testing the specific mechanisms of key theories being investigated (i.e. social comparison theory) or considers additional, interacting theories that might explain mechanisms where a current theoretical explanation is missing. Taking into account additional theories seems especially promising for future research that addresses the

lack of person variables being investigated in this topic, as individual limits on processing information quantity and frequency have been a relevant topic in both psychology and behavioral economics research (e.g. Gigerenzer & Goldstein, 2011). Since field research methods are helpful to enable theory development (Feldman & Orlikowski, 2011), we call for future experimental studies to advance the current body of theory by providing the basis to develop PF theories.

Despite the systematic process applied within this literature review to synthesize PF research, this paper has limitations which relate to the methodology and the scope. In terms of methodology, this review does not include all relevant articles on the effects of PF since we rely on a selection of the highest-ranked journals per discipline as a starting point. Still, our systematic approach allows us to synthesize a literature base that fulfills the highest standards in management research with articles that advance the understanding of its respective discipline with strict peer-review requirements. Prior reviews waived a quality threshold of this nature to strive for exhaustive results – an objective no literature review can fulfill due to the inherent restrictions deriving from choosing specific keywords. Moreover, we include additional synonyms to limit the probability of omitting relevant studies. Also, with regards to methodology, our review is qualitative in nature and might include limitations concerning the objectivity of our analyses. We address this issue of subjectivity by conducting a transparently documented review process and building our analysis on frameworks and findings from prior literature reviews. In addition, we discussed our methodology in detail, for instance each selection criteria, with fellow researchers to guarantee the validity of our approach. While conducting a meta-analysis can provide objective, quantitative comparisons of objective results, we assess the current state of research as too heterogeneous and aspire to organize the current body of PF research to provide the basis for meta-analyses in the future. In terms of scope, the three management related disciplines of accounting, management general and organizational behavior are not the only disciplines in which PF research has been conducted. Our aim, on the

contrary, is not to provide an overview of all PF research, but to allow for comparisons between three disciplines which emphasize applicable findings for managers in organizations. These limitations might present avenues for future research.

Overall, this literature review contributes to the accepted and growing stream of management research on PF and offers a synthesis of definitions, theories, research designs and findings for both researchers and practitioners. Nevertheless, additional research is necessary to advance the understanding of specific mechanisms and to enable organizations to manage PF effectively in specific settings.

3 Paper II: Less Information, More Comparison, and Better Performance: Evidence from a Field Experiment^{9 10}

3.1 Introduction

In this paper, we use a field experiment to understand whether providing relative, absolute, or both absolute *and* relative PF works best to motivate workers and improve performance. Many organizations across a range of settings select among these different types of PF (e.g. Casas-Arce et al., 2017; Y. Chen et al., 2010; Song et al., 2018). While the prevalence of PF in practice has motivated a growing literature on the topic, research has not yet compared these three alternative PF designs in the same setting (Schnieder, 2018). We address this gap with a field experiment in professional European football.

Theory from economics and psychology contends that RPF influences performance by facilitating peer comparison and, thereby, incentivizing individuals to outperform their peers and attain a positive self-image (Falk & Ichino, 2006; Festinger, 1954; Mas & Moretti, 2009). This suggests not only a mechanism by which RPF affects performance, but also specific conditions under which these performance effects should be most pronounced – i.e. in the presence of factors that further facilitate social comparison or heighten self-image concerns. Furthermore, standard economic theory predicts that among Bayesian rational agents, more information is weakly better at aiding decision-making than less information (Savage, 1954). Thus, to the extent that relative performance information facilitates social comparison and the addition of absolute performance information aids decision-making, providing absolute and relative performance information should lead to as large or larger performance effects than providing either type of feedback in isolation.

⁹ Eyring, H., Ferguson, P. F. & Koppers, S. (2019). Less Information, More Comparison, and Better Performance: Evidence from a Field Experiment. Unpublished Working Paper.

¹⁰ We list the co-authors of Paper II in accordance with the alphabetical-ordering rule which is generally applied for top-quality journals in the discipline of accounting (K. C. Chan et al., 2009).

However, we also know from behavioral theories of decision-making that mental processes that dictate performance may mean it is best to draw attention to certain information by omitting other information (Gigerenzer & Goldstein, 2011; Kahneman & Tversky, 1972). In a feedback context, such an approach is particularly beneficially if the provision of additional information induces information overload or inhibits social comparison by reducing the salience of relative performance information. This suggests that more information may not always be better. Thus, it is not obvious if feedback that contains relative information leads to higher or lower levels of performance than feedback that contains both relative and absolute information.

To test these conflicting predictions, we extend the literature on feedback by comparing the performance effects generated by absolute, relative, and absolute and relative PF. To understand the mechanisms that give rise to these performance effects, we also consider how the results depend on the number of measures included in PF, a participant's involvement in social comparison, and a participant's reports of information overload. Finally, we consider the role of the recipient's baseline psychological profile in order to explore cross-sectional variation in the advantage of providing either absolute, relative, or combined absolute and relative PF.

In collaboration with Anton Paar SportsTec – the producers of *skills.lab*, a state-of-the-art football training simulator used by elite athletes in Germany and Austria – we design and implement a field experiment in which we manipulate the PF received by 117 professional and semi-professional football players during training. We intervene in the feedback system utilized by *skills.lab* by varying the attributes of the PF users receive on two key metrics – passing accuracy and speed – after completing a standardized training drill. We do so by employing a 3x2 research design, where players are randomly assigned: 1) feedback on either their absolute, relative, or absolute *and* relative performance (as captured by their performance rank); and, 2) feedback that contains either aggregate or detailed performance information (i.e. a smaller or larger number of performance measures). To study the responses of the players, we use data on a range of actions taken by the athletes. We draw on survey evidence to understand how

behavioral responses to the feedback vary based on psychological profiles. The field setting and our intervention do not involve explicit performance incentives (although strong implicit incentives in the form of career concerns are present), which helps us to identify the distinct information effects of the PF.

We show that providing relative performance information in feedback is more effective at lifting performance than providing absolute performance information in feedback. Furthermore, we show across a range of different dimensions of performance that feedback that contains *only* relative performance information is more effective – or at least as effective – as feedback that contains *both* relative and absolute PF. Along the dimensions of performance we measure and provide feedback on, players who receive feedback that contains only relative performance information are roughly 14 percent more accurate and 17 percent faster than players that receive feedback that contains both relative and absolute performance information. Thus, among the types of displays we test, providing feedback that contains a subset of an information set dominates providing feedback that contains the full information set – a result consistent with behavioral theories of decision-making and in conflict with standard economic theory.

We do not find evidence that this result is explained by information overload that arises from presenting individuals with too many measures per se.¹¹ Specifically, we do not find that adding more measures while holding constant whether feedback is absolute or relative leads to a decrease in performance. On the contrary, only increasing the scope of feedback to contain absolute along with relative information leads to a decrease in performance. Regardless of the number of measures supplied to the subjects in our experiment, feedback that contains *only* relative performance information weakly dominates feedback that contains *both* absolute and relative performance information and strictly dominates feedback that contains *only* APF.

¹¹ Information overload literature describes how excess quantities of information can induce stress and mental exhaustion that may lead to poorer reasoning and decision-making (Miller, 1956; Simnett, 1996; Tuttle & Burton, 1999).

Furthermore, we do not find that individuals are more likely to report having received ‘too much’ information when exposed to feedback with both absolute and relative performance information.

To understand why increasing the scope of feedback beyond relative information results in a decrease in performance, we consider the role of social comparison. We use survey evidence to show that restricting PF to only relative information increases reported involvement in social comparison. In turn, we find that increased involvement in social comparison is associated with improved performance. This suggests that RPF in isolation is most effective because the mixture of different information types diminishes the salience of social comparison that occurs when relative information alone is displayed.

Our cross-sectional tests suggest that the benefits of relative performance information are not as strong for people who have lower levels of “conscientiousness”, a personality trait associated with diligence and thoroughness (Judge & Zapata, 2015). In addition, the benefits of relative performance information are stronger for participants who are highly interested in learning about areas for improvement. Finally, for these participants who are highly interested in learning about areas for improvement, we find less of a difference in performance effects of providing relative information alone versus relative information along with absolute information. This last cross-sectional result suggests that the salience of information may diminish at a slower rate for people who are more interested in what the information has to offer them – in this case, ranking-based information that may help in identifying areas with room for improvement.

This paper contributes to two main streams of literature. First, we contribute to the economics and accounting literature that addresses forces other than explicit contracts for inducing desired behavior. Studies have established that forces such as reciprocity, preference alignment, culture, and social comparison can be effective nonfinancial means of resolving agency conflicts (Campbell, 2012; Dyreng et al., 2012; Sobel, 2005). Within this stream, we extend the research

that focuses on social comparison as a motivating force. Our results document that restricting performance information to contain relative measures can amplify social comparison and benefit performance. Moreover, our cross-sectional results offer evidence on the interaction between preferences and social comparison, demonstrating that those who are more conscientious and more interested in identifying areas for improvement are those for whom the benefits of social comparison are strongest.

Second, this paper extends the growing body of research on feedback (S. E. Bonner & Sprinkle, 2002; Y. Chen et al., 2010; Song et al., 2018; Tafkov, 2013). Prior work has tested effects of feedback design elements, including feedback frequency, benchmarks for peer comparison, and messages of congratulations (Casas-Arce et al., 2017; Y. Chen et al., 2010; Eyring & Narayanan, 2018). However, while examples of PF displays that contain only relative, only absolute, or combined relative and absolute information are prevalent among these studies, a fundamental question remains in terms of which approach is most effective. We address this question by employing a field experiment that utilizes a real-effort task to compare these alternative feedback designs.

The remainder of this paper is organized as follows. Section 2 outlines the relevant theory and develops the hypotheses. Section 3 describes the field setting and provides background on the research site, skills.lab, and the industry in which it operates. The design of the field experiment as well as data is described in Section 4. Section 5 reports the results of our empirical analysis. Section 6 concludes the paper.

3.2 Theory and Hypotheses

RPF is information that allows individuals (or groups of individuals) to assess their performance in relation to peers (or peer groups) (Schnieder, 2018). RPF is regularly used in practice and often widely applied within organizations across a range of industries and settings. For instance, RPF is frequently used in the healthcare sector to motivate workers and facilitate

the exchange of best practices amongst employees (Song et al., 2018). RPF is also provided in settings as varied as retail banking (Gino & Staats, 2011), education (Eyring & Narayanan, 2018), and the public sector (Charness et al., 2014). Furthermore, firms commonly provide RPF to their employees even when employees' compensation is not explicitly tied to peer performance (Nordstrom et al., 1991; Tafkov, 2013).

A variety of academic disciplines have explored the mechanisms by which RPF influences performance. The economics, accounting, and psychology literatures argue that RPF triggers individual concerns for the self- and social image (Falk & Ichino, 2006; Mas & Moretti, 2009; Tafkov, 2013). By activating these image concerns, the provision of RPF leads to changes in the behavior of recipients. Social comparison theory and impression management theory are the two frameworks most commonly used in these literatures to explain how RPF activates image-related concerns and thereby influences performance (Schnieder, 2018). In this study, we focus on social comparison theory, which accounts for self-image concerns to explain how *private* RPF shapes behavior (in contrast, impression management theory draws on social-image concerns to understand how *public* RPF shapes behavior).

The core principle of social comparison theory is that individuals have a strong tendency to compare their own attributes or achievements with those of other individuals (Festinger, 1954; Suls & Wheeler, 2000). In doing so, individuals directly derive utility from possessing attributes or achievements that other individuals lack. This tendency towards social comparison is particularly pronounced with regard to an individual's abilities (D. J. Brown et al., 2007; Suls & Wheeler, 2000). As ability is often unobservable, individuals will use their performance on some relevant task as a proxy for their ability. RPF serves as a key input for individuals to compare their abilities in a 'social' context. Furthermore, seeking to maximize their utility by achieving and maintaining a positive self-image, individuals will behave in a manner so as to achieve high relative performance compared to their peers (D. J. Brown et al., 2007; Garcia &

Tor, 2007). In this way, RPF is argued to foster competition and thus lead individuals to improve their performance for the purposes of achieving and maintaining a positive self-image.

In terms of empirical findings, many studies have found evidence that RPF affects performance (Schnieder, 2018). In general, researchers – in both the field and lab – have found that RPF generates positive performance effects. These positive effects are typically attributed to individuals increasing the duration and intensity of their effort in response to RPF (Schnieder, 2018). Hannan et al. (2008) employs a lab experiment to show that providing RPF improves the mean performance of participants compensated under an individual incentive scheme. Similarly, Blanes i Vidal and Nossol (2011) uses personnel records to show that RPF in a tournament setting without prizes has a positive effect on worker output. Using a lab setting to explore the interaction between feedback and compensation schemes, Tafkov (2013) finds that performance pay increases the performance effects of RPF.

However, there is a small number of studies across a range of industries that find that RPF produces no effect (Eriksson et al., 2017) or negative effects. For instance, a field experiment in a school setting by Bursztyn and Jensen (2015) documents negative effects of RPF on performance. These authors find that RPF is especially ‘demotivating’ for high performers. Similarly, Ashraf (2018) finds that RPF has a negative effect on performance in a manufacturing setting. The author argues that this result is driven by the need for conformity in groups and the role of social pressure in dampening performance improvements.

Nonetheless, weighing up the wealth of evidence in support of relative feedback against the small number of studies that suggest the contrary, we propose the following hypothesis:

HYPOTHESIS 1a (H1a): *Providing RPF positively affects performance compared to providing APF.*

While the balance of theory and prior evidence suggests that we will find a positive effect of providing relative as compared to absolute PF, research has yet to compare those alternatives

with a third approach – the combination of both relative *and* absolute PF. A priori, the performance effects of providing feedback that contains both absolute and relative performance information is not clear. Most obviously, providing both absolute and relative performance information rather than *only* relative performance information results in a doubling of the quantity of information disclosed in a feedback intervention. According to standard economic theory – and frameworks in the decision sciences, more broadly – more information is strictly viewed as better than less information (Savage, 1954).

This perspective on the benefits of increased information is typically grounded in Bayesian models of decision-making. According to these models, an individual is able to consume and process all available information and use this to rationally update their beliefs (Arrow et al., 1949). In accordance with this theory, more information leads to better calibrated beliefs and by extension more informed decisions. In terms of mapping this to performance, more informed decisions improve performance by helping individuals better understand when, where and how to direct their effort when performing a task and by better facilitating learning and the accumulation of human capital (S. E. Bonner & Sprinkle, 2002; Schnieder, 2018). Thus, if individuals behave in a manner consistent with Bayesian rational agents, providing feedback that contains both absolute and relative performance information should lead to as large or larger performance effects than providing individuals with feedback that contains only relative performance information.

However, more recent work in behavioral economics and psychology has brought into question the view that more information is strictly better (Kahneman & Tversky, 1972; Simon, 1979; Tversky & Kahneman, 1974, 1983). According to this perspective, information is costly for decision makers to consume and process. In order to economize their limited cognitive resources, when faced with large amounts of information, individuals may employ a range of heuristics (Goldstein & Gigerenzer, 2002). As adaptive mechanisms, these heuristics may lead

to efficiency gains in information processing, but at the cost of introducing certain biases into the decision-making process (Gigerenzer & Goldstein, 2011).

For instance, large amounts of feedback (or feedback that is too detailed or too frequent) may inhibit an individual's ability to identify general trends or patterns in their performance information (Tversky & Kahneman, 1974). As a result, excess amounts of information may distract individuals' attention from informative signals (Gennaioli & Shleifer, 2010). By misdirecting the recipient's attention, these heuristics may lead to suboptimal assessments or decisions (Rabin, 2002).

Importantly, these types of biases can lead to behaviors that harm performance. Firstly, too much information may induce cognitive overload and introduce a distraction or anxiety effect, whereby the recipient disengages from the task they are required to perform, and instead solely focuses on trying to process the information they receive (Malhotra, 1982; Wurman, 1990, 2001). Secondly, the provision of additional information may crowd out the salience of relative performance information, and thus dampen social comparison. This would serve to mute the performance effects of relative feedback. As such, increases in the quantity or variety of information contained in PF may in fact harm decision making. This suggests that providing feedback that contains both absolute and relative performance information may actually dampen performance. As such, we propose the following hypothesis:

HYPOTHESIS 1b (H1b): *Providing feedback that contains only relative performance information positively affects performance compared to providing feedback that contains both absolute and relative performance information.*

As touched on above, information overload may cause individuals who receive feedback that contains only RPF to outperform individuals who receive feedback that contains both relative and absolute PF. In short, information overload arises when individuals are forced to process

‘too much’ information (Eppler & Mengis, 2004). In a seminal study of the topic, Schroder et al. (1967) propose a model for how information load affects individual decision-making. This model suggests an inverted-U relationship between information load and performance. Specifically, task performance will initially improve as more information is received. However, performance quickly begins to decline as the quantity of information exceeds some threshold determined by the decision-maker’s processing capacity.

Subsequent work on this topic in psychology has sought to pin down the specific mechanism by which information overload reduces performance. One such theory is that when presented with large amounts of information, individuals over-allocate cognitive resources and attention towards processing the available information (Schick et al., 1990). This can introduce stress, anxiety, or over-stimulation (Bawden, 2001). When this arises, additional information hinders an individual’s ability to not only execute the task or job at hand, but also to make any ‘use’ of the information provided.

There are a number of empirical studies that report findings consistent with information overload. Miller (1956) is an oft-cited study which suggests individuals may experience information overload when presented with as few as seven ‘cues’ or pieces of information. Similarly, Chewing and Harrell (1990) find that high-ability individuals may only be able to process upwards to nine pieces of information. Looking at learning and strategy development as the primary outcome of interest, Goodman et al. (2004) show that increasing the specificity of feedback is beneficial for initial performance but discourages exploration and undermines the learning needed for later, more independent (unsupervised) performance. Information overload has also been a topic of interest to applied researchers in management and accounting (Eppler & Mengis, 2004). In a lab experiment, Simnett (1996) finds that the decision-making of auditors deteriorates in information load when information processing is constrained. Tuttle and Burton (1999) find a similar set of results when conducting an investment analysis task in the lab.

If information overload is the factor that causes the treatment of RPF to yield better results than the treatment of combined relative and absolute PF, then we should also find that increasing the number of measures in either of these two treatments leads to a performance decrement. In short, information overload arises due to increases in the amount of information contained in the feedback, and not differences in the variety or types of information disclosed. From this line of reasoning, the following hypothesis is consistent with information overload:

HYPOTHESIS 2 (H2): *Holding constant whether measures are relative or absolute, increasing the number of measures that describe a given performance construct leads to a reduction in performance.*

However, information overload is not the only explanation for why relative feedback may outperform absolute and relative feedback. Alternatively, feedback that contains only relative performance information may also outperform feedback that contains both absolute and relative performance information if the addition of absolute performance information reduces the salience of relative performance information and thereby dampens social comparison. As social comparison has been shown to spur competition and boost performance, this ‘salience effect’ then serves to mute – or even potentially cancel out – the performance benefits associated with relative performance information.

Psychologists view salience detection as a key attentional mechanism enabling humans to focus their limited cognitive resources on a relevant subset of the available sensory data (Bordalo et al., 2012). In the literature on salience, decision makers do not take into account all the information available to them. Gennaioli and Shleifer (2010) call such decision makers local thinkers, because they neglect potentially important data. Studies in economics and finance have shown in the contexts of taxes and investing that the salience of relevant information for decision-making can be diminished by the presence of other information or by the context of the information display (Barber & Odean, 2008; Chetty et al., 2009). Relatedly, in accounting, Casas-Arce et al. (2017) find that providing more frequent feedback can cause the presence of

recent performance information to reduce the amount of attention paid to long-term performance information, and that this hampers learning.

Based on prior research that shows the performance benefit of RPF as compared to APF, if the latter directs attention away from the former in a way that dampens involvement in social comparison, we would expect to see a decrease in performance. Linking the salience of relative performance information to performance via social comparison, we propose the following set of interrelated hypotheses:

HYPOTHESIS 3a (H3a): *Providing feedback that contains only relative performance information increases involvement in social comparison compared to providing feedback that contains both absolute and relative performance information.*

HYPOTHESIS 3b (H3b): *Increased involvement in social comparison is positively associated with performance.*

3.3 Setting

3.3.1 Field Site

Our study is set at skills.lab, a training facility for professional football players. skills.lab is owned and operated by Anton Paar SportsTec – a subsidiary of the Austrian firm Anton Paar, a producer of analytical instruments for laboratories and other analytical technologies. Professional and semi-professional football players across Austria and Germany use the simulator. The simulator replicates game-like situations and common training drills in a highly controlled environment. The automated training sessions occur within the simulator's six-walled, 320 square meter, turf-covered playing field. The simulator uses automated ball machines, integrated video projectors, as well as camera and laser tracking technology to measure passing speed and accuracy (see Appendix 2).

A passing drill in the skills.lab simulator works as follows: a player positions themselves in the center of the training surface, a ball is then fired at the player from one of four ball machines, the player is required to chase down and control the ball, once the ball is under control the player must pass the ball by foot to a static or dynamic target that is projected onto the walls of the simulator. In a standard passing drill, a player is required to complete this passing task 12 times; in a standard session, a player will be required to complete the passing drill several times. The simulator measures and evaluates ball-handling and passing skills along two dimensions: passing accuracy, or how close the player was to hitting the target, and passing speed, or how quickly the player was able to control and pass the ball to the target. Once a player has completed the drill, the simulator provides feedback along these two performance dimensions. The feedback is projected onto a wall of the training simulator and provides the player with information on his or her passing accuracy and speed.

The passing drills at skills.lab are frequently used for testing and training purposes by professional and semi-professional football teams in Austria and Germany. The drills are designed to replicate training exercises and game scenarios commonly employed in ‘traditional’, ‘outdoor’ practice sessions. Coaches and players pay to use skills.lab, and report that it provides valuable PF and a unique setting within which to fine-tune technical skills. The CEO of Anton Paar SportsTec echoes this point: “Professional players are accustomed to being constantly evaluated, but they particularly appreciate an objective performance analysis in skills.lab, as they are usually evaluated only subjectively and thus possibly incorrectly.” The relevance of the simulator to coaches and players is reflected by the fact that first-division football teams in Europe have recently purchased and ordered skills.lab simulators of their own to be produced and installed by Anton Paar SportsTec.

The combination of a controlled environment and a real-effort task as used in practice provides a number of advantages for a field experiment. First, we are able to randomly assign feedback conditions within the context of an athlete carrying out his or her job responsibilities. Second,

as the simulator is fully programmable and automated, we are able to keep the other elements of the experiment finely tuned and perfectly replicable. Third, the advanced diagnostic focus of skills.lab allows for precise feedback and precise measurement of effects.

3.3.2 Industry and Incentives

Professional football is a \$31.6 billion business (Deloitte, 2019), bringing in roughly triple the revenue of any major American sport. In the 2018/2019 season, the average annual salary for an athlete playing first-division football in one of Europe’s top-five leagues was 2.4 million \$ (Ingle, 2018). At the same time as elite players receive high salaries, jobs in professional football are scarce. Only one out of every 2,000 registered football players in Germany (Giersch, 2014) – one of the largest and most-successful football-playing nations – earn a living playing the sport. Naturally, this creates intense competition amongst football players from a young age to secure and maintain a contract with a professional team. The competition is amplified by the fact that these athletes have relatively short careers and so need to achieve most of their financial payouts before their mid-thirties (PFA, n.d.). The relative scarcity of jobs in professional football means that professional and semi-professional players have very strong incentives to develop and maintain their skills and fitness. As such, these players view training less as ‘practice’ and more as a central driver of success in their career.

Athletes who use skills.lab belong to top-level teams in Austria and Germany, which sign players to performance-contingent contracts. Alongside their regular salary, players receive a number of bonuses – for instance, a bonus for being chosen to play in a game or for achieving specific performance criteria (e.g. for goals scored or assists played). If a player consistently performs well in games, that player can receive financial rewards aside from a bonus, such as re-signing a contract, obtaining more favorable contract terms, or signing under better terms with a rival team. Representatives at our field site noted that performance in training, including at skills.lab, can influence management decisions to enter a player in a game, yielding associated financial payouts. skills.lab provides coaches and managers with information on

their players' performance in the simulator that would allow comparison among players. Thus, while we do not provide pay for performance at skills.lab, our setting is one in which an effort to outperform peers plausibly derives in part from implicit financial incentives.

Our setting is also likely one in which social comparison is relevant due to self-image concerns. These players compete in large stadiums with television and social media audiences. Given the level of publicity around these careers and the amount of time and focus devoted from a young age, being a professional football player is generally a fundamental part of their self-image. Therefore, in addition to training and developing their skills to secure and maintain lucrative club contracts, football players also drive themselves to improve so as to protect their identity as elite athletes.

3.3.3 Performance Constructs: Handling and Passing

The simulator at skills.lab employs a range of testing and training drills designed to evaluate ball-handling and passing – fundamental technical skills required from all football players, regardless of playing position or style. In fact, ball-handling and passing are widely viewed as skills that distinguish the very best players even at elite levels. As such, there is a strong demand from professional teams for football players who are highly skilled in these areas. Accordingly, players will typically spend the bulk of their time in practice – upwards of two to three hours a day – performing training drills and exercises designed to hone these skills. For these same reasons, skills.lab has selected these highly relevant dimensions of performance in anticipation that coaches and players will be interested in the skills obtained and the PF provided.

3.4 Experiment Design

3.4.1 Subjects

In total, 117 professional football players participated in the experiment. Table 11 provides summary statistics on the demographics and football experience of these individuals. The age of the average player involved in the study was 21,44 years. The sample is composed of 18

female players and 99 male players, each of whom earn a salary from playing football either as a professional or semi-professional player.

Participants were sourced from 11 different professional clubs, with four clubs providing 79 percent of all participants. On average, the participants started playing football at five years of age and have approximately 15 years of experience competing in the sport. Players in the sample frequently practice their skills in training and matches, with the average player training just over five times a week and competing in matches most weeks during the season. 72 percent of the individuals in our sample were either starting or substitute players for a professional team in the 2018-2019 football season. Around ten percent of players in our sample are goalkeepers, 30 percent are defenders, 44 percent are midfielders, and 16 percent are forwards. This reflects the standard composition by position of players on a team's roster. Almost 85 percent of the sample had trained in the simulator prior to the experiment. Furthermore, players in the sample had trained in the simulator an average of just over five times.

3.4.2 Real-Effort Task

Participants in the experiment were required to complete a passing drill two times. This occurred at the skills.lab simulator and closely resembled the drills completed by professional football players in training (see Appendix 3). The training drill required the players to complete 12 passes. To complete a pass, the players were required to receive a pass fired sequentially from one of two ball cannons located on the left and right-sides of the simulator. After receiving the pass, the player then passed the ball to a dynamic target projected onto one of the walls of the simulator. The target consisted of a semicircle with a diameter of 185 cm at the feet of a projected teammate (see Appendix 4).

The specifications of the drill, including the speed of the passes and the target for passes, were adapted for our study so that no participant had played this specific drill before. To prevent participants from anticipating the location and movement of the target, a sequence of 12 starting

and end positions for the target (one for each ball) was randomly drawn for each round. Players had a maximum of about 11 seconds to complete a pass before the next ball was fired from the simulator's ball cannons.¹²

Participants completed the task once, then received feedback on their performance. Participants then completed the drill a second time and again received feedback on their performance. This feedback showed only the player's performance on the most-immediate iteration of the drill. Participants did not exit the simulator between drills and were not exposed to other participants during treatment.

3.4.3 Treatment Conditions

Each football player involved in the experiment was randomly allocated to one of six treatment groups that received different forms of feedback whilst training in the simulator. On the social-comparison dimension of treatment, we randomly assigned players to receive either relative, absolute, or both relative and absolute performance information. On the detail dimension of treatment, we randomly assigned a player to receive either a report of average performance for all passes during the round (i.e. aggregate feedback), or to receive a report of average performance for passes that required a right turn, average performance for passes that required a left turn, and average performance for all passes during the round (i.e. detailed feedback) (see Appendix 5). The breakdown by turn direction is relevant given that it requires the use of dominant and non-dominant feet in a different combination, which can lead to slower times for the player's less-preferred side. A player who can make passes effectively whether the ball comes from a direction that requires a right or left turn is more valuable to a team. The procedure of the experiment was based upon the regular flow of football players through the skills.lab training facility (see Appendix 6 and Appendix 7).

¹² Based on historical user data, this time limit is the standard setting employed by the site when conducting drills in the simulator.

Players in the *Absolute* performance information treatment arm received the following feedback on passing accuracy and speed of execution: percentage of passes on target, average distance of pass from the center of the target (in meters), fastest time handling ball (in seconds), and average time handling ball (in seconds). We describe these measures more fully in section 3.4.4 and in Appendix 9. Players in the *Relative* performance information treatment arm received feedback on the same performance measures but presented in the form of a performance rank for each of the measures (e.g. 19 of 32 for average time handling the ball, etc.). To arrive at this rank, each participant's performance was compared to the performances of players from a common reference group. The reference group consisted of 31 professional or semi-professional football players who completed the same training drill in the simulator at skills.lab during the pre-experiment period. No player from the reference group took part in the experiment. The players in the reference group were of a similar age, level of ability and level of experience as the subjects who took part in the experiment. Finally, players in the *Absolute&Relative* performance information treatment arm also received the same performance measures but presented in both absolute and relative terms.

As noted earlier, players in the *Aggregate* treatment arm received this information on these four performance measures as an average for all passes received during the drill. Players in the *Detail* treatment arm received this same information along with a breakdown showing the average for passes that required the player to turn to their right, and the average for passes that required the player to turn to their left. As such, players in the *Detail* treatment arm received three times the number of performance measures as players in the *Aggregate* treatment arm.

The interaction of these two treatment arms led to six treatments: *Absolute x Aggregate*, *Absolute x Detail*, *Relative x Aggregate*, *Relative x Detail*, *Absolute&Relative x Aggregate*, *Absolute&Relative x Detail* (see Appendix 5 and Appendix 8). In line with prior field-based research on PF, we do not have an experimental group that received no PF (Casas-Arce et al., 2017; Song et al., 2018). We use this approach because football players who train at skills.lab

expect to receive PF after each drill. The provision of objective performance measures is a key feature of the simulator experience, and so the absence of feedback would essentially defeat a key purpose of their attending the training and could cause them to feel that something went wrong. In terms of the total number of performance measures contained within the different feedback treatments, *Absolute x Aggregate* and *Relative x Aggregate* featured four measures; *Absolute x Detail* and *Relative x Detail* featured 12 measures; *Absolute&Relative x Aggregate* featured eight measures; and, *Absolut&Relative x Detail* featured 24 measures.

3.4.4 Measures

Our study employs player performance data at the drill-level. Our measures of performance on the drill are the main dependent variables and capture the accuracy and speed of the passes completed by each player. To describe the sample and address effect moderators and mechanisms, we gathered demographic data and self-reported measures of behavioral characteristics and responses to the feedback.

In our analysis, we use the two measures of passing accuracy reported to players in their PF while in the training simulator. The first, *Hitrates*, is defined as the percentage of passes on target a player completed during the drill. A score of 100 percent indicates that a player successfully hit the target for all 12 passes during the drill. *Distance* is a continuous measure of accuracy and is defined as the average distance (in meters) from the center of the target for the 12 passes completed by the player during the drill. The lower the score on this measure, the more accurate the player's passing.

We also use the two measures of ball-handling and passing speed recorded by the training simulator and included in the PF. *Fast_Time* is the quickest time (in seconds) it took a player to complete a pass during the drill. *Avg_Time* is the average time (in seconds) it took a player to complete a pass over the 12 passes completed during the drill. For both of these measures,

lower times reflect quicker ball-handling and passing, meaning better performance on dimensions of speed.

Our analysis also uses player demographics and responses to surveys. We observe the age, gender, height and weight, playing attributes, experience, and language of all subjects in our sample. We collected information on personality traits using a post-survey questionnaire that identified the Big Five personality trait ‘conscientiousness’ along with a variable regarding interest in learning about areas for improvement. Using the post-survey questionnaire, we also collected information on social comparison and on perceptions of the quantity of information provided.

Appendix 9 contains a full list of variable definitions. We use the four quantitative performance measures as the main dependent variables in our experiment. Table 11 reports descriptive statistics for our performance measures, player demographics, and post-experiment survey responses.¹³

¹³ An analysis of covariate balance shows that there are no statistically significant differences among the six assigned treatment groups on pre-feedback performance or on demographic variables at the .05 level.

Table 11. Descriptive Statistics

	N	Mean	SD
<i>Performance measures</i>			
<i>Hitrate (pct.)</i>	117	0.76	0.13
<i>Distance (m.)</i>	117	0.8	0.35
<i>Fast_Time (s.)</i>	117	2.27	0.13
<i>Avg_Time (s.)</i>	117	4.3	1.04
<i>Demographic variables</i>			
<i>Age</i>	117	21.44	4.65
<i>Gender</i>	117	0.15	0.36
<i>League level</i>	117	0.22	0.42
<i>% Games entered</i>	117	72.05	31.73
<i>Tenure</i>	117	4.02	3.03
<i>Simulator experience</i>	117	1.79	1
<i>Training</i>	117	5.54	1.09
<i>Height</i>	117	179.49	7.42
<i>Weight</i>	117	72.56	8.95
<i>Post-experiment survey</i>			
<i>Social comparison (standardized mean)</i>	117	0	0.31
<i>Too many measures</i>	117	2.52	1.32
<i>Conscientiousness</i>	117	3.9	0.84
<i>Interest</i>	117	5.93	1.08

Notes: This table shows the descriptive statistics for the measures of performance, demographic variables, and post-experiment survey questions. The performance measures are reported as in round 2, after the delivery of PF. The factor analysis used to compute *Social Comparison* yields a score that is standardized to mean zero. An analysis of covariate balance shows that there are no statistically significant differences among the six assigned treatment groups on pre-feedback performance or on demographic variables at the .05 level. Appendix 9 contains a full list of variable definitions.

3.5 Analysis

We employ OLS regression, with full interaction of treatments, to test for effects. We include controls for player demographics in line with the guidance from econometric research that, in field experiments where randomization is used, “it is customary to control for covariates to correct for chance associations between treatment status and applicant characteristics and to increase precision” (Angrist & Pischke, 2008). OLS has been applied in this manner in prior field experimental research in accounting and economics (Eyring & Narayanan, 2018; Krueger & Whitmore, 2001).

For each analysis, we address performance by using all four of the measures that were reported in the PF. As described earlier, two of the measures reflect accuracy – *Hitrate* and *Distance*, and two of the measures reflect speed – *Avg_Time* and *Fast_Time*.

Table 12 shows the OLS model that represents the performance effects of our 3x2 treatment design. The omitted treatment condition is *Absolute* x *Aggregate*. Thus, the coefficients on *Relative* and on *Absolute&Relative* represent the estimated effect of those types of feedback as compared to the *Absolute* type of feedback. Similarly, the coefficients on *Detail* represent the estimated effect of providing feedback at a detailed level rather than aggregate-only level. For each dependent variable, we run the model with main effects – *Relative*, *Absolute&Relative*, and *Detail* – in the odd-numbered columns, and then with both those main effects and their interactions in the even-numbered columns.

The statistically significant coefficients on *Relative* in Table 12 columns (1), (3), (5), and (7) show that there is an estimated statistically significant performance benefit of providing relative feedback, as compared to providing absolute feedback, for all four dependent variables. In reading these coefficients, note that a decrease in *Avg_Time* or *Fast_Time*, represented by a negative coefficient, reflects an improvement in time-related performance. Similarly, a decrease in *Distance*, also represented by a negative coefficient, reflects an improvement in passing accuracy. The magnitudes of these coefficients are substantial. Providing RPF leads to between a ten percent and 27 percent improvement, depending on the measure, as compared to the performance of players who receive APF alone. These performance effects are at least half of a standard deviation increase for each performance measure. In our setting, participants are highly skilled on the performance dimensions we report and so even a small percent improvement represents significant movement in the performance distribution. The results from this analysis offer support for H1a – that providing RPF leads to better performance than providing APF.

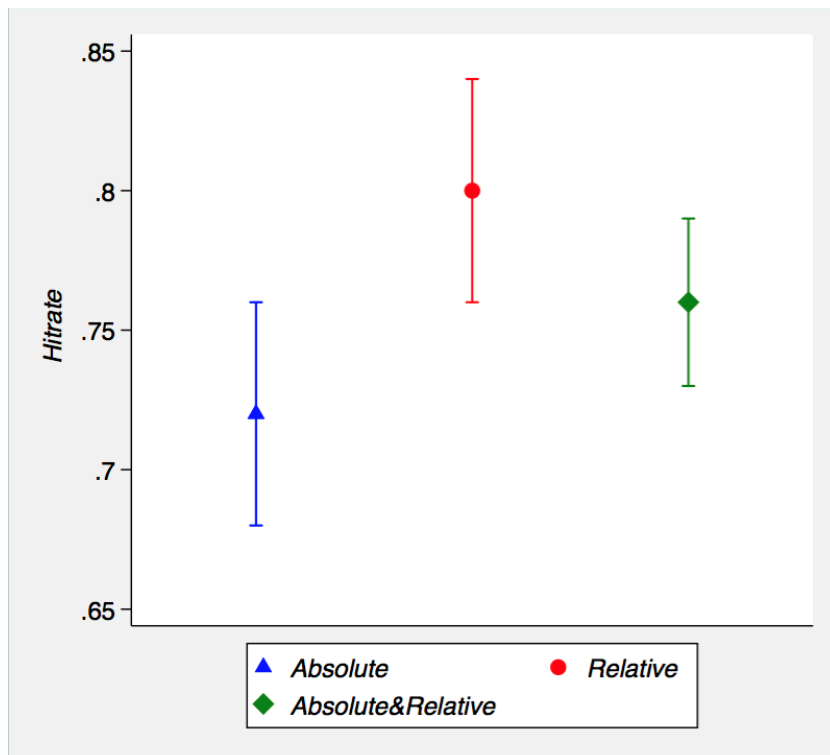
In columns (2), (4), and (6) of Table 12, we use a circle symbol to denote that the effect of providing relative is statistically significantly stronger than the effect of providing combined absolute and relative feedback for the three measures *Hitrates*, *Distance*, and *Avg_Time*. These results offer support for H1b – i.e., that, contrary to standard economic theory, performance is better when PF contains a more limited information set. The coefficient on *Absolute&Relative* in column (7) shows that providing combined absolute and relative feedback has a statistically significantly benefit for performance, as compared to providing absolute feedback, only for the variable *Fast_Time*. In subsequent analyses, we consider the difference between our four performance measures, particularly in terms of whether number-gaming could easily occur, to seek to understand why this effect appears only for one of our four measures. Figures 4-7 show the mean and a 90 percent confidence interval for each performance measure (*Hitrates*, *Distance*, *Average Time*, *Fastest Time*) in round two, after the provision of PF, by assignment to *Absolute*, *Relative*, or *Absolute&Relative* treatment condition.

Table 12. Effect of Relative Information and Detail in Performance Feedback

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	<i>Hitrate</i>	<i>Hitrate</i>	<i>Distance</i>	<i>Distance</i>	<i>Avg_Time</i>	<i>Avg_Time</i>	<i>Fast_Time</i>	<i>Fast_Time</i>
<i>Relative</i>	0.10***,◦ [2.74]	0.11**,◦ [2.32]	-0.25**,◦ [-2.36]	-0.28*,◦ [-1.92]	-0.80***,◦ [-2.90]	-0.84**,◦ [-2.35]	-0.07** [-2.31]	-0.06 [-1.41]
<i>Absolute&Relative</i>	0.04 [1.34]	0.04 [0.78]	-0.11 [-0.99]	-0.14 [-1.05]	-0.44* [-1.73]	-0.38 [-1.06]	-0.10*** [-3.32]	-0.14*** [-2.89]
<i>Detail</i>	-0.01 [-0.26]	-0.01 [-0.14]	0.04 [0.44]	-0.01 [-0.04]	0.05 [0.21]	0.06 [0.17]	0.02 [0.95]	-0.00 [-0.04]
<i>Relative x Detail</i>		-0.02 [-0.33]		0.07 [0.35]		0.10 [0.19]		-0.02 [-0.25]
<i>Absolute&Relative x Detail</i>		0.02 [0.26]		0.07 [0.34]		-0.13 [-0.27]		0.09 [1.48]
<i>Age</i>	-0.00 [-1.23]	-0.00 [-1.17]	0.01 [1.36]	0.01 [1.30]	0.04 [1.28]	0.04 [1.23]	0.00 [0.96]	0.00 [1.05]
<i>Gender</i>	-0.15*** [-2.81]	-0.15*** [-2.68]	0.15 [1.35]	0.15 [1.24]	1.17*** [2.95]	1.15*** [2.83]	0.08** [2.37]	0.09*** [2.62]
<i>Height</i>	-0.00 [-0.78]	-0.00 [-0.68]	-0.01 [-1.12]	-0.01 [-1.14]	0.02 [0.80]	0.02 [0.73]	0.00 [0.39]	0.00 [0.36]
<i>Weight</i>	0.00 [0.11]	0.00 [0.04]	0.01 [0.69]	0.01 [0.72]	-0.01 [-0.35]	-0.01 [-0.30]	-0.00 [-0.47]	-0.00 [-0.46]
<i>League level</i>	0.01 [0.40]	0.01 [0.36]	0.04 [0.32]	0.04 [0.33]	-0.11 [-0.44]	-0.10 [-0.42]	-0.03 [-1.14]	-0.03 [-1.24]
<i>% Games entered</i>	-0.01 [-1.04]	-0.01 [-0.88]	0.02 [1.18]	0.02 [1.09]	0.04 [0.97]	0.04 [0.86]	0.01 [0.88]	0.01 [1.11]
<i>Tenure</i>	0.00 [0.54]	0.00 [0.42]	0.00 [0.67]	0.00 [0.71]	-0.00 [-0.45]	-0.00 [-0.36]	0.00 [0.87]	0.00 [0.60]
<i>Simulator experience FE</i>	yes	yes	yes	yes	yes	yes	yes	yes
<i>Language FE</i>	yes	yes	yes	yes	yes	yes	yes	yes
<i>Training FE</i>	yes	yes	yes	yes	yes	yes	yes	yes
<i>Position FE</i>	yes	yes	yes	yes	yes	yes	yes	yes
<i>N</i>	117	117	117	117	117	117	117	117

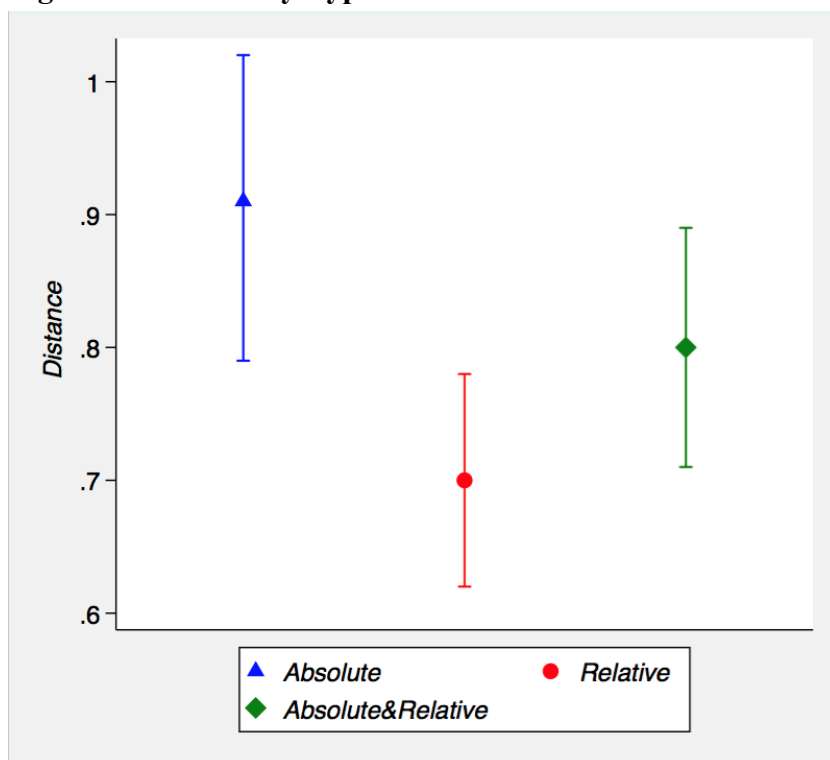
Notes: This table presents estimates of the effect of PF display type and level of detail on the performance measures *Hitrate*, *Distance*, *Avg_Time*, and *Fast_Time*. The coefficient on *Relative* represents the effect of showing only relative as compared to showing only absolute measures in feedback. The coefficient on *Absolute&Relative* represents the effect of showing both absolute and relative measures as compared to showing only absolute measures in feedback. The coefficient on *Detail* represents the effect of showing detailed rather than aggregate feedback. The interaction terms test whether showing relative feedback either alone or with absolute measures is more or less effective, as compared to showing absolute feedback alone, when the feedback is detailed. T-statistics are based on heteroscedastic-robust standard errors and are reported in brackets below each coefficient. *, **, *** denote significance at the .1, .05, and .01 levels, respectively. ◦ denotes significance at least the .1 level as indicated by a z-test for differences in coefficients, indicating that the coefficient on *Relative* is statistically significantly larger than the coefficient on *Absolute&Relative*.

Figure 4. *Hitrate* by Type of Measures in Performance Feedback



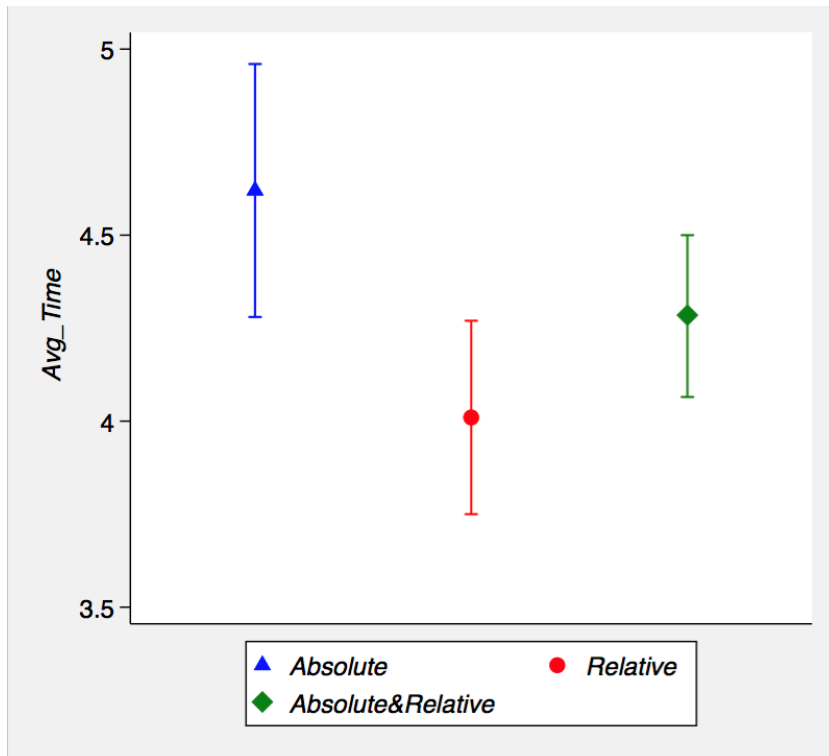
Notes: This figure displays performance on the measure *Hitrate*, or percent of passes that were on target, from the second round of the passing drill after the provision of PF. Performance is presented by the type of measures in feedback – absolute only, relative only, or combined absolute and relative. The symbols represent the mean for each treatment type and the lines represent 90 percent confidence intervals.

Figure 5. *Distance* by Type of Measures in Performance Feedback



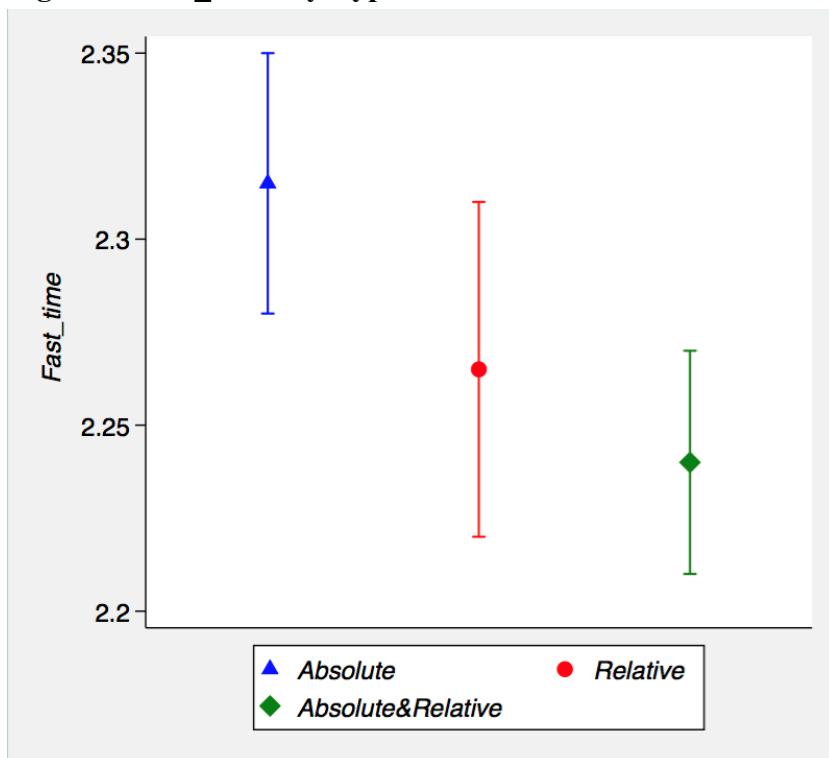
Notes: This figure displays performance on the measure *Distance*, or average distance in meters between passes and the target, from the second round of the passing drill after the provision of PF. Performance is presented by the type of measures in feedback – absolute only, relative only, or combined absolute and relative. The symbols represent the mean for each treatment type and the lines represent 90 percent confidence intervals.

Figure 6. *Avg_Time* by Type of Measures in Performance Feedback



Notes: This figure displays performance on the measure *Avg_Time*, or average time in seconds taken to complete a pass, from the second round of the passing drill after the provision of PF. Performance is presented by the type of measures in feedback – absolute only, relative only, or combined absolute and relative. The symbols represent the mean for each treatment type and the lines represent 90 percent confidence intervals.

Figure 7. *Fast_Time* by Type of Measures in Performance Feedback

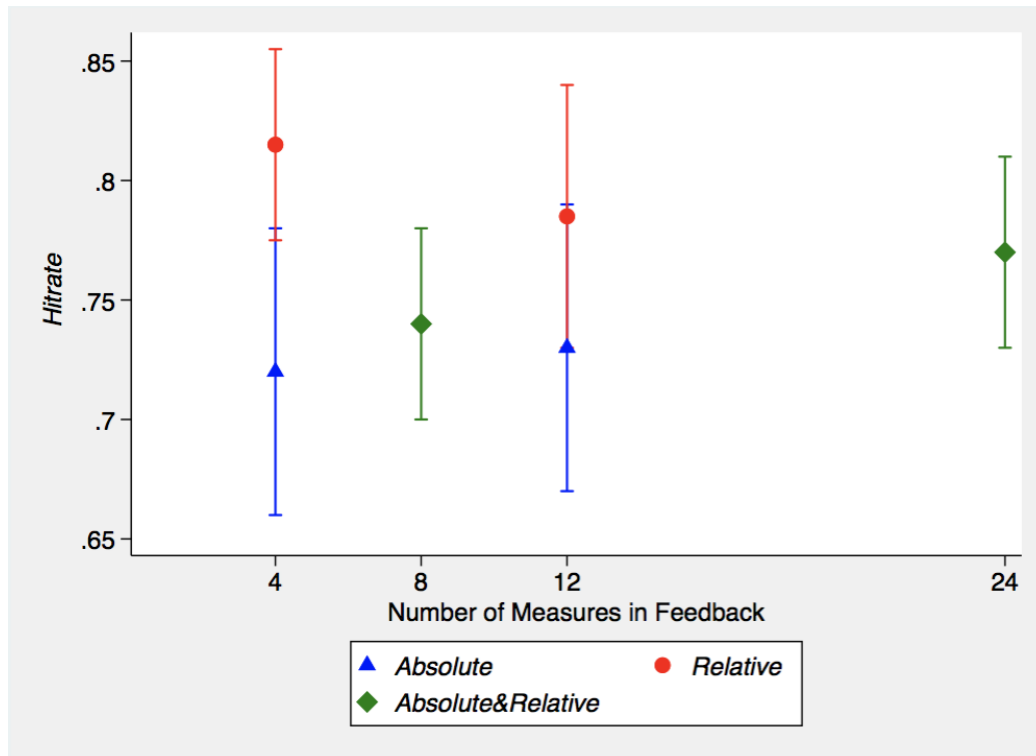


Notes: This figure displays performance on the measure *Fast_Time*, or minimum time in seconds taken to complete a pass, from the second round of the passing drill after the provision of PF. Performance is presented by the type of measures in feedback – absolute only, relative only, or combined absolute and relative. The symbols represent the mean for each treatment type and the lines represent 90 percent confidence intervals.

With regard to the *Detail* condition, the coefficients on this term in all columns of Table 12 document that we find no statistically significant evidence that increasing the number of measures per se while holding constant the scope of PF – relative, absolute, or combined absolute and relative – leads to worse performance. This fails to support H2, and suggests that the benefit of restricting PF to contain relative information alone is not a function of the quantity of measures displayed per se. It is worth noting that, if the quantity of measures were the issue causing *Relative* to perform better than *Absolute&Relative*, then *Detail* would reasonably have a more negative effect than *Absolute&Relative*. This is because *Detail* increases the number of measures by a factor of three (i.e., an average is shown along with its two component parts, rather than the average shown alone), while *Absolute&Relative* increases the number of measures by a factor of only two (i.e., an absolute measure is shown along with its associated rank, rather than either the absolute measure or associated rank alone). Furthermore, if the negative effect of *Absolute&Relative* as compared to *Relative* were driven by information quantity, we might also expect the negative effect to be moderated by *Detail*. The coefficients on the interaction terms in Table 12 do not provide evidence to support this.

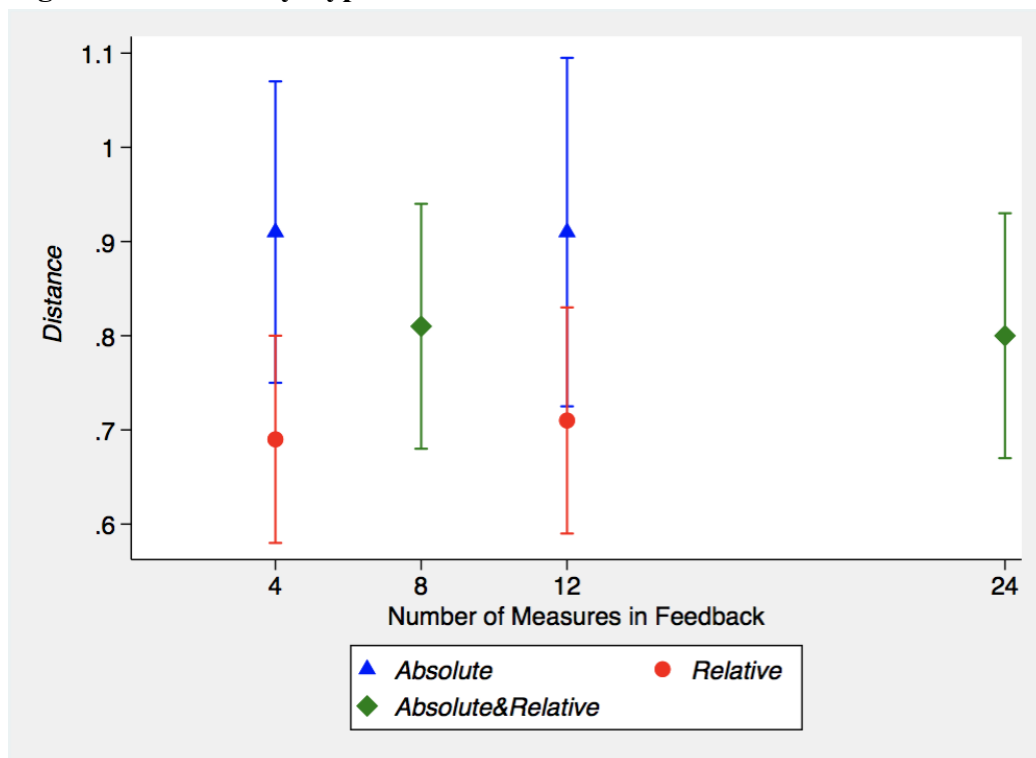
We use Figures 8-11 to demonstrate visually that there is no pattern whereby an increase in the number of measures in PF leads to lower performance. Specifically, for the given performance measure in Figures 8-11, we do not see a negative trend in performance as we move along the x-axis to a larger number of measures included in the PF. Rather, we see differences by color, which represents whether the feedback was *Absolute*, *Relative*, or *Absolute&Relative*, as indicated in the legend. This provides further support for the idea that it is the type of measures included in PF, rather than the number of measures included in PF, that is driving the performance effects that we observe.

Figure 8. *Hitrate* by Type and Number of Measures in Performance Feedback



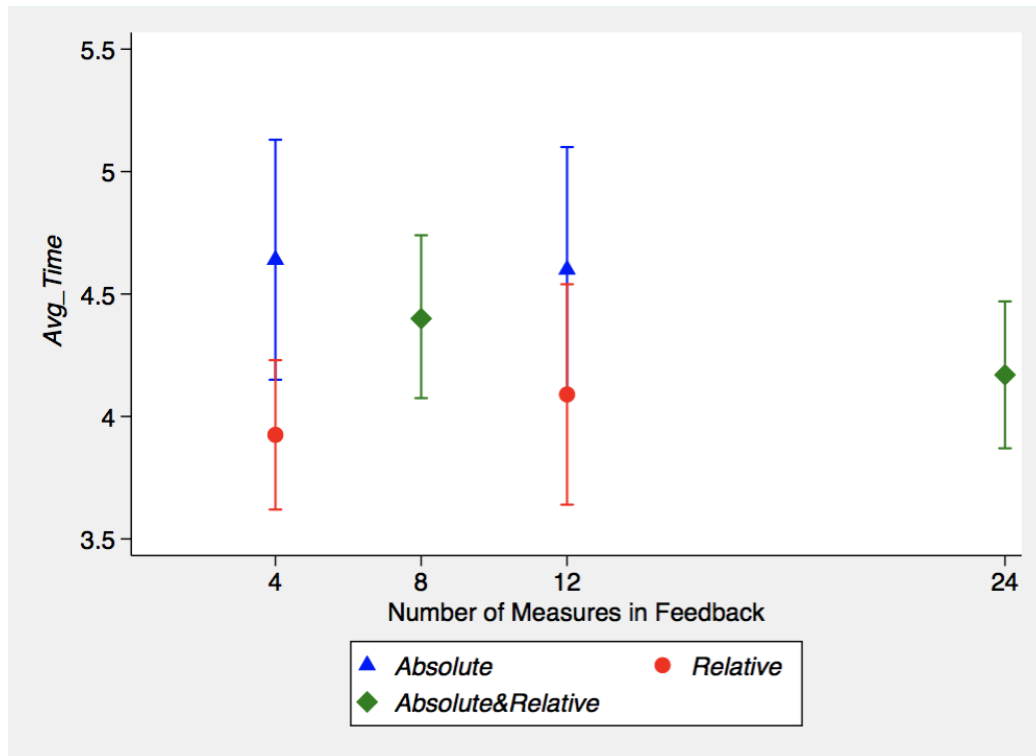
Notes: This figure displays performance on the measure *Hitrate*, or percent of passes that were on target, from the second round of the passing drill after the provision of PF. Performance is presented by the type of measures in feedback – absolute only, relative only, or combined absolute and relative – and the number of measures in feedback. The symbols represent the mean for each treatment type and the lines represent 90 percent confidence intervals.

Figure 9. *Distance* by Type and Number of Measures in Performance Feedback



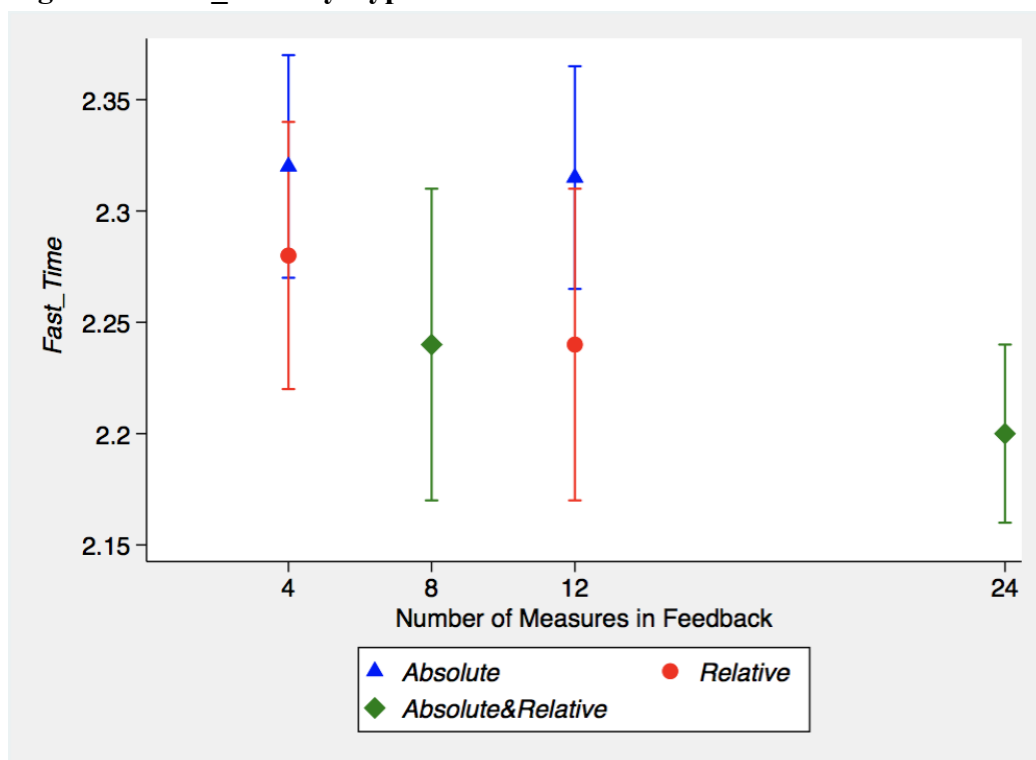
Notes: This figure displays performance on the measure *Distance*, or average distance in meters between passes and the target, from the second round of the passing drill after the provision of PF. Performance is presented by the type of measures in feedback – absolute only, relative only, or combined absolute and relative – and the number of measures in feedback. The symbols represent the mean for each treatment type and the lines represent 90 percent confidence intervals.

Figure 10. *Avg_Time* by Type and Number of Measures in Performance Feedback



Notes: This figure displays performance on the measure *Avg_Time*, or average time in seconds taken to complete a pass, from the second round of the passing drill after the provision of PF. Performance is presented by the type of measures in feedback – absolute only, relative only, or combined absolute and relative – and the number of measures in feedback. The symbols represent the mean for each treatment type and the lines represent 90 percent confidence intervals.

Figure 11. *Fast_Time* by Type and Number of Measures in Performance Feedback



Notes: This figure displays performance on the measure *Fast_Time*, or minimum time in seconds taken to complete a pass, from the second round of the passing drill after the provision of PF. Performance is presented by the type of measures in feedback – absolute only, relative only, or combined absolute and relative – and the number of measures in feedback. The symbols represent the mean for each treatment type and the lines represent 90 percent confidence intervals.

Table 13. Effect of Providing Both Absolute and Relative Performance Information on Reports of Number of Measures Being too Many

	(1)	(2)	(3)
	<i>Too many measures</i>	<i>Too many measures</i>	<i>Too many measures</i>
<i>Absolute&Relative</i>	-0.23 [-0.81]	-0.50 [-1.24]	-0.14 [-0.28]
<i>Age</i>	-0.01 [-0.35]	0.02 [0.30]	-0.00 [-0.04]
<i>Gender</i>	-0.02 [-0.03]	0.01 [0.01]	0.25 [0.34]
<i>Height</i>	-0.06 [-1.35]	0.03 [0.50]	-0.13* [-1.92]
<i>Weight</i>	0.07* [1.90]	0.05 [1.03]	0.11* [1.97]
<i>League level</i>	0.18 [0.43]	-0.30 [-0.50]	0.52 [0.75]
<i>% Games entered</i>	0.01 [1.36]	0.01 [1.46]	0.01 [0.99]
<i>Tenure</i>	0.01 [0.17]	0.00 [0.05]	0.01 [0.05]
<i>Simulator experience FE</i>	yes	yes	yes
<i>Language FE</i>	yes	yes	yes
<i>Training FE</i>	yes	yes	yes
<i>Position FE</i>	yes	yes	yes
<i>Sample</i>	Full	Aggregated information	Detailed information
<i>N</i>	117	59	58

Notes: This table presents estimates of the effect of displaying both absolute and relative measures in PF on post-experiment survey responses regarding participants' perceptions of having received too many measures. The survey responses are on a seven-point Likert scale. The coefficients on *Absolute&Relative* represent the effect of showing both absolute and relative as compared to showing only absolute or relative feedback. T-statistics are based on heteroscedastic-robust standard errors and are reported in brackets below each coefficient. *, **, *** denote significance at the .1, .05, and .01 levels, respectively.

Table 13 uses survey data to further document the lack of evidence in support of the idea that providing both absolute and relative information induces information overload. In the post-experiment survey, reports that the number of measures were ‘too many’ are not statistically significantly affected by whether the information displayed was absolute, relative, or both absolute and relative. Column (1) shows the result for the full sample, and columns (2) and (3) show that the result is similar whether the information is aggregated or detailed. Taken together, the results in Tables 12 and 13 provide evidence that information overload does not explain

why RPF outperforms feedback that contains both absolute and relative performance information.

To address another potential mechanism behind the benefit of restricting information to be in relative terms alone, we test whether providing only RPF increases the strength of social comparison. We follow prior research on RPF and social comparison by conducting a factor analysis of questions regarding social comparison involvement (Tafkov, 2013). We use the resulting factor, *Social Comparison*, as our measure of social comparison involvement. Table 14 shows that there is a statistically significant effect of providing relative information alone on *Social Comparison* as compared to providing either absolute information alone or the combination of absolute and relative information. Table 15 documents that our measure of social comparison is positively and statistically significantly related to performance for all four of our measures. These results provide support for H3a and H3b and suggest that the benefit of restricting information to be relative terms alone yields a performance benefit born of more salient social comparison in this treatment.

Table 14. Effect of Relative Performance Feedback on Social Comparison

	(1)
	<i>Social comparison</i>
<i>Relative</i>	0.25*** [○]
	[3.98]
<i>Absolute&Relative</i>	0.09
	[1.28]
<i>Age</i>	-0.01
	[-1.35]
<i>Gender</i>	0.11
	[1.04]
<i>Height</i>	0.01
	[0.90]
<i>Weight</i>	0.00
	[0.34]
<i>League level</i>	0.04
	[0.56]
<i>% Games entered</i>	0.00
	[0.73]
<i>Tenure</i>	-0.00
	[-0.33]
<i>Simulator experience FE</i>	yes
<i>Language FE</i>	yes
<i>Training FE</i>	yes
<i>Position FE</i>	yes
<i>N</i>	117

Notes: This table presents estimates of the effect of PF display on *Social Comparison*. *Social Comparison* is the result of a factor analysis of questions regarding the degree of social comparison that the participant felt involved in during the experiment. The coefficient on *Relative* represents the effect of showing relative as compared to showing only absolute feedback. The coefficient on *Absolute&Relative* represents the effect of showing both absolute and relative as compared to showing only absolute feedback. T-statistics are based on heteroscedastic-robust standard errors and are reported in brackets below each coefficient. *, **, *** denote significance at the .1, .05, and .01 levels, respectively. [○] denotes significance at least the .1 level as indicated by a z-test for differences in coefficients, indicating that the coefficient on *Relative* is statistically significantly larger than the coefficient on *Absolute&Relative*.

Table 15. Relationship Between Social Comparison and Performance

	(1)	(2)	(3)	(4)
	<i>Hirate</i>	<i>Distance</i>	<i>Avg_Time</i>	<i>Fast_Time</i>
<i>Social comparison</i>	0.06*	-0.14*	-0.49*	-0.07*
	[1.68]	[-1.68]	[-1.70]	[-1.76]
<i>Age</i>	0.00	-0.01	-0.01	-0.00
	[0.51]	[-1.10]	[-0.56]	[-0.22]
<i>Gender</i>	-0.13***	0.26**	1.07***	0.08***
	[-3.27]	[2.60]	[3.48]	[2.62]
<i>Height</i>	0.00	-0.01	-0.01	0.00
	[0.52]	[-1.63]	[-0.40]	[0.07]
<i>Weight</i>	-0.00	0.01*	0.01	-0.00
	[-0.84]	[1.88]	[0.60]	[-0.46]
<i>League level</i>	-0.01	0.05	0.07	-0.01
	[-0.31]	[0.46]	[0.35]	[-0.34]
<i>% Games entered</i>	0.00	0.00	-0.00	0.00
	[0.25]	[1.23]	[-0.15]	[0.29]
<i>Tenure</i>	-0.00	0.00	0.01	0.00
	[-0.22]	[0.19]	[0.15]	[0.11]
<i>Simulator experience FE</i>	yes	yes	yes	yes
<i>Language FE</i>	yes	yes	yes	yes
<i>Training FE</i>	yes	yes	yes	yes
<i>Position FE</i>	yes	yes	yes	yes
N	234	234	234	234

Notes: This table presents estimates of the relationships between *Social Comparison* and the performance measures *Hirate*, *Distance*, *Avg_Time*, and *Fast_Time* measured in rounds 1 and 2 of the passing drill. *Social Comparison* is the result of a factor analysis of questions regarding the degree of social comparison that the participant felt involved in during the experiment. T-statistics are based on standard errors clustered at the participant level and are reported in brackets below each coefficient. *, **, *** denote significance at the .1, .05, and .01 levels, respectively.

To understand drivers and moderators of our effects, we use cross-sectional analysis. Our first two cross-sectional analyses examine why *Absolute&Relative* is able to yield a positive performance effect for *Fast_Time* and not for the other three measures of performance. While the other three measures would be harder to game given that they are averages across a full round of balls in the drill, *Fast_Time* is an extremum and could be gamed by making a quick but loosely accurate pass. The occurrence of gaming may partly explain why we see a positive effect on *Fast_Time* even in the *Absolute&Relative* treatment. Specifically, *Absolute&Relative* produces lower levels of social comparison than *Relative*, but those lower levels may be

sufficient to yield improvement when relatively less effort is required to improve performance by a measure.

To explore whether number gaming may be occurring for *Fast_Time*, we partition our analysis of performance effects by whether a player scored at a medium or lower level of *Conscientiousness* along the Big Five personality traits scale. These players are, as measured by this psychometric index, relatively less diligent and less prone to do a thorough job. If number gaming is occurring for *Fast_Time*, we would expect that the effect of providing relative performance information is stronger *Fast_Time* time among players who are less conscientious. Table 16 column (4) provides evidence to support this idea, showing statistically significant estimated effects of both *Relative* and *Absolute&Relative* that are of larger magnitude among these relatively less conscientious players than they were among the full sample shown in Table 12 column (7). Moreover, the effect sizes of *Relative* and *Absolute&Relative* for the other three performance measures, shown in Table 16 columns (1)-(3), are each smaller among these relatively less conscientious players than they were among the full sample shown in Table 12 columns (1), (3), and (5). This provides some evidence to suggest that our estimated performance effects for measures that require more substantive improvement are partly dependent on the conscientiousness of the players.

Table 16. Effect of Relative Information and Detail in Performance Feedback, Partitioned by Conscientiousness

	(1)	(2)	(3)	(4)
	<i>Hitrates</i>	<i>Distance</i>	<i>Avg_Time</i>	<i>Fast_Time</i>
<i>Relative</i>	0.06 [1.08]	-0.12 [-0.78]	-0.45 [-1.18]	-0.10** [-2.31]
<i>Absolute&Relative</i>	0.01 [0.24]	0.03 [0.17]	-0.22 [-0.57]	-0.15*** [-3.22]
<i>Age</i>	-0.00 [-0.50]	-0.01 [-0.50]	0.02 [0.39]	-0.00 [-0.39]
<i>Gender</i>	-0.24*** [-3.09]	0.20 [1.18]	1.84*** [3.20]	0.09 [1.62]
<i>Height</i>	-0.01 [-1.16]	-0.02 [-1.14]	0.04 [1.15]	-0.00 [-0.26]
<i>Weight</i>	-0.00 [-0.08]	0.02 [1.57]	-0.00 [-0.00]	0.00 [0.79]
<i>League level</i>	0.01 [0.31]	-0.08 [-0.62]	-0.10 [-0.34]	-0.07** [-2.01]
<i>% Games entered</i>	-0.00 [-0.04]	0.00 [0.88]	0.00 [0.17]	0.00 [1.00]
<i>Tenure</i>	-0.00 [-0.26]	0.01 [0.24]	0.01 [0.14]	0.01 [0.86]
<i>Simulator experience FE</i>	yes	yes	yes	yes
<i>Language FE</i>	yes	yes	yes	yes
<i>Training FE</i>	yes	yes	yes	yes
<i>Position FE</i>	yes	yes	yes	yes
<i>Conscientiousness partition</i>	Med-Low	Med-Low	Med-Low	Med-Low
N	72	72	72	72

Notes: This table presents estimates of the effect of PF display on the performance measures *Hitrates*, *Distance*, *Avg_Time*, and *Fast_Time*. The coefficients on *Relative* represent the effect of showing relative as compared to showing absolute feedback. The coefficients on *Absolute&Relative* represent the effect of showing both absolute and relative as compared to showing absolute feedback. The partitioning variable, *Conscientious* is scored on a scale of 1-5 using the Big Five personality test methodology. *Conscientious* represents an individual's tendency towards being thorough and not lazy. T-statistics are based on heteroscedastic-robust standard errors and are reported in brackets below each coefficient. *, **, *** denote significance at the .1, .05, and .01 levels, respectively.

To provide a direct measure of number-gaming in the form of trading off accuracy for speed, we test whether there is a change in the relationship between *Hitrates* and *Fast_Time* after the delivery of RPF. If players begin trading off *Hitrates* by reducing it in order to achieve a reduced *Fast_Time*, we would expect to see a more positive correlation between these two variables in round two than in round one. Table 17 supports this idea, showing that in a regression on *Hitrates*

there is a positive and statistically significant interaction between *Fast_Time* and *Round 2*. This interaction only occurs for the easier-to-game measure of time – *Fast_Time* – and not for *Avg_Time*. Moreover, the interaction does not occur unless the feedback includes relative information. Collectively, these results provide some evidence that the improvement in *Fast_Time* is gamed to a degree, which is a potential reason why *Absolute&Relative* is able to produce performance improvement in the full sample only for *Fast_Time*. As noted earlier, while the *Absolute&Relative* treatment dilutes social comparison when compared to the *Relative* treatment, both may induce sufficient levels of social comparison to lead to improvement if the improvement can be gamed.

Table 17. Evidence of Attempt to Reduce *Fast_Time* by Reducing *Hitrate*

	(1)	(2)	(3)	(4)
	<i>Hitrate</i>	<i>Hitrate</i>	<i>Hitrate</i>	<i>Hitrate</i>
<i>Fast_Time</i>	-0.23 [-1.66]	-0.39* [-1.89]		
<i>Average_Time</i>			-0.13*** [-88.63]	-0.13*** [-40.07]
<i>Round 2</i>	-0.58 [-1.62]	-0.24 [-0.31]	-0.01 [-1.41]	0.00 [0.04]
<i>Fast_Time x Round 2</i>	0.27* [1.72]	0.09 [0.28]		
<i>Average_Time x Round 2</i>			0.00 [0.72]	-0.00 [-0.15]
Sample	<i>Relative or Absolute&Relative</i>	<i>Absolute</i>	<i>Relative or Absolute&Relative</i>	<i>Absolute</i>
N	156	78	156	78

Notes: This table shows results of a test for whether there was a change in the relationship between *Hitrate* and *Fast_Time* or between *Hitrate* and *Average_Time* that occurs between rounds 1 and 2 of the passing drill, after the provision of PF. The coefficient on the interaction terms represents whether there was a change in the correlation after the feedback was provided. T-statistics are based on standard errors clustered at the player level and are reported in brackets below each coefficient. *, **, *** denote significance at the .1, .05, and .01 levels, respectively.

Furthering our cross-sectional tests, we document in Table 18 that the benefits of *Relative* and *Absolute&Relative* vis-à-vis *Absolute* are generally much stronger for players who report a high level of interest in their weaker areas where there is room for improvement. The effects of *Relative* and *Absolute&Relative* are not statistically significantly different for these players.

This suggests that players’ interest in identifying their weaknesses, which rank-based information can reveal, may preserve the salience and effects of relative information even when absolute information is provided in addition.

Table 18. Effect of Relative Information and Detail in Performance Feedback, Partitioned by Interest in Identifying Areas for Improvement

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	<i>Hirate</i>	<i>Hirate</i>	<i>Distance</i>	<i>Distance</i>	<i>Avg_Time</i>	<i>Avg_Time</i>	<i>Fast_Time</i>	<i>Fast_Time</i>
<i>Relative</i>	0.17*** [4.35]	0.07 [1.17]	-0.42** [-2.34]	-0.19 [-1.23]	-1.33*** [-4.32]	-0.56 [-1.31]	-0.03 [-0.54]	-0.06* [-1.72]
<i>Absolute&Relative</i>	0.12*** [2.85]	-0.01 [-0.14]	-0.30* [-2.00]	0.03 [0.19]	-1.02*** [-3.23]	-0.06 [-0.14]	-0.11** [-2.35]	-0.09** [-2.24]
<i>Age</i>	-0.01 [-1.12]	-0.01 [-1.20]	0.01 [0.40]	0.02 [0.85]	0.04 [1.24]	0.05 [1.09]	0.01 [1.37]	-0.00 [-0.25]
<i>Gender</i>	-0.16* [-2.01]	-0.17* [-1.84]	-0.10 [-0.28]	-0.01 [-0.04]	1.39** [2.33]	1.29* [1.89]	0.23*** [2.81]	0.09** [2.05]
<i>Height</i>	0.00 [0.17]	-0.01 [-1.21]	-0.04 [-1.39]	-0.00 [-0.12]	0.00 [0.04]	0.03 [0.91]	0.02** [2.59]	-0.01 [-1.57]
<i>Weight</i>	-0.00 [-0.13]	0.00 [0.25]	0.03 [1.08]	-0.01 [-0.49]	-0.00 [-0.01]	-0.01 [-0.23]	-0.01 [-1.46]	0.00 [0.49]
<i>League level</i>	-0.02 [-0.45]	0.02 [0.32]	0.09 [0.54]	-0.08 [-0.49]	0.17 [0.44]	-0.16 [-0.38]	-0.06 [-1.15]	0.02 [0.48]
<i>% Games entered</i>	0.00 [0.55]	0.00 [1.06]	0.00 [1.24]	-0.00 [-1.09]	-0.00 [-0.67]	-0.00 [-0.89]	-0.00 [-1.12]	0.00 [0.97]
<i>Tenure</i>	-0.02 [-1.65]	-0.00 [-0.13]	0.06** [2.10]	-0.01 [-0.34]	0.12 [1.65]	0.01 [0.12]	-0.00 [-0.25]	0.01 [1.33]
<i>Simulator experience FE</i>	yes	yes	yes	yes	yes	yes	yes	yes
<i>Language FE</i>	yes	yes	yes	yes	yes	yes	yes	yes
<i>Training FE</i>	yes	yes	yes	yes	yes	yes	yes	yes
<i>Position FE</i>	yes	yes	yes	yes	yes	yes	yes	yes
<i>Interest partition</i>	High	Med-Low	High	Med-Low	High	Med-Low	High	Med-Low
N	36	81	36	81	36	81	36	81

Notes: This table presents estimates of the effect of PF display on the performance measures *Hirate*, *Distance*, *Avg_Time*, and *Fast_Time* for samples partitioned by reported interest in identifying areas of performance with room for improvement. The coefficient on *Relative* represents the effect of showing only relative as compared to showing only absolute measures in feedback. The coefficient on *Absolute&Relative* represents the effect of showing both absolute and relative measures as compared to showing only absolute measures in feedback. The coefficient on *Detail* represents the effect of showing detailed rather than aggregate feedback. The interaction terms test whether showing relative feedback either alone or with absolute measures is more or less effective, as compared to showing absolute feedback alone, when the feedback is detailed. The partitioning variable, *Interest* is scored on a scale of 1-7 and indicates the degree of interest in learning about one's weaknesses. T-statistics are based on heteroscedastic-robust standard errors and are reported in brackets below each coefficient. *, **, *** denote significance at the .1, .05, and .01 levels, respectively.

3.6 Conclusion

We use a field experiment in professional sports to address the trade-offs in changing the scope of PF to include absolute information, relative information, or both types of information together. Our analysis furthers research on the design of PF. While PF interventions typically use relative and absolute information either alone or in some combination, prior work has not assigned these conditions in a randomized controlled trial to identify which is most effective.

Contrary to the notion that increased information is more advantageous for performance, we find that restricting PF to relative information alone yields the best effects. Our tests of effect mechanisms indicate that this benefit derives from restricting the scope of information to be in relative terms alone, rather than from restricting the number of displayed measures per se. In particular, restricting PF to relative performance information leads to higher levels of social comparison and associated performance improvement. We contribute most directly to the growing economic and accounting literature on the effective design of PF. Our results also complement recent studies from economics-based literature that show how behavioral considerations can lead to relationships between information, incentives, and performance that are not predicted by traditional economic models.

4 Paper III: Providing Ranks as a Managerial Practice to Reduce Stress: How Performance Feedback Affects Individual Recipients¹⁴

4.1 Introduction

To increase productivity, firms rely on competitive employees but competition triggers individual levels of stress and thereby affects individual behavior and elicits physiological consequences (Zhong et al., 2018). With regards to individual behavior, higher levels of stress have been found to affect decision-making, for example by increasing risk aversion (Cahlíková & Cingl, 2017) or lowering competitiveness overall (Cahlíková et al., 2019, hereafter CCL). Regarding physiological consequences, robust evidence has been established between higher levels of stress and multiple severe diseases such as clinical depression, cardiovascular disease, or HIV/AIDS (Cohen et al., 2007). From a management perspective, each individual case of an employee with stress-related consequences can have large financial and performance implications. Still, the economic implications at a superordinate level are even more severe: In the U.S. alone, five to eight percent of the annual healthcare costs are expected to be associated with workplace stressors (Goh et al., 2016).

Even though recent literature in the field of economics has documented that the adoption of advanced management practices (e.g. PF, target setting or financial incentives) has a robust relationship with firm-level productivity (e.g. Bender et al., 2018) by causally improving the individual productivity of employees (Gosnell et al., 2020), these management practices also represent key stressors for employees in the workplace (Goh et al., 2016). For example, providing detailed PF can lead to higher levels of performance because it provides enough information for individuals to interpret their actions (Casas-Arce et al., 2017), but providing

¹⁴ Koppers, S., Schmidt, S. L. & Schreyer, D. (2020). Providing Ranks as a Managerial Practice to Reduce Stress: How Performance Feedback Affects Individual Recipients. Unpublished Working Paper.

more information than an individual can process can also induce stress and anxiety manifested through physical conditions (Ettema & Zielhuis, 1971; Malhotra, 1982). Therefore, performance-focused managers face a dilemma because the intent of improving organizational productivity through higher levels of competition can be contradicted by the consequences of increasing levels of individual stress. Still, current research is lacking causal evidence about the underlying mechanisms of management practices, specifically for the adoption of PF (Gosnell et al., 2020).

Within the last two decades, the research on PF, i.e. the deliberate presentation of information to improve the effectiveness of specific aspects, has resulted in an extant body of studies (Koppers & Schmidt, 2020). One of the most common characteristics of PF is to foster competition by ranking individual performance via the provision of RPF (Gill et al., 2019). In line with social comparison theory, individuals rely on the display of RPF to compare their performance to their peers' (Festinger, 1954). Providing this specific type of PF has been found to lead to individual performance improvements in general, but also produced no or even negative effects (Schnieder, 2018). Although prior research has suggested that these mixed findings for RPF are due to diverging perceptions by individual recipients (e.g. E. W. Chan, 2018; Eyring & Narayanan, 2018), empirical research has yet to explore how individual differences in processing stress are related to the type of PF received and how the level of stress as an underlying mechanism can potentially contribute to advance the understanding of the mixed results of PF research.

In Eyring et al. (2019, hereafter EFK), the authors introduce two important extensions to the literature on PF. In a field experiment set in professional European football, the design provides individuals with RPF but also includes a second treatment group where individuals receive APF as well as a third treatment group that receives both absolute and relative PF. EFK find that RPF alone is most effective in improving individual performance along multiple performance metrics on speed and accuracy of passing the ball.

In this paper, we extend EFK's findings by investigating whether providing relative, absolute or both absolute and relative PF affects the individual stress level of the feedback recipient as proxied by the individual's heart rate (Buckert et al., 2017; CCL; Halko & Sääksvuori, 2017). We also examine how the perception of PF, trust in the feedback environment as well as individual cognitive traits impact the stress level of the feedback recipient. As such, we provide novel empirical evidence on the individual perception of PF as a management practice by combining experimental data with survey data (Schnieder, 2018). Moreover, we address the current lack of experimental research focusing on person variables in the relationship between PF and behavioral responses (Koppers & Schmidt, 2020) by investigating the mechanism if distinct types of PF affect stress levels experienced by recipients. We aim to narrow this gap since prior research settings have been limited in the ability to account for potentially unintended effects of providing PF (Song et al., 2018).

4.2 Background

We examine the effects of relative versus absolute PF (and both) on individual stress levels in the highly competitive context of professional football in Europe. Professional sports have been identified as a unique laboratory to conduct empirical analysis to investigate the labor market (Kahn, 2000). Hence, sports data has been applied in economics to investigate a variety of research topics in the context of strong incentives (e.g. Massey & Thaler, 2013). Economic incentives are particularly present in the professional football industry since players have naturally occurring incentives to perform both through individual career concerns, based on short career spans with intense competition for roster spots, and self-image concerns, since they derive parts of their identity from being elite performers in the public perception in stadiums, on televisions or on social media (EFK).

Because the industry of professional football is highly suitable to investigate RQs in the field of economics (e.g. Apesteguia & Palacios-Huerta, 2010; Kassis et al., 2017; Kocher et al., 2012)

with relevance for management, we can leverage data from a field experiment to contribute to the understanding of the effects of PF on stress. The choice of setting is especially fitting since many events throughout an individuals' professional career involve competition and stress (CCL) and causal evidence has not been provided for highly skilled workers in developed economies (Gosnell et al., 2020).

4.3 Experimental Design and Data

Our data was collected as part of the pre-registered randomized controlled trial with the AEA RCT Registry (Koppers, 2019). Therefore, this paper builds closely on multiple parts of the EFK paper: the description of the field site skills.lab, a football training simulator, the details on the real-effort task as well as the procedure of the field experiment. Unlike EFK, which focuses on investigating the effects of the type of PF and amount of information on individual performance, we assess the effects of PF on individual stress levels through a diverging set of variables from the pre-registered data set displayed in a transparency table (see Appendix 1).

While taking part in the field experiment, participants played a real-effort task in the football training simulator skills.lab and provided entries to a pre- and post-experimental survey. The real-effort task consisted of a competitive passing drill with 12 balls to be passed to a projected virtual teammate in each round. Participants were randomly assigned to one of the three PF treatments and were provided with either absolute (APF), relative (RPF) or absolute and relative display of PF in between rounds of the passing drill (see EFK and Appendix 2 to Appendix 8 for more details and visualizations of the PF design and treatments). As an additional component of the training, administrators at skills.lab also collected heart rate data with the Garmin HRM-Run heart rate monitor¹⁵ for a subsample of 41 professional football players.¹⁶

¹⁵ The Garmin HRM-Run is a system regularly used at skills.lab and in regular training sessions to monitor physical workload data.

¹⁶ Administrators at skills.lab were not able to collect a larger subsample of participants with heart rate data due to sudden and unforeseeable technological defects of the heart rate measurement system in the collection phase of the field experiment. Therefore, the selection of participants can be considered random.

In line with our quasi-blocked randomization design, the three PF types are evenly distributed.¹⁷ The heart rate data was provided on a ball-by-ball level for each pass with 12 respective heart rate values for a round played. In addition, our data meets the two conditions by CCL to make causal inferences on individual stress.¹⁸

Using this data, we employ OLS regressions with White (1980) standard errors robust to heteroscedasticity to investigate the influence of three PF types on individual stress levels. The dependent variable, mean heart rate difference (*MHRD*), is based on the objectively measured difference in the average heart rate of an individual player between round two and round one of the experimental task, i.e. two rounds of the same training task. Heart rate is an established proxy for individual stress (e.g. CCL) since stress leads to an immediate increase in heart rate (Kirschbaum et al., 1993). We use the difference of heart rates between rounds to account for general individual differences in heart rate and because competitive situations – like football training – can cause stress (Buckert et al., 2017; Buser et al., 2017), so we take the difference between two competitive trainings to investigate the change.

We include six independent variables in three categories. First, we investigate how the type of PF influences individual stress levels. Second, we examine how the individual perception of the feedback recipient, i.e. the perceived value of the feedback and the trust in the simulator measurement, affects individual stress levels. Third, we analyze the effect of individual cognitive traits on stress, i.e. by using the individual cognitive reflection test (*CRT*) score measuring correct answers to three simple math problems that each have an intuitively persuasive but wrong answer (Frederick, 2005). We are specifically interested in the influence of the role individual heuristics might play since prior articles on the effects of PF only rarely

¹⁷ The 41 participants are composed as follows: 14 in the *Absolute* treatment group, 14 in the *Relative* treatment group and 13 in the *Absolute&Relative* treatment group.

¹⁸ CCL (p.2) identify that experimental research needs to address two problems of observational data to make causal inferences on stress through laboratory experiments: (1) to prevent participants from self-selection into competitive (and stressful) situations and (2) to isolate the effects of experimental manipulations to understand the distinctive effects.

account for the influence of cognitive traits (Koppers & Schmidt, 2020) and instead mostly rely on general proxies like ACT or SAT scores used in college admissions test in the United States (e.g. DeShon et al., 2004). We also include several control variables following econometric advice (Angrist & Pischke, 2008) on results from the first round of the experimental task and specific player characteristics. Because we control for speed and accuracy performance metrics, i.e. both *Hitrates* and *Fastest Time*, our results are not biased by the success of the participant in the first round. We also control for the duration of the first round (*Round Duration*) as a proxy for a diverging physical intensity between participants. In addition, we control for several player characteristics, for instance age of the participant or prior simulator experience. Table 19 presents all variables investigated, how they are defined and a rationale for why we include them in our research design.

Table 19. Variable Definition, Measurement and Rationale for use.

Variable	Definition and measurement	Rationale for including variable
Dependent variable		
<i>Mean heart rate difference (MHRD)</i>	Difference between the mean heart rate of the second round of the task minus the mean heart rate of the first round of the task. The mean heart rate per round is based on averaging the heart rate measured for each of the 12 balls played.	Proxy for an individual’s immediate physiological response to a stressor that is related to individual competitiveness (CCL).
Independent variables		
Performance feedback		
<i>Absolute</i>	Indicator variable equal to one if the participant received absolute performance information on passing accuracy and speed of execution.	Three alternative types of PF commonly displayed across a range of organizational settings which influence individual performance (EFK).
<i>Relative</i>	Indicator variable equal to one if the participant received performance ranks relative to a reference group on passing accuracy and speed of execution.	
<i>Absolute&Relative</i>	Indicator variable equal to one if the participant received absolute performance information and performance ranks relative to a reference group on passing accuracy and speed of execution.	
Perception		
<i>Perceived value of feedback</i>	Mean value for responses to three survey questions ($\alpha = 0.80$) taken from Ashford (1986) using a seven-point Likert scale.	Perceiving PF as valuable influences feedback processing (Zingoni & Byron, 2017).
<i>Trust in the simulator</i>	Mean value for responses to three survey questions ($\alpha = 0.78$) taken from Kizilcec (2016) using a seven-point Likert scale.	Individual trust is a key concern in the perception of technology (Kizilcec, 2016).
Cognitive traits		
<i>CRT Score</i>	Number of items answered correctly (Frederick, 2005) based on three updated Cognitive Reflection Test items by Shenhav et al. (2012).	Proxy to identify more intuitive (low score) or more deliberative (high score) individuals’ (Peysakhovich & Rand, 2016).
Controls		
<i>Hitrate</i>	Percentage of 12 balls played in the first round where a participant has hit the target area.	Control for individual performance level in accuracy.
<i>Fast_Time</i>	Fastest time (in seconds) that it took a player to complete a pass among the 12 balls played in the first round.	Control for individual performance level in speed.
<i>Round duration</i>	Total time (in seconds) that it took a player to complete all 12 balls played in the first round.	Control for physical intensity.
<i>% Games entered</i>	Response to survey question, “How many games did you play for your team in the last season (both in the starting line-up and as a substitute)?” using a percentage from 0 % to 100%.	Control for performance level.
<i>Age</i>	Individual age of the participant.	Control for age differences.
<i>Simulator experience</i>	Indicator variable equal to one if the participant has trained in the simulator before.	Control for diverging familiarization with the simulator setting.
<i>Gender</i>	Indicator variable equal to one if the participant is female.	Control for gender differences.

Table 20 presents the descriptive statistics for all variables. The average decrease of *MHRD* between the second and the first round of the task was 3.97 beats per minute (bpm, $SD_{MHRD} = 12.89$). The player specific control variables show that our subsample is composed of young professional players ($Mean(M)_{Age} = 22.34$; $SD_{Age} = 4.45$) with a high baseline of skills since they regularly play in competitive matches ($M\% Games Entered = 64.44$; $SD\% Games Entered = 35.51$).

The majority of participants has trained in the skills.lab simulator before ($M_{\text{Simulator Experience}} = 0.80$; $SD_{\text{Simulator Experience}} = 0.40$) and is predominantly male ($M_{\text{Gender}} = 0.34$; $SD_{\text{Gender}} = 0.48$), even though female players participated as well.

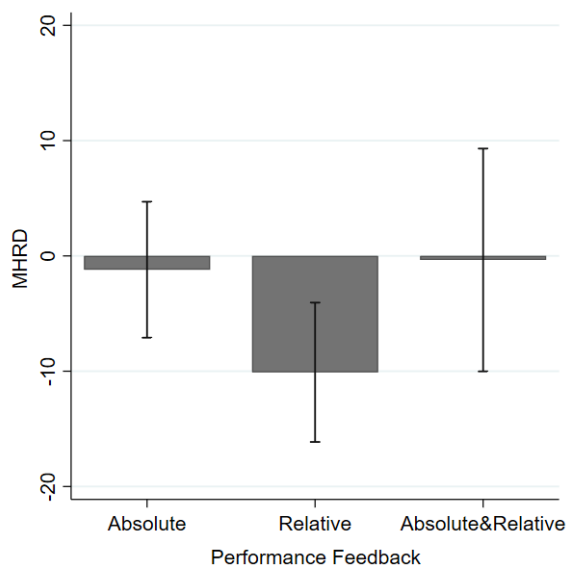
Table 20. Descriptive Statistics

Variable	N	Mean	SD	Min	Max
<i>MHRD</i>	41	-3.97	12.89	-30.25	40.75
<i>Perceived value of feedback</i>	41	5.37	1.42	1.00	7.00
<i>Trust in the simulator</i>	41	5.67	0.87	3.33	7.00
<i>CRT score</i>	41	1.32	1.17	0.00	3.00
<i>Hitrate</i>	41	0.73	0.16	0.17	1.00
<i>Fastest time</i>	41	2.35	0.14	2.13	2.91
<i>Round duration</i>	41	50.87	3.36	45.77	60.99
<i>% Games entered</i>	41	64.44	35.14	0.00	100.00
<i>Age</i>	41	22.34	4.45	16.00	31.00
<i>Simulator experience</i>	41	0.80	0.40	0.00	1.00
<i>Gender</i>	41	0.34	0.48	0.00	1.00

4.4 Results

To assess the effects of PF on individual stress, we start with presenting visual evidence. As Figure 12 shows, providing RPF with ranks (*Relative*) leads to a large decrease in stress. More specifically, the decrease is especially strong compared to the two other types of PF, *Absolute* and *Absolute&Relative* that only lead to a minor decrease in individual stress.

Figure 12. Mean and Confidence Interval of *MHRD*



Notes: This figure visualizes the mean through bars and the 95 percent confidence interval through lines of *MHRD*.

To measure the effects of PF, perception and cognitive traits on stress levels we estimate five OLS models: (1) a PF model including only the three treatment levels of *Absolute*, *Relative* and *Absolute&Relative*, (2) an extended model including all six independent variables of PF, perception and cognitive traits, (3) a PF model additionally containing all variables from the first model with controls, (4) an extended model including all independent variables from the second model as well as controls and (5) a PF model including only the treatment levels of *Relative* and *Absolute&Relative* and all controls. In Table 21 we present the results for each OLS regression model.

Table 21. Determinants of MHRD

	MHRD				
	(1)	(2)	(3)	(4)	(5)
Performance feedback					
<i>Absolute</i>	REF	REF	REF	REF	
<i>Relative</i>	-8.8988** (3.9116)	-8.3999** (3.5200)	-9.3621** (4.4697)	-9.8108** (3.5547)	-10.919* (5.1402)
<i>Absolute&Relative</i>	0.8439 (5.2006)	-1.6929 (4.6437)	1.8420 (5.7340)	-2.2622 (3.8582)	REF
Perception					
<i>Perceived value of feedback</i>		-1.6138 (1.3115)		-2.8323* (1.4503)	-1.5703 (2.2785)
<i>Trust in the simulator</i>		-5.6379*** (1.9923)		-8.4452*** (2.6463)	-9.3462 (5.3238)
Cognitive traits					
<i>CRT score 0</i>		REF		REF	REF
<i>CRT score 1</i>		1.1460 (5.8989)		1.0667 (4.9021)	-8.0475 (7.6444)
<i>CRT score 2</i>		-1.9575 (4.6047)		-3.8477 (4.6217)	-11.286 (10.811)
<i>CRT score 3</i>		8.6374** (4.2256)		7.8465* (4.0271)	-2.7822 (11.883)
Controls					
<i>Hirate</i>			3.7886 (12.340)	-7.8619 (10.984)	-2.8916 (23.395)
<i>Fastest time</i>			2.3811 (20.634)	33.168 (23.473)	8.1999 (29.311)
<i>Round duration</i>			-0.2859 (1.5353)	-1.8532 (1.4247)	-0.6495 (2.1819)
<i>% Games entered</i>			0.1108** (0.05036)	0.1211* (0.06200)	0.1543* (0.08397)
<i>Age</i>			-0.9304 (0.5542)	-1.3143** (0.5182)	-1.1352 (0.9358)
<i>Simulator experience</i>			YES	YES	YES
<i>Gender</i>			YES	YES	YES
<i>N</i>	41	41	41	41	27
<i>R²</i>	0.1208	0.3220	0.2282	0.5244	0.5356

Notes: Robust standard errors in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Overall, the results presented in both Figure 12 and Table 21 present causal evidence that the type of PF affects individual stress. More specifically, we find that providing *Relative* leads to a robust and strong negative effect on *MHRD* compared to providing either *Absolute* or *Absolute&Relative*, i.e. reducing individual stress levels, across all five OLS models. Interestingly, we find no such effect in models (1) to (4) for the PF treatment of *Absolute&Relative*. By using *Absolute&Relative* as the reference group in model (5) to investigate the effects of providing the *Relative* treatment, we find a significant effect.¹⁹ Hence, providing both ranking information and absolute performance metrics together (*Absolute&Relative*) has no significant effect on individual stress compared to providing only *Relative*. This result is particularly noteworthy and consistent with findings by EFK that RPF alone yields the best result on improving individual performance.

The results in model (4) further show that the perception of the individual feedback recipient has a significant effect on *MHRD* when comparing the effects of *Relative* versus *Absolute&Relative* to only *Absolute*. Both increasing levels of the perceived value of feedback and the individual trust in the simulator lead to a decrease in stress when all other variables are held constant. With regards to cognitive traits, the maximum level of the *CRT* score has a significant relation with *MHRD*. Individuals, both in model (2) as well as in model (4), that are more deliberate thinkers (*CRT* score of 3) and are more prone against relying on heuristics have higher stress levels than intuitive thinkers after receiving the PF. Interestingly, perception and cognitive traits of the feedback recipient have no significant effect in model (5) to explain the effects of comparing only *Relative* versus *Absolute&Relative*. We therefore conclude that the effects of both perception and cognitive traits are related to the inclusion of ranks in contrast to only providing absolute values, both isolated or in combination with absolute values. In total,

¹⁹ We also estimated an OLS-model to test if providing *Absolute* versus the *Absolute&Relative* treatment level has an effect on *MHRD* without significant results ($p = 0.629$).

we provide further evidence that the type of information provided in PF does not only matter for performance but also for recipients' stress levels.

4.5 Conclusion

In this paper, we present analytic results based on unique field experimental data to investigate the effects of PF on individual recipients' stress levels. As a result, our paper strives to advance the understanding of PF as an established management practice. In contrast to prior studies on PF investigating the effects on performance, we focus on three categories of person variables (namely *perceived value of PF*, *trust in the simulator* and *CRT* scores) to provide additional evidence for underlying PF mechanisms and reveal three findings.

First, we find that RPF – in the form of providing ranks – has a strong and negative effect on the stress level of the participant, in a way that participants in this treatment group experience significantly less stress. On the contrary, providing APF or both absolute and relative PF has no statistically significant effect on stress. These results complement recent findings by EFK that not all types of PFs are equal and that RPF is most effective in improving individual performance. In line with EFK's findings, these results on stress might be explained by the salience of social comparison in RPF. In this treatment, RPF is assumed to be the salient "top of the head-phenomena" (Taylor & Fiske, 1978, p. 252) that is available to individuals instantly without having its effect weakened by the presence of additional (absolute) informational cues (Hirshleifer & Teoh, 2003). Since social comparison theory predicts that individuals shape their self-image based on comparisons of performance with others (Suls & Wheeler, 2000), RPF might decrease stress levels since it allows individuals to evaluate abilities and to seek positive feelings (R. H. Smith, 2000). This finding is relevant since CCL find that stress decreases competitiveness (which is related to performance in their study) and our evidence suggests that RPF both increases performance (EFK) and simultaneously decreases stress levels.

Second, we show that the recipients' perception has a negative relationship with stress when comparing APF to RPF. This finding holds for both perceived value of feedback and trust in the feedback environment, here a training simulator, in a way that higher levels of perceived value and trust decrease stress. While Zingoni and Byron (2017) identify a positive relationship between perceived value of feedback with effort and learning in a laboratory experiment, our paper provides additional evidence for the positive individual effects of perceiving feedback as valuable. Because prior studies which find that designing technologies that promote users' trust is related to the type of information being provided (Kizilcec, 2016), our results further complement this stream of literature by fostering a better understanding of how the level of trust also affects individual stress levels.

Finally, individual cognitive traits have a positive relationship with stress when comparing APF to RPF. More deliberate thinkers experience significantly higher levels of stress compared to recipients who rather stop thinking to follow an available heuristic response. This finding provides additional evidence for recent findings that differentiating between more intuitive and more deliberative individuals with regards to applying heuristics is important for understanding changes of preference (Peysakhovich & Rand, 2016). Interestingly, recipients' perceptions and individual cognitive traits only have a significant effect on stress when all three PF types are included in the OLS model. Both categories of variables are not significantly different when comparing only RPF to absolute and relative PF. This finding further supports EFK's findings that the type of information in PF design choices is of high importance for researchers and business practitioners alike.

This paper provides novel insights for the effective application of RPF as a management practice in organizations. Our results show that RPF can potentially not only improve employee performance but also reduce stress levels of the workforce. Furthermore, we provide insights that highlight the importance of adapting PF design choices to the individual recipient since we quantify the relationship between PF recipients' stress levels for specific types of PF and

identify which additional person variables, namely perception and cognitive traits, have an effect on individual stress. Overall, these results provide organizations with a range of variables which should be considered and measured on an individual basis to enable managers to estimate the performance and stress-related implications of different types of PF.

In line with prior studies relying on individual level data from professional football (e.g. Kassis et al., 2017), our paper is certainly limited by a relatively small sample size due to the intricate nature of data collected and a smaller share of participating female players. To address these limitations, future research should explore the effects of PF, perception, and cognitive traits on stress through larger data sets, also in other environments, and with a larger share of women. Future research should, for example, measure individuals' perception of salience to provide evidence for the hypothesis that a reduction of salience might explain the different effects.

5 Conclusion

5.1 Summary of Findings

In this dissertation, I provided three research papers to advance our understanding on *what* type of PF information to choose and *how* to display them effectively based on empirical evidence. Therefore, three distinctive, yet interrelated RQs were addressed which contribute to specific research gaps in PF research:

RQ I: What is the state of knowledge on the effects of PF based on empirical evidence in the three management-related disciplines of accounting, management general and organizational behavior and which definitions, theories and research designs are applied?

RQ II: What type of display of PF information – absolute, relative, or absolute and relative – and which quantity of information works best to motivate individuals and improve performance?

RQ III: How does the type of display of PF information – absolute, relative, or absolute and relative – as well as individual person variables affect stress levels of individual PF recipients?

These three RQs were investigated through the course of three stand-alone papers. The following summary of findings strives to provide a condensed synthesis of the findings from each research of the three papers of this dissertation.

In Paper I, a systematic review of 87 quantitative studies on the effects of PF, published in the highest-ranked management-related journals in three disciplines, allowed me to conclude that the overall PF literature is still fragmented. The current state of knowledge further perpetuates a pre-existing ambiguity in the application of PF as a management practice and which underlying mechanisms cause which types of effects.

The overall finding of heterogeneity of the literature is based on the synthesis of three main findings. First, I found that there has not been a research debate in empirical articles or review articles on how to define the key term of PF in a concise way so that interdisciplinary studies can map its findings to a consolidating general definition of PF. Currently, articles mostly refrain from defining PF and from linking it back to the concept of PF, but rather define the specific PF variables and environment they investigate. Therefore, I proposed an integrative PF definition that builds upon recent conceptual PF frameworks (e.g. Schnieder, 2018) which intends to align current synonyms and allow researchers to map specific PF studies to parts of that definition: *PF is the deliberate presentation of information, with varying characteristics, on the task performance within a specific environment to recipients (individuals or groups), to allow them to assess and improve the effectiveness of specific aspects of their future performance.*

Second, I found that the current body of theories being investigated is still as incoherent as Kluger and DeNisi (1996) stated 24 years ago. Even though the specific split of theories being investigated slightly varies by discipline, the majority of articles (58 percent) relies on theories which are only applied less than three times across all articles or on exploratory theory sections. Exploratory theory sections combine multiple theories or make no predictions about direction of hypothesized effects due to the “lack of definitive theory” (Loftus & Tanlu, 2018, p. 278). Only social comparison theory is investigated in a double-digit percentage (16 percentage) and produces a consistent stream of research with cross-citations that refine its predictive power.

Third, I found that the body of articles across disciplines still produces either positive effects (e.g. Eyring & Narayanan, 2018), no effects (e.g. Lourenço, 2016) or negative effects (e.g. E. W. Chan, 2018). Therefore, I summarized the findings of each article and found a current lack of cross-fertilization between disciplines, which I traced back to the immanent nature of the three disciplines and the heterogenous use of PF definitions and key terms. With regards to the research designs applied, I synthesized the elements of the suggested integrative definition for

each discipline and found a common lack of investigating person variables, for instance personality traits or cognitive abilities/traits, in PF research designs. While I found that empirical evidence on the effects of PF across all three disciplines is mainly based on laboratory experiments (80 percent) with student samples (74 percent), the type of incentivization in these studies varies largely between disciplines. In general, the findings of Paper I suggest that additional, integrative research is required to advance the understanding of specific PF variables to enable organizations to apply PF as an effective management practice in heterogenous settings.

In Paper II, I built upon the findings of Paper I by providing causal, field experimental evidence that the PF design choice of *what* type of PF information to display and *how* to do so effectively has a distinct effect on performance. Using a field experiment in professional football, I found that presenting RPF works better than APF to motivate individuals to improve performance. While this finding is in line with the hypothesized direction of the effect based on social comparison theory (Festinger, 1954; Suls & Wheeler, 2000), I also found, across a range of measures, that a subset of information (RPF) alone is more effective to increase performance than the full set of information (absolute and relative PF) – contrary to Bayesian economic theory that predicts that more information is better for performance. The performance effect of providing different types of PF are substantial: players who receive only RPF are almost 14 percent more precise in completing passes and finish the passing drill 17 percent faster.

The fully-interacted 3x2 design, including both three types of PF information (absolute, relative, or both absolute and relative) as well as two dimensions of the number of measures (aggregate, detail), enabled me to show that the performance improvement is not caused by restricting the number of measures in itself, since I do not find that adding more measures – while holding the type of PF constant – leads to a decrease in performance. Furthermore, participants in the absolute and relative treatment did not report having received ‘too much’ information in the post-survey of the field experiment. These findings suggest that the results

are not driven by information overload. Instead, I used additional survey evidence to show that limiting the scope of PF to relative measures makes social comparison more salient. Thus, behavioral economic forces make it optimal to focus attention on specific relative information and to omit absolute information. To address the lack of accounting for person variables when investigating the performance effects of PF, I conducted cross-sectional tests and found that individuals with lower levels of conscientiousness, a Big Five personality trait associated with diligence and thoroughness (Judge & Zapata, 2015), have weaker performance improvements based on the provision of RPF. In addition, I found that participants who are highly interested in learning about areas for improvement have stronger performance benefits from RPF, but also less of a difference in effects on performance between RPF alone and absolute and relative PF. This cross-sectional finding suggests that the salience of relative information may diminish less if people are more prone to pay attention to the available information types.

Paper III further builds upon Paper I and extends Paper II by exploring the effects of the type of PF information on individual recipients' stress levels. Using unique heart rate data, I presented causal evidence that participants who are only provided with ranks (RPF) experienced significantly less stress than participants who receive APF. In addition, I found that providing absolute and relative PF, that is the full set of PF information, has no statistically significant effect on stress. This result complements the findings of Paper II and further supports the line of reasoning that RPF alone emphasizes ranks as salient "top of the head-phenomena" (Taylor & Fiske, 1978, p. 252) for social comparison which then allows individuals to evaluate abilities and derive positive feelings (R. H. Smith, 2000). Additionally, I found strong support that the PF recipients' perception of the PF display and the PF environment has a negative relationship with stress. Based on survey evidence, I showed that higher levels of perceived value of PF and trust in the simulator decrease stress levels. This finding emphasizes the importance of *how* to display PF effectively depends on a holistic assessment of the PF setting. Finally, I specifically tested for the effects of individual cognitive traits, based on the cognitive reflection test as a

proxy for the tendency to apply intuitive and potentially flawed thinking (Peysakhovich & Rand, 2016), and found support for the existence of an effect on stress levels. The findings suggest that deliberate thinkers, who are prone to override heuristic responses, experience significantly higher levels of stress than individuals who are prone to rely on heuristics. This cross-sectional result provided evidence for the findings from Paper I, that integrating cognitive abilities or traits into PF research designs is necessary to advance the current understanding.

Overall, the findings from all three papers contribute to the PF literature by providing answers to all three RQs of this dissertation. First, Paper I demonstrates that the current state of knowledge on the effects of PF derived from empirical evidence in management-related disciplines is heterogenous, based on a mostly non-consecutive stream of theories and definitions and conducted through an overemphasis of standard laboratory experiments with limited external validity. Second, Paper II parses out that the choice of PF information matters to motivate individual performance through social comparison, that is, a smaller number of more salient RPF information is at least as or even more effective than a larger number of RPF information and the combination of both relative and absolute PF information. Finally, Paper III establishes that the type of PF information affects stress levels of PF recipients in a way that RPF information leads to a decrease of the individual heart rate, and that the individually perceived value of PF and trust in the PF environment reduces stress levels.

5.2 Theoretical and Practical Implications

5.2.1 Theoretical Implications

The novel findings of this dissertation make several theoretical contributions to the growing stream of literature which provides empirical evidence on the effects of PF as a managerial practice (e.g. Azmat et al., 2019; Gill et al., 2019; Hannan et al., 2019).

First, they provide the methodological and theoretical basis for future research which aims to quantify the relationship between specific PF variables and the causal effects on individual

performance and reactions. Through systematically exploring how PF is currently defined, how different variables types of PF variables have been investigated, which theories have been applied and which research designs have been conducted, I acknowledge the complexity of the PF process but make the current state of knowledge transparent, so that future research can derive research designs to investigate the effects of PF and the underlying mechanisms, specifically with regards to the individual recipient (e.g. personality traits or cognitive traits). Second, the findings provide evidence that the type of PF measure matters and that PF design choices affect individual performance as well as stress levels. While prior research has mostly focused on comparing the performance effects of different applications of RPF (Kachelmeier, 2019), I find that limiting PF to relative information displayed on few measures motivates performance improvements and reduces stress levels. By considering cross-sectional results on Big Five personality traits and cognitive traits from the CRT, I find support that these person variables can explain underlying mechanisms of RPF which have not been tested (Schnieder, 2018). Third, my findings have important implications for theories that should be considered to predict the effects of how individuals respond to PF. While the findings are consistent with prior PF literature in describing the predictive power of social comparison to explain the effects of RPF (e.g. Tafkov, 2013), I also provide evidence for the behavioral economic assumption that the salience of PF information predicts the positive PF effects of limiting the amount of PF (Casas-Arce et al., 2017) and show that this effect is not caused by information overload.

5.2.2 Practical Implications

The findings of this dissertation also have practical implications for applying PF as a managerial practice by providing evidence for *what* type of information to choose and *how* to display them effectively within the new nature of work that organizations operate in (Wegman et al., 2018).

First, the findings suggest that managers should keep ranks as the preferred type of PF information, contrary to highly publicized and recent decisions by large companies to drop this type of PF information (Cappelli & Tavis, 2016; Mercer, 2019). By relying on RPF, managers can improve performance and cater to the immanent need for development of the modern workforce and simultaneously reduce stress levels which allows employees to stay engaged and productive while it prevents them from suffering severe physiological consequences to their health. Second, the findings from this dissertation should encourage managers to make conscious decisions about the amount of measures included in PF systems in the midst of rapid technological advancements that allow to track and display multiple PF information in real-time. As such, managers should generally emphasize RPF without diluting the salience through adding additional types of information, since the type of PF information drives the performance effects I observed. Even though the findings demonstrate that adding performance measures per se does not lead to an increase or decrease in performance, managers should rather focus on understanding the specific effects of the types of PF information contained within their PF system than including all available measures. Third, the findings emphasize the importance for managers to consider the PF recipients within their organization to adapt their PF designs. Because the cross-sectional findings show that, among others, personality traits, cognitive traits, interest in improving performance, perceived value of PF and trust in the PF system affect how PF motivates performance or reduces stress levels, managers should gather survey data or other assessments from their employees to individualize the PF design choices to their specifications – within their organizational setting – to heighten motivation and improve performance.

5.3 Future Research Directions

While the findings of this dissertation contribute to the understanding of the effects of PF as a managerial practice, three limitations of the three stand-alone papers on the type of review

analysis, the experimental design and the field setting need to be acknowledged and offer promising avenues for future research on the effects of PF.

First, future reviews on PF should perform meta-analytical approaches to calculate an overall effect size of specific PF variables. In order to perform a meta-analysis, studies have to be similar in their methodological designs and share statistical measures (DerSimonian & Laird, 1986; Glass, 1976). As the findings of the systematic literature review in Paper I have shown, the current state of empirical PF research is fragmented and therefore impedes meta-analytical research designs, especially across management-related disciplines. To cope with this current state of research, the systematic literature review in Paper I allowed to identify the empirical evidence on the effects of PF in the management-related literature based on a qualitative and descriptive analysis. Because systematic literature reviews have been underrepresented in managerial research and the analytical strategy is effective in informing research and practice (Snyder, 2019; Tranfield et al., 2003), I hope that the synthesis of PF research in Paper I and recent literature reviews on specific types of PF (e.g. Lechermeier & Fassnacht, 2018; Schnieder, 2018) facilitate the methodological alignment of more homogenous management-related research on PF by using comparable variables and methodologies to make meta-analyses more feasible in the future.

Second, future quantitative research on the effects of PF should build upon the experimental design from Paper II and Paper III by including additional variables to replicate and extend upon my findings based on salience and social comparison theory. Potential experimental iterations could therefore include to make the APF, RPF and both absolute and relative PF treatments either public or private (Song et al., 2018), to display them as either re-start or consecutive PF displays over multiple sessions (Hannan et al., 2019) or to include different reference points for high or low performers (Eyring & Narayanan, 2018).

Third, future empirical research on the effects of PF can consider investigating different field settings and thereby samples, potentially also with longitudinal data to investigate long-term

effects of PF. The experimental evidence for Paper II and Paper III was observed during a single training session without additional data from longer periods of time or from in-game performance and team results. Replication studies that consider longer time horizons of PF treatments and multiple PF frequencies can provide evidence if the findings in my dissertation are long-lasting, and to which extent (Casas-Arce et al., 2017). In addition, the field setting utilizes precise and objective ranks. Research in other settings should examine if displaying subjective ranks in RPF leads to the same results, maybe by extending upon the findings from Paper III on the influence of trust in the PF environment. While the field site was able to include the Big Five personality traits and the CRT on cognitive traits during the limited available time for each training session, future research should add person variables and cognitive tests to provide evidence for the underlying mechanisms of PF effects (Schnieder, 2018). Because the sample in Paper II and III is from football clubs that already trained at the field-site or from clubs which already were in prior talks about using the simulator, I was only able to include one female team in the sample. Since prior research has shown that gender differences exist in performance and tournament entries in RPF environments (Wozniak, 2012), with regards to responses to wins and losses (Gill & Prowse, 2014) and also in career success in predominantly male environments (Huntington-Klein & Rose, 2018), future studies should especially investigate gender effects with regards to the findings of Paper II and Paper III. To replicate the findings in non-football settings and to confirm the external validity for different office settings in organizations, future research should also build upon the large variety of field settings to confirm my findings – e.g. through continuing the diversity of settings from prior PF research on university students (Bandiera et al., 2015), fruit pickers (Bandiera et al., 2013), home repair professionals (Casas-Arce et al., 2017) or marathon runners (Allen et al., 2017).

Overall, I would welcome to see the continuous growth in the management-related stream of quantitative PF research to be maintained by continuing to adapt the research designs and methodologies to the changing nature of work while advancing the understanding of the core

theories and underlying mechanisms of PF effects. By doing so, researchers can provide actionable empirical evidence to guide managers in implementing PF designs as a management practice (Gosnell et al., 2020) to optimize the positive impact of PF on performance by selecting the right information and displaying them effectively.

References

- Adomavicius, G., Curley, S. P., Gupta, A., & Sanyal, P. (2012). Effect of information feedback on bidder behavior in continuous combinatorial auctions. *Management Science*, *58*(4), 811–830. <https://doi.org/10.1287/mnsc.1110.1443>
- Alder, G. S., & Ambrose, M. L. (2005). An examination of the effect of computerized performance monitoring feedback on monitoring fairness, performance, and satisfaction. *Organizational Behavior and Human Decision Processes*, *97*(2), 161–177. <https://doi.org/10.1016/j.obhdp.2005.03.003>
- Allcott, H. (2011). Social norms and energy conservation. *Journal of Public Economics*, *95*(9–10), 1082–1095. <https://doi.org/10.1016/j.jpubeco.2011.03.003>
- Allen, E. J., Dechow, P. M., Pope, D. G., & Wu, G. (2017). Reference-Dependent Preferences: Evidence from Marathon Runners. *Management Science*, *63*(6), 1657–1672. <https://doi.org/10.1287/mnsc.2015.2417>
- Alvero, A. M., Bucklin, B. R., & Austin, J. (2001). An Objective Review of the Effectiveness and Essential Characteristics of Performance Feedback in Organizational Settings (1985–1998). *Journal of Organizational Behavior Management*, *21*(1), 3–29. https://doi.org/10.1300/J075v21n01_02
- Andiola, L. 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. <https://doi.org/10.1016/j.acclit.2014.07.001>
- Angrist, J. D., & Pischke, J.-S. (2008). *Mostly harmless econometrics: An empiricist's companion*. Princeton university press.
- Anseel, F., Beatty, A. S., Shen, W., Lievens, F., & Sackett, P. R. (2015). How Are We Doing After 30 Years? A Meta-Analytic Review of the Antecedents and Outcomes of Feedback-Seeking Behavior. *Journal of Management*, *41*(1), 318–348. <https://doi.org/10.1177/0149206313484521>
- Anseel, F., Lievens, F., & Schollaert, E. (2009). Reflection as a strategy to enhance task performance after feedback. *Organizational Behavior and Human Decision Processes*, *110*(1), 23–35. <https://doi.org/10.1016/j.obhdp.2009.05.003>
- Apestequia, J., & Palacios-Huerta, I. (2010). Psychological Pressure in Competitive Environments: Evidence from a Randomized Natural Experiment. *American Economic Review*, *100*(5), 2548–2564. <https://doi.org/10.1257/aer.100.5.2548>
- Arps, G. F. (1920). Work with knowledge of results versus work without knowledge of results: Awareness and partial awareness as factors conditioning efficiency. *Psychological Monographs*, *28*(3), i–41. <https://doi.org/10.1037/h0093152>
- Arrow, K. J., Blackwell, D., & Girshick, M. A. (1949). Bayes and Minimax Solutions of Sequential Decision Problems. *Econometrica*, *17*(3), 213–244. <https://doi.org/10.2307/1905525>
- Ashford, S. J. (1986). Feedback-Seeking in Individual Adaptation: A Resource Perspective. *Academy of Management Journal*, *29*(3), 465–487. <https://doi.org/10.5465/256219>
- Ashford, S. J., Blatt, R., & VandeWalle, D. (2003). Reflections on the looking glass: A review of research on feedback-seeking behavior organizations. *Journal of Management*, *29*(6),

- 773–799. [https://doi.org/10.1016/S0149-2063\(03\)00079-5](https://doi.org/10.1016/S0149-2063(03)00079-5)
- Ashford, S. J., & Cummings, L. L. (1983). Feedback as an individual resource: Personal strategies of creating information. *Organizational Behavior and Human Performance*, 32(3), 370–398. [https://doi.org/10.1016/0030-5073\(83\)90156-3](https://doi.org/10.1016/0030-5073(83)90156-3)
- Ashraf, A. (2018). *Do Performance Ranks Increase Productivity? Evidence from a Field Experiment*. [https://doi.org/10.1016/S0925-5214\(98\)00091-X](https://doi.org/10.1016/S0925-5214(98)00091-X)
- Atkins, P. W. B., Wood, R. E., & Rutgers, P. J. (2002). The effects of feedback format on dynamic decision making. *Organizational Behavior and Human Decision Processes*, 88(2), 587–604. [https://doi.org/10.1016/S0749-5978\(02\)00002-X](https://doi.org/10.1016/S0749-5978(02)00002-X)
- Azmat, G., Bagues, M., Cabrales, A., & Iriberry, N. (2019). What You Don't Know...Can't Hurt You? A Natural Field Experiment on Relative Performance Feedback in Higher Education. *Management Science*, 65(8), 3714–3736. <https://doi.org/10.1287/mnsc.2018.3131>
- Bachrach, D. G., Bendoly, E., & Podsakoff, P. M. (2001). Attributions of the “causes” of group performance as an alternative explanation of the relationship between organizational citizenship behavior and organizational performance. *Journal of Applied Psychology*, 86(6), 1285–1293. <https://doi.org/10.1037/0021-9010.86.6.1285>
- Balafoutas, L., Chowdhury, S. M., & Plessner, H. (2019). Applications of sports data to study decision making. *Journal of Economic Psychology*. <https://doi.org/10.1016/j.joep.2019.02.009>
- Balcazar, F., Hopkins, B. L., & Suarez, Y. (1985). A Critical, Objective Review of Performance Feedback. *Journal of Organizational Behavior Management*, 7(3–4), 65–89. https://doi.org/10.1300/J075v07n03_05
- Bandiera, O., Barankay, I., & Rasul, I. (2013). Team Incentives: Evidence From A Firm Level Experiment. *Journal of the European Economic Association*, 11(5), 1079–1114. <https://doi.org/10.1111/jeea.12028>
- Bandiera, O., Larcinese, V., & Rasul, I. (2015). Blissful ignorance? A natural experiment on the effect of feedback on students' performance. *Labour Economics*, 34, 13–25. <https://doi.org/10.1016/j.labeco.2015.02.002>
- Barber, B. M., & Odean, T. (2008). All That Glitters: The Effect of Attention and News on the Buying Behavior of Individual and Institutional Investors. *Review of Financial Studies*, 21(2), 785–818. <https://doi.org/10.1093/rfs/hhm079>
- Barkan, R. (2002). Using a signal detection safety model to simulate managerial expectations and supervisory feedback. *Organizational Behavior and Human Decision Processes*, 89(2), 1005–1031. [https://doi.org/10.1016/S0749-5978\(02\)00007-9](https://doi.org/10.1016/S0749-5978(02)00007-9)
- Barley, S. R. (2006). When I Write My Masterpiece: Thoughts On What Makes A Paper Interesting. *Academy of Management Journal*, 49(1), 16–20. <https://doi.org/10.5465/amj.2006.20785495>
- Bartol, K. M., Durham, C. C., & Poon, J. M. L. (2001). Influence of performance evaluation rating segmentation on motivation and fairness perceptions. *Journal of Applied Psychology*, 86(6), 1106–1119. <https://doi.org/10.1037/0021-9010.86.6.1106>
- Baumeister, R. F., & Heatherton, T. F. (1996). Self-Regulation Failure: An Overview. *Psychological Inquiry*, 7(1), 1–15. https://doi.org/10.1207/s15327965pli0701_1
- Bawden, D. (2001). Information overload. *Library & information briefings*, 92, 1–15.

- Bender, S., Bloom, N., Card, D., Van Reenen, J., & Wolter, S. (2018). Management practices, workforce selection, and productivity. *Journal of Labor Economics*, 36(S1), 371–409. <https://doi.org/10.1086/694107>
- Berger, L., Libby, T., & Webb, A. (2018). The effects of tournament horizon and the percentage of winners on social comparisons and performance in multi-period competitions. *Accounting, Organizations and Society*, 64, 1–16. <https://doi.org/10.1016/j.aos.2017.12.006>
- Bernstein, E. S., & Li, S. (2017). Seeing Where You Stand: From Performance Feedback to Performance Transparency. *Academy of Management Annual Meeting Proceedings*, 2017(1), 14752. <https://doi.org/10.5465/ambpp.2017.14752abstract>
- Bezuijen, X. M., van den Berg, P. T., van Dam, K., & Thierry, H. (2009). Pygmalion and Employee Learning: The Role of Leader Behaviors. *Journal of Management*, 35(5), 1248–1267. <https://doi.org/10.1177/0149206308329966>
- Bhave, D. P. (2014). The invisible eye? Electronic performance monitoring and employee job performance. *Personnel Psychology*, 67(3), 605–635. <https://doi.org/10.1111/peps.12046>
- Blanes i Vidal, J., & Nossol, M. (2011). Tournaments Without Prizes: Evidence from Personnel Records. *Management Science*, 57(10), 1721–1736. <https://doi.org/10.1287/mnsc.1110.1383>
- Bloomfield, R., Nelson, M. W., & Soltes, E. (2016). Gathering Data for Archival, Field, Survey, and Experimental Accounting Research. *Journal of Accounting Research*, 54(2), 341–395. <https://doi.org/10.1111/1475-679X.12104>
- Bol, J. C. (2011). The Determinants and Performance Effects of Managers' Performance Evaluation Biases. *The Accounting Review*, 86(5), 1549–1575. <https://doi.org/10.2308/accr-10099>
- Bonner, B. L., Baumann, M. R., & Dalal, R. S. (2002). The effects of member expertise on group decision-making and performance. *Organizational Behavior and Human Decision Processes*, 88(2), 719–736. [https://doi.org/10.1016/S0749-5978\(02\)00010-9](https://doi.org/10.1016/S0749-5978(02)00010-9)
- Bonner, S. E., & Sprinkle, G. B. (2002). The effects of monetary incentives on effort and task performance: Theories, evidence, and a framework for research. *Accounting, Organizations and Society*, 27(4–5), 303–345. [https://doi.org/10.1016/S0361-3682\(01\)00052-6](https://doi.org/10.1016/S0361-3682(01)00052-6)
- Bordalo, P., Gennaioli, N., & Shleifer, A. (2012). Salience Theory of Choice Under Risk. *The Quarterly Journal of Economics*, 127(3), 1243–1285. <https://doi.org/10.1093/qje/qjs018>
- Bouncken, R. B., Gast, J., Kraus, S., & Bogers, M. (2015). Coopetition: a systematic review, synthesis, and future research directions. *Review of Managerial Science*, 9(3), 577–601. <https://doi.org/10.1007/s11846-015-0168-6>
- Bragger, J. D., Hantula, D. A., Bragger, D., Kirnan, J., & Kutcher, E. (2003). When success breeds failure: History, hysteresis, and delayed exit decisions. *Journal of Applied Psychology*, 88(1), 6–14. <https://doi.org/10.1037/0021-9010.88.1.6>
- Brand, J. E. (1905). The effect of verbal suggestion upon the estimation of linear magnitudes. *Psychological Review*, 12(1), 41–49. <https://doi.org/10.1037/h0073182>
- Brown, D. J., Ferris, D. L., Heller, D., & Keeping, L. M. (2007). Antecedents and consequences of the frequency of upward and downward social comparisons at work. *Organizational Behavior and Human Decision Processes*, 102(1), 59–75.

- <https://doi.org/10.1016/j.obhdp.2006.10.003>
- Brown, J. L., Farrington, S., & Sprinkle, G. B. (2016). Biased self-assessments, feedback, and employees' compensation plan choices. *Accounting, Organizations and Society*, 54, 45–59. <https://doi.org/10.1016/j.aos.2016.08.003>
- Buckert, M., Schwieren, C., Kudielka, B. M., & Fiebach, C. J. (2017). How stressful are economic competitions in the lab? An investigation with physiological measures. *Journal of Economic Psychology*, 62, 231–245. <https://doi.org/10.1016/j.joep.2017.07.004>
- Budescu, D. V., & Maciejovsky, B. (2005). The effect of payoff feedback and information pooling on reasoning errors: Evidence from experimental markets. *Management Science*, 51(12), 1829–1843. <https://doi.org/10.1287/mnsc.1050.0426>
- Bursztyjn, L., & Jensen, R. (2015). How Does Peer Pressure Affect Educational Investments?. *The Quarterly Journal of Economics*, 130(3), 1329–1367. <https://doi.org/10.1093/qje/qjv021>
- Buser, T., & Dreber, A. (2016). The Flipside of Comparative Payment Schemes. *Management Science*, 62(9), 2626–2638. <https://doi.org/10.1287/mnsc.2015.2257>
- Buser, T., Dreber, A., & Mollerstrom, J. (2017). The impact of stress on tournament entry. *Experimental Economics*, 20(2), 506–530. <https://doi.org/10.1007/s10683-016-9496-x>
- Cahlíková, J., & Cingl, L. (2017). Risk preferences under acute stress. *Experimental Economics*, 20(1), 209–236. <https://doi.org/10.1007/s10683-016-9482-3>
- Cahlíková, J., Cingl, L., & Lively, I. (2019). How Stress Affects Performance and Competitiveness Across Gender. *Management Science*. <https://doi.org/10.1287/mnsc.2019.3400>
- Campbell, D. (2008). Nonfinancial Performance Measures and Promotion-Based Incentives. *Journal of Accounting Research*, 46(2), 297–332. <https://doi.org/10.1111/j.1475-679X.2008.00275.x>
- Campbell, D. (2012). Employee Selection as a Control System. *Journal of Accounting Research*, 50(4), 931–966. <https://doi.org/10.1111/j.1475-679X.2012.00457.x>
- Campbell, J. P. (1990). The role of theory in industrial and organizational psychology. In M. D. Dunnette & L. M. Hough (Eds.), *Handbook of industrial and organizational psychology* (pp. 39–73). Consulting Psychologists Press.
- Cappelli, P., & Tavis, A. (2016). *The Performance Management Revolution*. Harvard Business Review. <https://hbr.org/2016/10/the-performance-management-revolution>
- Casas-Arce, P., Lourenço, S. M., & Martínez-Jerez, F. A. (2017). The Performance Effect of Feedback Frequency and Detail: Evidence from a Field Experiment in Customer Satisfaction. *Journal of Accounting Research*, 55(5), 1051–1088. <https://doi.org/10.1111/1475-679X.12184>
- Chan, E. W. (2018). Promotion, Relative Performance Information, and the Peter Principle. *The Accounting Review*, 93(3), 83–103. <https://doi.org/10.2308/accr-51890>
- Chan, K. C., Chen, C. R., & Cheng, L. T. W. (2009). Co-authorship Patterns in Accounting Research. *International Review of Accounting, Banking and Finance*, 1(2), 1–10. https://www.researchgate.net/publication/267196467_Co-authorship_Patterns_in_Accounting_Research
- Charness, G., Masclet, D., & Villeval, M. C. (2014). The Dark Side of Competition for Status.

- Management Science*, 60(1), 38–55. <https://doi.org/10.1287/mnsc.2013.1747>
- Checkland, P. B., & Scholes, P. C. (1990). *Soft systems methodology in action*. John Wiley & Sons.
- Chen, G., & Mathieu, J. E. (2008). Goal orientation dispositions and performance trajectories: The roles of supplementary and complementary situational inducements. *Organizational Behavior and Human Decision Processes*, 106(1), 21–38. <https://doi.org/10.1016/j.obhdp.2007.11.001>
- Chen, Y., Harper, F. M., Konstan, J., & Li, S. X. (2010). Social Comparisons and Contributions to Online Communities: A Field Experiment on MovieLens. *American Economic Review*, 100(4), 1358–1398. <https://doi.org/10.1257/aer.100.4.1358>
- Chetty, R., Looney, A., & Kroft, K. (2009). Saliency and Taxation: Theory and Evidence. *American Economic Review*, 99(4), 1145–1177. <https://doi.org/10.1257/aer.99.4.1145>
- Chewning, E. G., & Harrell, A. M. (1990). The effect of information load on decision makers' cue utilization levels and decision quality in a financial distress decision task. *Accounting, Organizations and Society*, 15(6), 527–542. [https://doi.org/10.1016/0361-3682\(90\)90033-Q](https://doi.org/10.1016/0361-3682(90)90033-Q)
- Christ, M. H., Emett, S. A., Summers, S. L., & Wood, D. A. (2012). The Effects of Preventive and Detective Controls on Employee Performance and Motivation. *Contemporary Accounting Research*, 29(2), 432–452. <https://doi.org/10.1111/j.1911-3846.2011.01106.x>
- Christ, M. H., Emett, S. A., Tayler, W. B., & Wood, D. A. (2016). Compensation or feedback: Motivating performance in multidimensional tasks. *Accounting, Organizations and Society*, 50, 27–40. <https://doi.org/10.1016/j.aos.2016.03.003>
- Chun, J. S., Brockner, J., & De Cremer, D. (2018). How temporal and social comparisons in performance evaluation affect fairness perceptions. *Organizational Behavior and Human Decision Processes*, 145, 1–15. <https://doi.org/10.1016/j.obhdp.2018.01.003>
- Cohen, S., Janicki-Deverts, D., & Miller, G. E. (2007). Psychological Stress and Disease. *JAMA*, 298(14), 1685–1687. <https://doi.org/10.1001/jama.298.14.1685>
- Colquitt, J. A., & Zapata-Phelan, C. P. (2007). Trends in Theory Building and Theory Testing: A Five-Decade Study of the Academy of Management Journal. *Academy of Management Journal*, 50(6), 1281–1303. <https://doi.org/10.5465/amj.2007.28165855>
- Dai, H., Dietvorst, B. J., Tuckfield, B., Milkman, K. L., & Schweitzer, M. E. (2018). Quitting When the Going Gets Tough: A Downside of High Performance Expectations. *Academy of Management Journal*, 61(5), 1667–1691. <https://doi.org/10.5465/amj.2014.1045>
- Dauten, P. M. J. (1958). Preface. *The Journal of the Academy of Management*, 1(1), 5–6. www.jstor.org/stable/254573
- Davidson, S. (1966). Editor's Preface. *Journal of Accounting Research*, 4, iii. www.jstor.org/stable/2490161
- DellaVigna, S. (2009). Psychology and Economics: Evidence from the Field. *Journal of Economic Literature*, 47(2), 315–372. <https://doi.org/10.1257/jel.47.2.315>
- Deloitte. (2019). *Annual Review of Football Finance 2019*. <https://www2.deloitte.com/uk/en/pages/sports-business-group/articles/annual-review-of-football-finance.html>
- DerSimonian, R., & Laird, N. (1986). Meta-analysis in clinical trials. *Controlled Clinical*

- Trials*, 7(3), 177–188. [https://doi.org/10.1016/0197-2456\(86\)90046-2](https://doi.org/10.1016/0197-2456(86)90046-2)
- DeShon, R. P., Kozlowski, S. W. J., Schmidt, A. M., Milner, K. R., & Wiechmann, D. (2004). A multiple-goal, multilevel model of feedback effects on the regulation of individual and team performance. *Journal of Applied Psychology*, 89(6), 1035–1056. <https://doi.org/10.1037/0021-9010.89.6.1035>
- DeZoort, T., Harrison, P., & Taylor, M. (2006). Accountability and auditors' materiality judgments: The effects of differential pressure strength on conservatism, variability, and effort. *Accounting, Organizations and Society*, 31(4–5), 373–390. <https://doi.org/10.1016/j.aos.2005.09.001>
- Dimotakis, N., Mitchell, D., & Maurer, T. (2017). Positive and negative assessment center feedback in relation to development self-efficacy, feedback seeking, and promotion. *Journal of Applied Psychology*, 102(11), 1514–1527. <https://doi.org/10.1037/apl0000228>
- Dyreng, S. D., Mayew, W. J., & Williams, C. D. (2012). Religious Social Norms and Corporate Financial Reporting. *Journal of Business Finance and Accounting*, 39(7–8), 845–875. <https://doi.org/10.1111/j.1468-5957.2012.02295.x>
- Earley, C. E. (2001). Knowledge Acquisition in Auditing: Training Novice Auditors to Recognize Cue Relationships in Real Estate Valuation. *The Accounting Review*, 76(1), 81–97. <https://doi.org/10.2308/accr.2001.76.1.81>
- Engelbrecht-Wiggans, R., & Katok, E. (2008). Regret and feedback information in first-price sealed-bid auctions. *Management Science*, 54(4), 808–819. <https://doi.org/10.1287/mnsc.1070.0806>
- Eppler, M. J., & Mengis, J. (2004). The Concept of Information Overload: A Review of Literature from Organization Science, Accounting, Marketing, MIS, and Related Disciplines. *The Information Society*, 20(5), 325–344. <https://doi.org/10.1080/01972240490507974>
- Eriksson, T., Mao, L., & Villeval, M. C. (2017). Saving face and group identity. *Experimental Economics*, 20(3), 622–647. <https://doi.org/10.1007/s10683-016-9502-3>
- ERIM. (2018). *ERIM Journals List (EJL) 2016-2021*. <https://www.irim.eur.nl/about-erim/erim-journals-list-ejl/>
- Ettema, J. H., & Zielhuis, R. L. (1971). Physiological parameters of mental load. *Ergonomics*, 14(1), 137–144. <https://doi.org/10.1080/00140137108931232>
- Ewenstein, B., Hancock, B., & Komm, A. (2016). Ahead of the curve: The future of performance management. *McKinsey Quarterly*, 2, 64–73. <https://www.mckinsey.com/business-functions/organization/our-insights/ahead-of-the-curve-the-future-of-performance-management>
- Eyring, H., Ferguson, P. F. & Koppers, S. (2019). Less Information, More Comparison, and Better Performance: Evidence from a Field Experiment. *Unpublished Working Paper*.
- Eyring, H., & Narayanan, V. G. (2018). Performance Effects of Setting a High Reference Point for Peer-Performance Comparison. *Journal of Accounting Research*, 56(2), 581–615. <https://doi.org/10.1111/1475-679X.12199>
- Falk, A., & Ichino, A. (2006). Clean Evidence on Peer Effects. *Journal of Labor Economics*, 24(1), 39–57. <https://doi.org/10.1086/497818>
- Feldman, M. S., & Orlikowski, W. J. (2011). Theorizing Practice and Practicing Theory. *Organization Science*, 22(5), 1240–1253. <https://doi.org/10.1287/orsc.1100.0612>

- Fernbach, P. M., Hagmayer, Y., & Sloman, S. A. (2014). Effort denial in self-deception. *Organizational Behavior and Human Decision Processes*, 123(1), 1–8. <https://doi.org/10.1016/j.obhdp.2013.10.013>
- Festinger, L. (1954). A Theory of Social Comparison Processes. *Human Relations*, 7(2), 117–140. <https://doi.org/10.1177/001872675400700202>
- Fischer, I., & Budescu, D. V. (2005). When do those who know more also know more about how much they know? The development of confidence and performance in categorical decision tasks. *Organizational Behavior and Human Decision Processes*, 98(1), 39–53. <https://doi.org/10.1016/j.obhdp.2005.04.003>
- Fisher, J. G., Maines, L. A., Peffer, S. A., & Sprinkle, G. B. (2002). Using Budgets for Performance Evaluation: Effects of Resource Allocation and Horizontal Information Asymmetry on Budget Proposals, Budget Slack, and Performance. *The Accounting Review*, 77(4), 847–865. <https://doi.org/10.2308/accr.2002.77.4.847>
- Floyd, E., & List, J. A. (2016). Using Field Experiments in Accounting and Finance. *Journal of Accounting Research*, 54(2), 437–475. <https://doi.org/10.1111/1475-679X.12113>
- Fong, C. J., Patall, E. A., Vasquez, A. C., & Stautberg, S. (2019). A Meta-Analysis of Negative Feedback on Intrinsic Motivation. *Educational Psychology Review*, 31, 121–162. <https://doi.org/10.1007/s10648-018-9446-6>
- Frederick, S. (2005). Cognitive Reflection and Decision Making. *Journal of Economic Perspectives*, 19(4), 25–42. <https://doi.org/10.1257/089533005775196732>
- Frederickson, J. R. (1992). Relative Performance Information: The Effects of Common Uncertainty and Contract Type on Agent Effort. *The Accounting Review*, 67(4), 647–669. <https://doi.org/10.2307/248317>
- Garcia, S. M., & Tor, A. (2007). Rankings, standards, and competition: Task vs. scale comparisons. *Organizational Behavior and Human Decision Processes*, 102(1), 95–108. <https://doi.org/10.1016/j.obhdp.2006.10.004>
- Gennaioli, N., & Shleifer, A. (2010). What Comes to Mind. *The Quarterly Journal of Economics*, 125(4), 1399–1433. <https://doi.org/10.1162/qjec.2010.125.4.1399>
- Gibson, F. P. (2000). Feedback Delays: How Can Decision Makers Learn Not to Buy a New Car Every Time the Garage Is Empty?. *Organizational Behavior and Human Decision Processes*, 83(1), 141–166. <https://doi.org/10.1006/obhd.2000.2906>
- Giersch, T. (2014, May 25). *Vom Bolzplatz zum Fußballprofi*. Handelsblatt. <https://www.handelsblatt.com/sport/fussball/der-grosse-traum-vom-bolzplatz-zum-fussballprofi/9918966-all.html>
- Gigerenzer, G., & Goldstein, D. G. (2011). The recognition heuristic : A decade of research. *Judgment and Decision Making*, 6(1), 100–121. <http://www.sas.upenn.edu/~baron/journal/11/rh15/rh15.pdf>
- Gill, D., Kissova, Z., Lee, J., & Prowse, V. (2019). First-Place Loving and Last-Place Loathing: How Rank in the Distribution of Performance Affects Effort Provision. *Management Science*, 65(2), 494–507. <https://doi.org/10.1287/mnsc.2017.2907>
- Gill, D., & Prowse, V. (2014). Gender differences and dynamics in competition: The role of luck. *Quantitative Economics*, 5(2), 351–376. <https://doi.org/10.3982/qe309>
- Gino, F., & Staats, B. R. (2011). Driven by Social Comparisons: How Feedback about Coworkers' Effort Influences Individual Productivity. *Harvard Business School NOM*

- Unit Working Paper*. <https://doi.org/10.2139/ssrn.1755051>
- Gist, M. E., & Mitchell, T. R. (1992). Self-Efficacy: A Theoretical Analysis of Its Determinants and Malleability. *The Academy of Management Review*, 17(2), 183–211. <https://doi.org/10.2307/258770>
- Glass, G. V. (1976). Primary, Secondary, and Meta-Analysis of Research. *Educational Researcher*, 5(10), 3–8. <https://doi.org/10.3102/0013189X005010003>
- Gneezy, U., & Imas, A. (2017). Lab in the field: Measuring preferences in the wild. *Handbook of Field Experiments*, 1, 1–40. <https://doi.org/10.1017/CBO9781107415324.004>
- Goh, J., Pfeffer, J., & Zenios, S. A. (2016). The relationship between workplace stressors and mortality and health costs in the United State. *Management Science*, 62(2), 608–628. <https://doi.org/10.1287/mnsc.2014.2115>
- Goldstein, D. G., & Gigerenzer, G. (2002). Models of ecological rationality: The recognition heuristic. *Psychological Review*, 109(1), 75–90. <https://doi.org/10.1037//0033-295X.109.1.75>
- Gonzalez, C. (2005). Decision support for real-time, dynamic decision-making tasks. *Organizational Behavior and Human Decision Processes*, 96(2), 142–154. <https://doi.org/10.1016/j.obhdp.2004.11.002>
- Gonzalez-Mulé, E., Courtright, S. H., DeGeest, D., Seong, J.-Y., & Hong, D.-S. (2016). Channeled Autonomy. *Journal of Management*, 42(7), 2018–2033. <https://doi.org/10.1177/0149206314535443>
- Goodman, J. S., & Wood, R. E. (2004). Feedback Specificity, Learning Opportunities, and Learning. *Journal of Applied Psychology*, 89(5), 809–821. <https://doi.org/10.1037/0021-9010.89.5.809>
- Goodman, J. S., Wood, R. E., & Chen, Z. (2011). Feedback specificity, information processing, and transfer of training. *Organizational Behavior and Human Decision Processes*, 115(2), 253–267. <https://doi.org/10.1016/j.obhdp.2011.01.001>
- Goodman, J. S., Wood, R. E., & Hendrickx, M. (2004). Feedback Specificity, Exploration, and Learning. *Journal of Applied Psychology*, 89(2), 248–262. <https://doi.org/10.1037/0021-9010.89.2.248>
- Gosnell, G. K., List, J. A., & Metcalfe, R. D. (2020). The Impact of Managerial Practices on Employee Productivity: A Field Experiment with Airline Captains. *Journal of Political Economy*, 128(4), 1195–1233. <https://doi.org/10.1086/705375>
- Halko, M. L., & Sääksvuori, L. (2017). Competitive behavior, stress, and gender. *Journal of Economic Behavior and Organization*, 141, 96–109. <https://doi.org/10.1016/j.jebo.2017.06.014>
- Hannan, R. L., Krishnan, R., & Newman, A. H. (2008). The Effects of Disseminating Relative Performance Feedback in Tournament and Individual Performance Compensation Plans. *The Accounting Review*, 83(4), 893–913. <https://doi.org/10.2308/accr.2008.83.4.893>
- Hannan, R. L., McPhee, G. P., Newman, A. H., & Tafkov, I. D. (2013). The Effect of Relative Performance Information on Performance and Effort Allocation in a Multi-Task Environment. *The Accounting Review*, 88(2), 553–575. <https://doi.org/10.2308/accr-50312>
- Hannan, R. L., McPhee, G. P., Newman, A. H., & Tafkov, I. D. (2019). The Informativeness of Relative Performance Information and its Effect on Effort Allocation in a Multi-Task

- Environment. *Contemporary Accounting Research*, 36(3), 1607–1633. <https://doi.org/10.1111/1911-3846.12482>
- Hartwell, C. J., & Campion, M. A. (2016). Getting on the same page: The effect of normative feedback interventions on structured interview ratings. *Journal of Applied Psychology*, 101(6), 757–778. <https://doi.org/10.1037/apl0000099>
- Hauser, O. P., Linos, E., & Rogers, T. (2017). Innovation with field experiments: Studying organizational behaviors in actual organizations. *Research in Organizational Behavior*, 37, 185–198. <https://doi.org/10.1016/j.riob.2017.10.004>
- Hazels, B., & Sasse, C. (2008). Forced Ranking: A Review. *SAM Advanced Management Journal*, 73(2), 35–39. <https://search.proquest.com/docview/231246704/fulltextPDF/6C8E45D09C67403CPQ/1?accountid=33964>
- Hirshleifer, D., & Teoh, S. H. (2003). Limited attention, information disclosure, and financial reporting. *Journal of Accounting and Economics*, 36(1–3), 337–386. <https://doi.org/10.1016/j.jacceco.2003.10.002>
- Hodge, F. D., Hopkins, P. E., & Wood, D. A. (2010). The Effects of Financial Statement Information Proximity and Feedback on Cash Flow Forecasts. *Contemporary Accounting Research*, 27(1), 101–133. <https://doi.org/10.1111/j.1911-3846.2010.01003.x>
- Hoever, I. J., Zhou, J., & van Knippenberg, D. (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), 2159–2181. <https://doi.org/10.5465/amj.2016.0642>
- Holmstrom, B. (1979). Moral Hazard and Observability. *The Bell Journal of Economics*, 10(1), 74–91. <https://doi.org/10.2307/3003320>
- Huntington-Klein, N., & Rose, E. (2018). Gender Peer Effects in a Predominantly Male Environment: Evidence from West Point. *AEA Papers and Proceedings*, 108, 392–395. <https://doi.org/10.1257/pandp.20181114>
- Ilgen, D. R., Fisher, C. D., & Taylor, M. S. (1979). Consequences of individual feedback on behavior in organizations. *Journal of Applied Psychology*, 64(4), 349–371. <https://doi.org/10.1037/0021-9010.64.4.349>
- Ilies, R., & Judge, T. A. (2005). Goal regulation across time: The effects of feedback and affect. *Journal of Applied Psychology*, 90(3), 453–467. <https://doi.org/10.1037/0021-9010.90.3.453>
- Indjejikian, R. J. (1999). Discussion of “Optimal Contracting, Accounting Standards, and Market Structures”. *Contemporary Accounting Research*, 16(2), 277–280. <https://doi.org/10.1111/j.1911-3846.1999.tb00581.x>
- Ingle, S. (2018, November 25). *Barcelona become first sports team to average £10m a year in wages*. The Guardian. <https://www.theguardian.com/football/2018/nov/25/barcelona-team-average-10m-year-wages-sporting-intelligence>
- Jermias, J. (2001). Cognitive dissonance and resistance to change: The influence of commitment confirmation and feedback on judgment usefulness of accounting systems. *Accounting, Organizations and Society*, 26(2), 141–160. [https://doi.org/10.1016/S0361-3682\(00\)00008-8](https://doi.org/10.1016/S0361-3682(00)00008-8)
- Johnson, M. D., Hollenbeck, J. R., Scott DeRue, D., Barnes, C. M., & Jundt, D. (2013).

- Functional versus dysfunctional team change: Problem diagnosis and structural feedback for self-managed teams. *Organizational Behavior and Human Decision Processes*, 122(1), 1–11. <https://doi.org/10.1016/j.obhdp.2013.03.006>
- Judge, T. A., & Zapata, C. P. (2015). The Person–Situation Debate Revisited: Effect of Situation Strength and Trait Activation on the Validity of the Big Five Personality Traits in Predicting Job Performance. *Academy of Management Journal*, 58(4), 1149–1179. <https://doi.org/10.5465/amj.2010.0837>
- Jung, J. H., Schneider, C., & Valacich, J. (2010). Enhancing the motivational affordance of information systems: The effects of real-time performance feedback and goal setting in group collaboration environments. *Management Science*, 56(4), 724–742. <https://doi.org/10.1287/mnsc.1090.1129>
- Kachelmeier, S. J. (2019). Discussion of “The Informativeness of Relative Performance Information and Its Effect on Effort Allocation in a Multitask Environment”. *Contemporary Accounting Research*, 36(3), 1–11. <https://doi.org/10.1111/1911-3846.12482>
- Kahn, L. M. (2000). The Sports Business as a Labor Market Laboratory. *Journal of Economic Perspectives*, 14(3), 75–94. <https://doi.org/10.1257/jep.14.3.75>
- Kahneman, D., & Tversky, A. (1972). Subjective probability: A judgment of representativeness. *Cognitive Psychology*, 3(3), 430–454. [https://doi.org/10.1016/0010-0285\(72\)90016-3](https://doi.org/10.1016/0010-0285(72)90016-3)
- Karagözoğlu, E., & Riedl, A. (2015). Performance information, production uncertainty, and subjective entitlements in bargaining. *Management Science*, 61(11), 2611–2626. <https://doi.org/10.1287/mnsc.2014.2012>
- Kassis, M., Schmidt, S. L., Schreyer, D., & Torgler, B. (2017). Who gets promoted? Personality factors leading to promotion in highly structured work environments: evidence from a German professional football club. *Applied Economics Letters*, 24(17), 1222–1227. <https://doi.org/10.1080/13504851.2016.1267841>
- Kelly, K. (2007). Feedback and Incentives on Nonfinancial Value Drivers: Effects on Managerial Decision Making. *Contemporary Accounting Research*, 24(2), 523–556. <https://doi.org/10.1506/Q5X8-5912-W613-0UTL>
- Kim, P. H., Diekmann, K. A., & Tenbrunsel, A. E. (2003). Flattery may get you somewhere: The strategic implications of providing positive vs. negative feedback about ability vs. ethicality in negotiation. *Organizational Behavior and Human Decision Processes*, 90(2), 225–243. [https://doi.org/10.1016/S0749-5978\(02\)00522-8](https://doi.org/10.1016/S0749-5978(02)00522-8)
- King, E. B., Botsford, W., Hebl, M. R., Kazama, S., Dawson, J. F., & Perkins, A. (2012). Benevolent Sexism at Work: Gender Differences in the Distribution of Challenging Developmental Experiences. *Journal of Management*, 38(6), 1835–1866. <https://doi.org/10.1177/0149206310365902>
- Kirschbaum, C., Pirke, K.-M., & Hellhammer, D. H. (1993). The ‘Trier Social Stress Test’ – A Tool for Investigating Psychobiological Stress Responses in a Laboratory Setting. *Neuropsychobiology*, 28(1–2), 76–81. <https://doi.org/10.1159/000119004>
- Kizilcec, R. F. (2016). How Much Information?: Effects of Transparency on Trust in an Algorithmic Interface. *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, 2390–2395. <https://doi.org/10.1145/2858036.2858402>
- Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: A

- historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin*, *119*(2), 254–284. <https://doi.org/10.1037/0033-2909.119.2.254>
- Kocher, M. G., Lenz, M. V., & Sutter, M. (2012). Psychological pressure in competitive environments: New evidence from randomized natural experiments. *Management Science*, *58*(8), 1585–1591. <https://doi.org/10.1287/mnsc.1120.1516>
- Koppers, S. (2019). The Performance Effects of Relative Performance Information and Feedback Detail: Evidence from a Field Experiment in Austria. *AEA RCT Registry*. July 01. <https://doi.org/10.1257/rct.4141-3.0>
- Koppers, S., & Schmidt, S. L. (2020). A Systematic Review of Performance Feedback Literature from Accounting, Management General and Organizational Behavior. *Unpublished Working Paper*.
- Koppers, S., Schmidt, S. L. & Schreyer, D. (2020). Providing Ranks as a Managerial Practice to Reduce Stress: How Performance Feedback Affects Individual Recipients. *Unpublished Working Paper*.
- Kramer, S., Maas, V. S., & Rinsum, M. van. (2016). Relative performance information, rank ordering and employee performance: A research note. *Management Accounting Research*, *33*, 16–24. <https://doi.org/10.1016/j.mar.2016.03.004>
- Krueger, A. B., & Whitmore, D. M. (2001). The Effect of Attending a Small Class in the Early Grades on College-test Taking and Middle School Test Results: Evidence from Project Star. *The Economic Journal*, *111*(468), 1–28. <https://doi.org/10.1111/1468-0297.00586>
- Kuhnen, C. M., & Tymula, A. (2012). Feedback, Self-Esteem, and Performance in Organizations. *Management Science*, *58*(1), 94–113. <https://doi.org/10.1287/mnsc.1110.1379>
- Lam, C. F., DeRue, D. S., Karam, E. P., & Hollenbeck, J. R. (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. <https://doi.org/10.1016/j.obhdp.2011.05.002>
- Lam, S. S. K., Schaubroeck, J., & Brown, A. D. (2004). Esteem maintenance among groups: Laboratory and field studies of group performance cognitions. *Organizational Behavior and Human Decision Processes*, *94*(2), 86–101. <https://doi.org/10.1016/j.obhdp.2004.03.004>
- Lam, S. S. K., Yik, M. S. M., & Schaubroeck, J. (2002). Responses to formal performance appraisal feedback: the role of negative affectivity. *The Journal of Applied Psychology*, *87*(1), 192–201. <https://doi.org/10.1037/0021-9010.87.1.192>
- Latham, G. P., & Locke, E. A. (1991). Self-regulation through goal setting. *Organizational Behavior and Human Decision Processes*, *50*(2), 212–247. [https://doi.org/10.1016/0749-5978\(91\)90021-K](https://doi.org/10.1016/0749-5978(91)90021-K)
- Lechermeier, J., & Fassnacht, M. (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), 145–193. <https://doi.org/10.1007/s11301-018-0136-8>
- Leiblein, M. J., Chen, J. S., & Posen, H. E. (2017). Resource Allocation in Strategic Factor Markets: A Realistic Real Options Approach to Generating Competitive Advantage. *Journal of Management*, *43*(8), 2588–2608. <https://doi.org/10.1177/0149206316683778>

- Lejarraga, T., & Gonzalez, C. (2011). Effects of feedback and complexity on repeated decisions from description. *Organizational Behavior and Human Decision Processes*, *116*(2), 286–295. <https://doi.org/10.1016/j.obhdp.2011.05.001>
- LePine, J. A., Hollenbeck, J. R., Ilgen, D. R., Colquitt, J. A., & Ellis, A. (2002). Gender composition, situational strength, and team decision-making accuracy: A criterion decomposition approach. *Organizational Behavior and Human Decision Processes*, *88*(1), 445–475. <https://doi.org/10.1006/obhd.2001.2986>
- Leung, P. W., & Trotman, K. T. (2005). The effects of feedback type on auditor judgment performance for configural and non-configural tasks. *Accounting, Organizations and Society*, *30*(6), 537–553. <https://doi.org/10.1016/j.aos.2004.11.003>
- Levitt, S. D., & List, J. A. (2009). Field experiments in economics: The past, the present, and the future. *European Economic Review*, *53*(1), 1–18. <https://doi.org/10.1016/j.eurocorev.2008.12.001>
- Linderbaum, B. A., & Levy, P. E. (2010). The development and validation of the feedback orientation scale (FOS). *Journal of Management*, *36*(6), 1372–1405. <https://doi.org/10.1177/0149206310373145>
- Locke, E. A. (1968). Toward a theory of task motivation and incentives. *Organizational Behavior and Human Performance*, *3*(2), 157–189. [https://doi.org/10.1016/0030-5073\(68\)90004-4](https://doi.org/10.1016/0030-5073(68)90004-4)
- Locke, E. A., Motowidlo, S. J., & Bobko, P. (1986). Using Self-Efficacy Theory to Resolve the Conflict Between Goal-Setting Theory and Expectancy Theory in Organizational Behavior and Industrial/Organizational Psychology. *Journal of Social and Clinical Psychology*, *4*(3), 328–338. <https://doi.org/10.1521/jscp.1986.4.3.328>
- Loftus, S., & Tanlu, L. J. (2018). Because of “Because”: Examining the Use of Causal Language in Relative Performance Feedback. *The Accounting Review*, *93*(2), 277–297. <https://doi.org/10.2308/accr-51830>
- Lourenço, S. M. (2016). Monetary Incentives, Feedback, and Recognition—Complements or Substitutes? Evidence from a Field Experiment in a Retail Services Company. *The Accounting Review*, *91*(1), 279–297. <https://doi.org/10.2308/accr-51148>
- Lourenço, S. M., Greenberg, J. O., Littlefield, M., Bates, D. W., & Narayanan, V. G. (2018). The performance effect of feedback in a context of negative incentives: Evidence from a field experiment. *Management Accounting Research*, *40*, 1–14. <https://doi.org/10.1016/j.mar.2018.01.002>
- Luft, J. (2016). Cooperation and competition among employees: Experimental evidence on the role of management control systems. *Management Accounting Research*, *31*, 75–85. <https://doi.org/10.1016/j.mar.2016.02.006>
- Lurie, N. H., & Swaminathan, J. M. (2009). Is timely information always better? The effect of feedback frequency on decision making. *Organizational Behavior and Human Decision Processes*, *108*(2), 315–329. <https://doi.org/10.1016/j.obhdp.2008.05.005>
- Maciejovsky, B., Sutter, M., V. Budescu, D., & Bernau, P. (2013). Teams make you smarter: How exposure to teams improves individual decisions in probability and reasoning tasks. *Management Science*, *59*(6), 1255–1270. <https://doi.org/10.1287/mnsc.1120.1668>
- Mak, V., Zwick, R., Rao, A. R., & Pattaratanakun, J. A. (2015). “Pay what you want” as threshold public good provision. *Organizational Behavior and Human Decision Processes*, *127*, 30–43. <https://doi.org/10.1016/j.obhdp.2014.11.004>

- Malhotra, N. K. (1982). Information Load and Consumer Decision Making. *Journal of Consumer Research*, 8(4), 419–430. <https://doi.org/10.1086/208882>
- Mas, A., & Moretti, E. (2009). Peers at Work. *American Economic Review*, 99(1), 112–145. <https://doi.org/10.1257/aer.99.1.112>
- Massey, C., & Thaler, R. H. (2013). The loser's curse: Decision making and market efficiency in the National Football League draft. *Management Science*, 59(7), 1479–1495. <https://doi.org/10.1287/mnsc.1120.1657>
- Meeker, M. (2018, May 30). *Internet Trends Report 2018*. <https://www.kleinerperkins.com/perspectives/internet-trends-report-2018/>
- Mercer. (2019). *Mercer's 2019 Global Performance Management Survey*. <https://www.imercer.com/uploads/common/HTML/LandingPages/AnalyticalHub/june2019-mercero-2019-global-performance-management-survey-executive-summary.pdf>
- Miller, G. A. (1956). The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychological Review*, 63(2), 81–97. <https://doi.org/10.1037/h0043158>
- Moore, D. A., & Klein, W. M. P. (2008). Use of absolute and comparative performance feedback in absolute and comparative judgments and decisions. *Organizational Behavior and Human Decision Processes*, 107(1), 60–74. <https://doi.org/10.1016/j.obhdp.2008.02.005>
- Moreland, R. L., & Myaskovsky, L. (2000). Exploring the performance benefits of group training: Transactive memory or improved communication?. *Organizational Behavior and Human Decision Processes*, 82(1), 117–133. <https://doi.org/10.1006/obhd.2000.2891>
- Moreno, K. K., Bhattacharjee, S., & Brandon, D. M. (2007). The effectiveness of alternative training techniques on analytical procedures performance. *Contemporary Accounting Research*, 24(3), 983–1014. <https://doi.org/10.1506/car.24.3.11>
- Narayanan, S., Balasubramanian, S., & Swaminathan, J. M. (2009). A Matter of Balance: Specialization, Task Variety, and Individual Learning in a Software Maintenance Environment. *Management Science*, 55(11), 1861–1876. <https://doi.org/10.1287/mnsc.1090.1057>
- Nelson, M. W., Krische, S. D., & Bloomfield, R. J. (2003). Confidence and Investors' Reliance on Disciplined Trading Strategies. *Journal of Accounting Research*, 41(3), 503–523. <https://doi.org/10.1111/1475-679X.00114>
- Newman, A. H., & Tafkov, I. D. (2014). Relative performance information in tournaments with different prize structures. *Accounting, Organizations and Society*, 39(5), 348–361. <https://doi.org/10.1016/j.aos.2014.05.004>
- Nordstrom, R., Lorenzi, P., & Hall, R. V. (1991). A Review of Public Posting of Performance Feedback in Work Settings. *Journal of Organizational Behavior Management*, 11(2), 101–124. https://doi.org/10.1300/J075v11n02_07
- Northcraft, G. B., Schmidt, A. M., & Ashford, S. J. (2011). Feedback and the Rationing of Time and Effort Among Competing Tasks. *Journal of Applied Psychology*, 96(5), 1076–1086. <https://doi.org/10.1037/a0023221>
- Oc, B., Bashshur, M. R., & Moore, C. (2015). Speaking truth to power: The effect of candid feedback on how individuals with power allocate resources. *Journal of Applied Psychology*, 100(2), 450–463. <https://doi.org/10.1037/a0038138>

- Oldham, G. R., & Hackman, J. R. (2010). Not what it was and not what it will be: The future of job design research. *Journal of Organizational Behavior*, *31*(2–3), 463–479. <https://doi.org/10.1002/job.678>
- Palmatier, R. W., Houston, M. B., & Hulland, J. (2018). Review articles: purpose, process, and structure. *Journal of the Academy of Marketing Science*, *46*(1), 1–5. <https://doi.org/10.1007/s11747-017-0563-4>
- Peterson, N. (1982). Feedback is Not a New Principle of Behavior. *The Behavior Analyst*, *5*(1), 101–102. <https://doi.org/10.1007/BF03393144>
- Peysakhovich, A., & Rand, D. G. (2016). Habits of Virtue: Creating Norms of Cooperation and Defection in the Laboratory. *Management Science*, *62*(3), 631–647. <https://doi.org/10.1287/mnsc.2015.2168>
- PFA. (n.d.). *Age*. Retrieved November 19, 2019, from <https://www.thepfa.com/equalities/commitment-to-all/age>
- Pillutla, M. M., & Chen, X. P. (1999). Social Norms and Cooperation in Social Dilemmas: The Effects of Context and Feedback. *Organizational Behavior and Human Decision Processes*, *78*(2), 81–103. <https://doi.org/10.1006/obhd.1999.2825>
- Pope, D. G. (2009). Reacting to rankings: Evidence from “America’s Best Hospitals”. *Journal of Health Economics*, *28*(6), 1154–1165. <https://doi.org/10.1016/j.jhealeco.2009.08.006>
- Prendergast, C. (1999). The Provision of Incentives in Firms. *Journal of Economic Literature*, *37*(1), 7–63. <https://doi.org/10.1257/jel.37.1.7>
- Rabin, M. (2002). Inference by Believers in the Law of Small Numbers. *The Quarterly Journal of Economics*, *117*(3), 775–816. <https://doi.org/10.1162/003355302760193896>
- Ramaprasad, A. (1983). On the definition of feedback. *Behavioral Science*, *28*(1), 4–13. <https://doi.org/10.1002/bs.3830280103>
- Rammstedt, B., & John, O. P. (2007). Measuring personality in one minute or less: A 10-item short version of the Big Five Inventory in English and German. *Journal of Research in Personality*, *41*(1), 203–212. <https://doi.org/10.1016/j.jrp.2006.02.001>
- Ray, K. (2007). Performance evaluations and efficient sorting. *Journal of Accounting Research*, *45*(4), 839–882. <https://doi.org/10.1111/j.1475-679X.2007.00253.x>
- Reinsel, D., Gantz, J., & Rydning, J. (2017). *Data Age 2025: The Evolution of Data to Life-Critical*. Seagate. <http://www.seagate.com/www-content/our-story/trends/files/Seagate-WP-DataAge2025-March-2017.pdf>
- Rolison, J. J., Evans, J. S. B. T., Dennis, I., & Walsh, C. R. (2012). Dual-processes in learning and judgment: Evidence from the multiple cue probability learning paradigm. *Organizational Behavior and Human Decision Processes*, *118*(2), 189–202. <https://doi.org/10.1016/j.obhdp.2012.03.003>
- Rose, J. P., & Windschitl, P. D. (2008). How egocentrism and optimism change in response to feedback in repeated competitions. *Organizational Behavior and Human Decision Processes*, *105*(2), 201–220. <https://doi.org/10.1016/j.obhdp.2007.08.003>
- Ruttan, R. L., & Nordgren, L. F. (2016). The strength to face the facts: Self-regulation defends against defensive information processing. *Organizational Behavior and Human Decision Processes*, *137*, 86–98. <https://doi.org/10.1016/j.obhdp.2016.06.006>
- Rynes, S. L. (2005). Taking Stock and Looking Ahead. *Academy of Management Journal*,

- 48(1), 9–15. <https://doi.org/10.5465/amj.2005.15993108>
- Savage, L. (1954). *The foundations of statistics*. Dover Publications.
- Schick, A. G., Gordon, L. A., & Haka, S. (1990). Information overload: A temporal approach. *Accounting, Organizations and Society*, 15(3), 199–220. [https://doi.org/10.1016/0361-3682\(90\)90005-F](https://doi.org/10.1016/0361-3682(90)90005-F)
- Schnieder, C. (2018). How Relative Performance Information Affects Employee Performance: A Systematic Review of Empirical Research. Available at SSRN. <https://doi.org/10.2139/ssrn.3327238>
- Schroder, H. M., Driver, M. J., & Streufert, S. (1967). *Human information processing—Individuals and groups functioning in complex social situations*. Holt, Rinehart and Winston.
- Schultze, T., Mojzisch, A., & Schulz-Hardt, S. (2012). Why groups perform better than individuals at quantitative judgment tasks: Group-to-individual transfer as an alternative to differential weighting. *Organizational Behavior and Human Decision Processes*, 118(1), 24–36. <https://doi.org/10.1016/j.obhdp.2011.12.006>
- Schultze, T., Pfeiffer, F., & Schulz-Hardt, S. (2012). Biased information processing in the escalation paradigm: Information search and information evaluation as potential mediators of escalating commitment. *Journal of Applied Psychology*, 97(1), 16–32. <https://doi.org/10.1037/a0024739>
- Seifert, C. F., Yukl, G., & McDonald, R. A. (2003). Effects of multisource feedback and a feedback facilitator on the influence behavior of managers toward subordinates. *Journal of Applied Psychology*, 88(3), 561–569. <https://doi.org/10.1037/0021-9010.88.3.561>
- Sengul, M., & Obloj, T. (2017). Better Safe Than Sorry: Subsidiary Performance Feedback and Internal Governance in Multiunit Firms. *Journal of Management*, 43(8), 2526–2554. <https://doi.org/10.1177/0149206316677298>
- Seo, M. G., & Ilies, R. (2009). The role of self-efficacy, goal, and affect in dynamic motivational self-regulation. *Organizational Behavior and Human Decision Processes*, 109(2), 120–133. <https://doi.org/10.1016/j.obhdp.2009.03.001>
- Shea, C., & Howell, J. M. (2000). Efficacy-performance spirals: An empirical test. *Journal of Management*, 26(4), 791–812. [https://doi.org/10.1016/S0149-2063\(00\)00056-8](https://doi.org/10.1016/S0149-2063(00)00056-8)
- Sheldon, O. J., Dunning, D., & Ames, D. R. (2014). Emotionally unskilled, unaware, and uninterested in learning more: Reactions to feedback about deficits in emotional intelligence. *Journal of Applied Psychology*, 99(1), 125–137. <https://doi.org/10.1037/a0034138>
- Shenhav, A., Rand, D. G., & Greene, J. D. (2012). Divine intuition: Cognitive style influences belief in God. *Journal of Experimental Psychology: General*, 141(3), 423–428. <https://doi.org/10.1037/a0025391>
- Shunko, M., Niederhoff, J., & Rosokha, Y. (2018). Humans are not machines: The behavioral impact of queueing design on service time. *Management Science*, 64(1), 453–473. <https://doi.org/10.1287/mnsc.2016.2610>
- Simnett, R. (1996). The effect of information selection, information processing and task complexity on predictive accuracy of auditors. *Accounting, Organizations and Society*, 21(7–8), 699–719. [https://doi.org/10.1016/0361-3682\(96\)00006-2](https://doi.org/10.1016/0361-3682(96)00006-2)
- Simon, H. A. (1979). Rational Decision Making in Business Organizations. *The American*

- Economic Review*, 69(4), 493–513. <https://www.jstor.org/stable/1808698>
- Smith, E. (2018, April 3). *The Future of HR? Real-Time Feedback Tools Take Off At Amazon, Elsewhere*. Associations Now. <https://associationsnow.com/2018/04/future-hr-real-time-feedback-tools-take-off-amazon-elsewhere/>
- Smith, R. H. (2000). Assimilative and Contrastive Emotional Reactions to Upward and Downward Social Comparisons. *Handbook of Social Comparison*, 173–200. https://doi.org/10.1007/978-1-4615-4237-7_10
- Snyder, H. (2019). Literature review as a research methodology: An overview and guidelines. *Journal of Business Research*, 104, 333–339. <https://doi.org/10.1016/j.jbusres.2019.07.039>
- Sobel, J. (2005). Interdependent Preferences and Reciprocity. *Journal of Economic Literature*, 43(2), 392–436. <https://doi.org/10.1257/0022051054661530>
- Song, H., Tucker, A. L., Murrell, K. L., & Vinson, D. R. (2018). Closing the Productivity Gap: Improving Worker Productivity Through Public Relative Performance Feedback and Validation of Best Practices. *Management Science*, 64(6), 2628–2649. <https://doi.org/10.1287/mnsc.2017.2745>
- Stajkovic, A. D., & Luthans, F. (2001). Differential Effects Of Incentive Motivators On Work Performance. *Academy of Management Journal*, 44(3), 580–590. <https://doi.org/10.2307/3069372>
- Stone, E. R., & Opel, R. B. (2000). Training to Improve Calibration and Discrimination: The Effects of Performance and Environmental Feedback. *Organizational Behavior and Human Decision Processes*, 83(2), 282–309. <https://doi.org/10.1006/obhd.2000.2910>
- Sully De Luque, M. F., & Sommer, S. M. (2000). The Impact of Culture on Feedback-Seeking Behavior: An Integrated Model and Propositions. *Academy of Management Review*, 25(4), 829–849. <https://doi.org/10.5465/amr.2000.3707736>
- Suls, J., & Wheeler, L. (Eds.) (2000). *Handbook of social comparison: Theory and research*. Springer Science & Business Media.
- Tafkov, I. D. (2013). Private and Public Relative Performance Information under Different Compensation Contracts. *The Accounting Review*, 88(1), 327–350. <https://doi.org/10.2308/accr-50292>
- Tallon, P. P., Queiroz, M., Coltman, T., & Sharma, R. (2018). Information technology and the search for organizational agility: A systematic review with future research possibilities. *The Journal of Strategic Information Systems*. <https://doi.org/10.1016/j.jsis.2018.12.002>
- Taylor, S. E., & Fiske, S. T. (1978). Saliency, attention, and attribution: Top of the head phenomena. *Advances in Experimental Social Psychology*, 11, 249–288. [https://doi.org/10.1016/S0065-2601\(08\)60009-X](https://doi.org/10.1016/S0065-2601(08)60009-X)
- Thaler, R. H. (2017a). Behavioral Economics. *Journal of Political Economy*, 125(6), 1799–1805. <https://doi.org/10.1086/694640>
- Thaler, R. H. [@R_Thaler] (2017b, August 11). *Make your research about the world, not the literature* [Tweet]. Twitter. https://twitter.com/R_Thaler/status/895865905564876800?s=20
- The Economist. (2012, November 10). *No More than just a game*. <https://www.economist.com/business/2012/11/10/more-than-just-a-game>

- The Economist. (2016, February 20). *The measure of a man*. <https://www.economist.com/business/2016/02/20/the-measure-of-a-man>
- Thomas, T. F. (2016). Motivating revisions of management accounting systems: An examination of organizational goals and accounting feedback. *Accounting, Organizations and Society*, 53, 1–16. <https://doi.org/10.1016/j.aos.2016.07.001>
- Thornock, T. A. (2016). How the timing of performance feedback impacts individual performance. *Accounting, Organizations and Society*, 55, 1–11. <https://doi.org/10.1016/j.aos.2016.09.002>
- Tiefenbeck, V., Goette, L., Degen, K., Tasic, V., Fleisch, E., Lalive, R., & Staake, T. (2018). Overcoming salience bias: How real-time feedback fosters resource conservation. *Management Science*, 64(3), 1458–1476. <https://doi.org/10.1287/mnsc.2016.2646>
- Tranfield, D., Denyer, D., & Smart, P. (2003). Towards a Methodology for Developing Evidence-Informed Management Knowledge by Means of Systematic Review. *British Journal of Management*, 14(3), 207–222. <https://doi.org/10.1111/1467-8551.00375>
- Tuttle, B., & Burton, F. G. (1999). The effects of a modest incentive on information overload in an investment analysis task. *Accounting, Organizations and Society*, 24(8), 673–687. [https://doi.org/10.1016/S0361-3682\(99\)00017-3](https://doi.org/10.1016/S0361-3682(99)00017-3)
- Tversky, A., & Kahneman, D. (1974). Judgment under Uncertainty: Heuristics and Biases. *Science*, 185(4157), 1124–1131. <https://doi.org/10.1126/science.185.4157.1124>
- Tversky, A., & Kahneman, D. (1983). Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological Review*, 90(4), 293–315. <https://doi.org/10.1037/0033-295X.90.4.293>
- Vancouver, J. B., & Tischner, C. E. (2004). The effect of feedback sign on task performance depends on self-concept discrepancies. *Journal of Applied Psychology*, 89(6), 1092–1098. <https://doi.org/10.1037/0021-9010.89.6.1092>
- VandeWalle, D., Cron, W. L., & Slocum, J. W. (2001). The role of goal orientation following performance feedback. *Journal of Applied Psychology*, 86(4), 629–640. <https://doi.org/10.1037/0021-9010.86.4.629>
- Wang, L. W. (2017). Recognizing the Best: The Productive and Counterproductive Effects of Relative Performance Recognition. *Contemporary Accounting Research*, 34(2), 966–990. <https://doi.org/10.1111/1911-3846.12292>
- Webster, J., & Watson, R. T. (2002). Analyzing the Past to Prepare for the Future: Writing a Literature Review. *MIS Quarterly*, 26(2), xiii–xxiii. <https://doi.org/10.2307/4132319>
- Wegman, L. A., Hoffman, B. J., Carter, N. T., Twenge, J. M., & Guenole, N. (2018). Placing Job Characteristics in Context: Cross-Temporal Meta-Analysis of Changes in Job Characteristics Since 1975. *Journal of Management*, 44(1), 352–386. <https://doi.org/10.1177/0149206316654545>
- White, H. (1980). A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity. *Econometrica*, 48(4), 817–838. <https://doi.org/10.2307/1912934>
- Wozniak, D. (2012). Gender differences in a market with relative performance feedback: Professional tennis players. *Journal of Economic Behavior and Organization*, 83(1), 158–171. <https://doi.org/10.1016/j.jebo.2011.06.020>
- Wurman, R. S. (1990). *Information anxiety. What to do when information doesn't tell you what*

you need to know. Bantam Books.

Wurman, R. S. (2001). *Information anxiety 2*. Macmillan.

Zhong, S., Shalev, I., Koh, D., Ebstein, R. P., & Chew, S. H. (2018). Competitiveness and Stress. *International Economic Review*, 59(3), 1263–1281. <https://doi.org/10.1111/iere.12303>

Zingoni, M., & Byron, K. (2017). How beliefs about the self influence perceptions of negative feedback and subsequent effort and learning. *Organizational Behavior and Human Decision Processes*, 139, 50–62. <https://doi.org/10.1016/j.obhdp.2017.01.007>

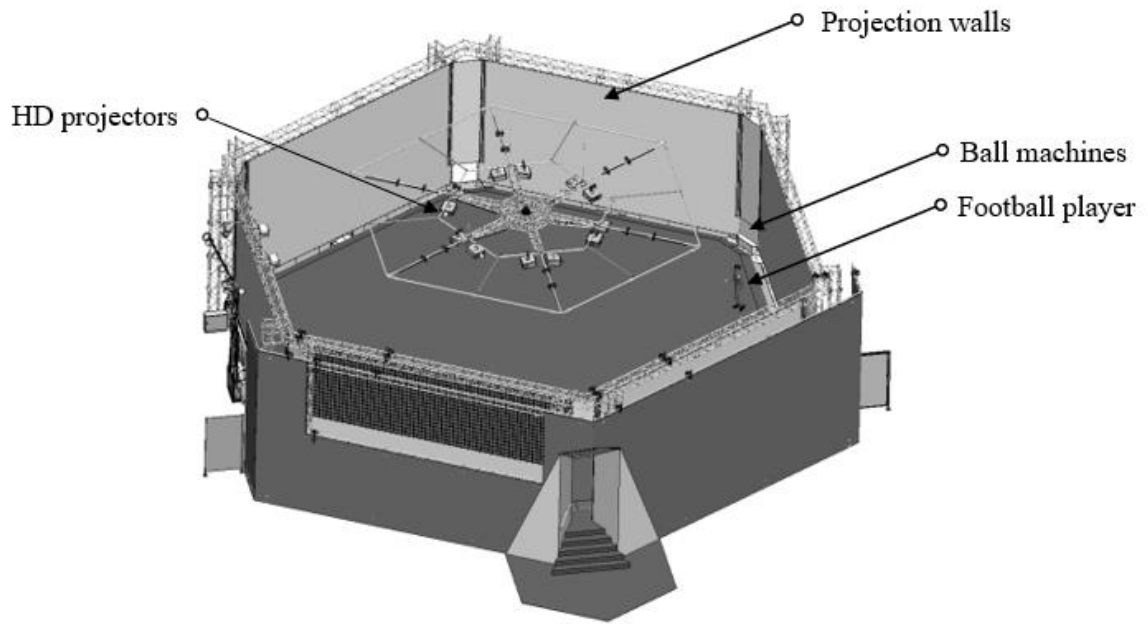
Appendix

Appendix 1. [Introduction] Transparency Table for Data Collected Based on the Pre-Registered Report (Koppers, 2019)

Variables in the complete dataset	Paper II: STATUS = Unpublished working paper currently under review	Paper III: STATUS = Unpublished working paper
Hitrate	x	x
Distance	x	
Average time (or Avg_Time)	x	
Fastest time (or Fast_Time)	x	x
Round duration		x
Heart rate		x
Absolute	x	x
Relative	x	x
Absolute&Relative	x	x
Aggregate	x	
Detail	x	
Too many measures	x	
Social comparison	x	
Perceived value		x
Trust		x
Conscientiousness	x	
CRT		x
Interest	x	
Age	x	x
Gender	x	
Height	x	
Weight	x	
League level	x	
% Games entered	x	x
Tenure	x	
Simulator experience	x	x
Language	x	
Training	x	
Position	x	

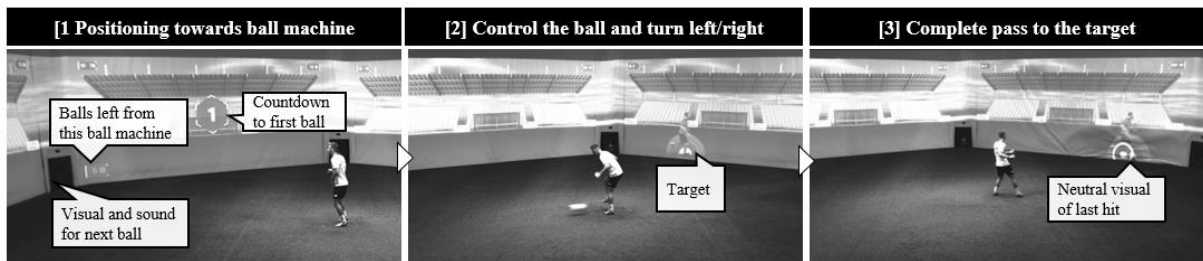
This exhibit lists all performance data from round one and two as well as pre- and post-survey data used within this dissertation from the pre-registered report with the AEA RCT Registry (Koppers, 2019).

Appendix 2. [Paper II] Layout of the Football Training Simulator skills.lab



This exhibit shows the layout of the skills.lab football training simulator. The integrated projectors display the passing target onto the walls of the simulator. One of the four ball machines is shown on the right side of the simulator. For scale, a football player is shown in the right corner of the simulator.

Appendix 3. [Paper II] Real-Effort Task



This exhibit describes in detail the real-effort task “Passes after controlling the ball with a 180° turn”. In the first step of the task, the football player waits for a ball to be passed from one of the simulator’s ball machines. In the second step of the task, the player must control the ball and locate the passing target. In the final step of the task, the player passes the ball to the target.

Appendix 4. [Paper II] Target Visualization in the Real-Effort Task



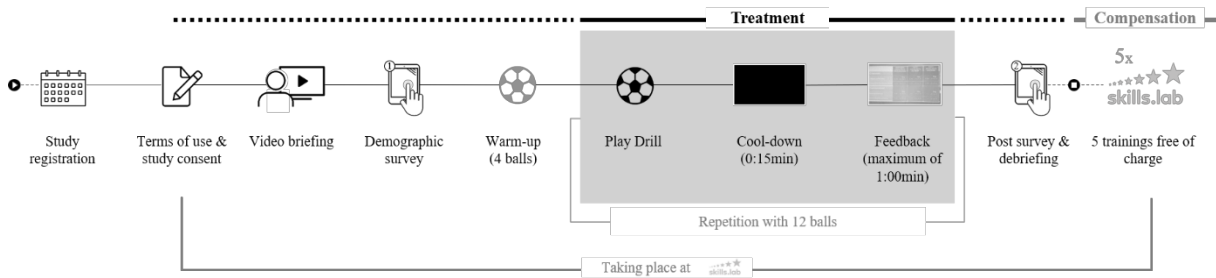
This exhibit shows the moving target in the real-effort task. Users were instructed to aim for the bullseye centered on the foot of the projected player.

Appendix 5. [Paper II] 3x2 Experiment Design

		Number of measures							
		Aggregate				Detail			
Performance Information	<i>Absolute</i>	Zielgenauigkeit	Ergebnis	Zielgenauigkeit	Gesamtergebnis: Auftritten über die linke & rechte Schulter	Ergebnis: Auftritten über die linke Schulter	Ergebnis: Auftritten über die rechte Schulter	Zielgenauigkeit	Ergebnis: Auftritten über die linke & rechte Schulter
	<i>Relative</i>	Zielgenauigkeit	Ergebnis	Zielgenauigkeit	Gesamtergebnis: Auftritten über die linke & rechte Schulter	Ergebnis: Auftritten über die linke Schulter	Ergebnis: Auftritten über die rechte Schulter	Zielgenauigkeit	Ergebnis: Auftritten über die linke & rechte Schulter
	<i>Absolute&Relative</i>	Zielgenauigkeit	Ergebnis	Zielgenauigkeit	Gesamtergebnis: Auftritten über die linke & rechte Schulter	Ergebnis: Auftritten über die linke Schulter	Ergebnis: Auftritten über die rechte Schulter	Zielgenauigkeit	Ergebnis: Auftritten über die linke & rechte Schulter
		Trifferrate (auf Zielscheibe)	88.33 %	Trifferrate (auf Zielscheibe)	88.33 %	90.00 %	86.87 %	Trifferrate (auf Zielscheibe)	88.33 % 19 von 22
		Durchschnittliche Entfernung (zur Zielscheibe)	0.93 m	Durchschnittliche Entfernung (zur Zielscheibe)	0.93 m	0.99 m	0.86 m	Durchschnittliche Entfernung (zur Zielscheibe)	0.93 m 19 von 22
		Handlungsgeschwindigkeit	2.45 s	Handlungsgeschwindigkeit	2.45 s	2.19 s	2.86 s	Handlungsgeschwindigkeit	2.45 s 19 von 22
		Reaktion (Ballverarbeitung)	2.79 s	Reaktion (Ballverarbeitung)	2.79 s	2.48 s	3.00 s	Reaktion (Ballverarbeitung)	2.79 s 9 von 22
		Durchschnittswert (Ballverarbeitung)	2.79 s	Durchschnittswert (Ballverarbeitung)	2.79 s	2.48 s	3.00 s	Durchschnittswert (Ballverarbeitung)	2.79 s 9 von 22
		Trifferrate (auf Zielscheibe)	Rang	Trifferrate (auf Zielscheibe)	Rang	Rang	Rang	Trifferrate (auf Zielscheibe)	88.33 % 19 von 22
		Durchschnittliche Entfernung (zur Zielscheibe)	19 von 22	Durchschnittliche Entfernung (zur Zielscheibe)	19 von 22	22 von 22	11 von 22	Durchschnittliche Entfernung (zur Zielscheibe)	0.93 m 19 von 22
		Handlungsgeschwindigkeit	Rang	Handlungsgeschwindigkeit	Rang	Rang	Rang	Handlungsgeschwindigkeit	2.45 s 19 von 22
		Reaktion (Ballverarbeitung)	19 von 22	Reaktion (Ballverarbeitung)	19 von 22	8 von 22	13 von 22	Reaktion (Ballverarbeitung)	2.79 s 9 von 22
		Durchschnittswert (Ballverarbeitung)	9 von 22	Durchschnittswert (Ballverarbeitung)	9 von 22	7 von 22	10 von 22	Durchschnittswert (Ballverarbeitung)	2.79 s 9 von 22
		Zielgenauigkeit	Ergebnis	Zielgenauigkeit	Gesamtergebnis: Auftritten über die linke & rechte Schulter	Ergebnis: Auftritten über die linke Schulter	Ergebnis: Auftritten über die rechte Schulter	Zielgenauigkeit	Ergebnis: Auftritten über die linke & rechte Schulter
		Trifferrate (auf Zielscheibe)	88.33 % 19 von 22	Trifferrate (auf Zielscheibe)	88.33 % 19 von 22	90.00 % 22 von 22	86.87 % 18 von 22	Trifferrate (auf Zielscheibe)	88.33 % 19 von 22
		Durchschnittliche Entfernung (zur Zielscheibe)	0.93 m 19 von 22	Durchschnittliche Entfernung (zur Zielscheibe)	0.93 m 19 von 22	0.99 m 19 von 22	0.86 m 11 von 22	Durchschnittliche Entfernung (zur Zielscheibe)	0.93 m 19 von 22
		Handlungsgeschwindigkeit	2.45 s 19 von 22	Handlungsgeschwindigkeit	2.45 s 19 von 22	2.19 s 8 von 22	2.86 s 12 von 22	Handlungsgeschwindigkeit	2.45 s 19 von 22
		Reaktion (Ballverarbeitung)	2.79 s 9 von 22	Reaktion (Ballverarbeitung)	2.79 s 9 von 22	2.48 s 7 von 22	3.00 s 10 von 22	Reaktion (Ballverarbeitung)	2.79 s 9 von 22
		Durchschnittswert (Ballverarbeitung)	2.79 s 9 von 22	Durchschnittswert (Ballverarbeitung)	2.79 s 9 von 22	2.48 s 7 von 22	3.00 s 10 von 22	Durchschnittswert (Ballverarbeitung)	2.79 s 9 von 22

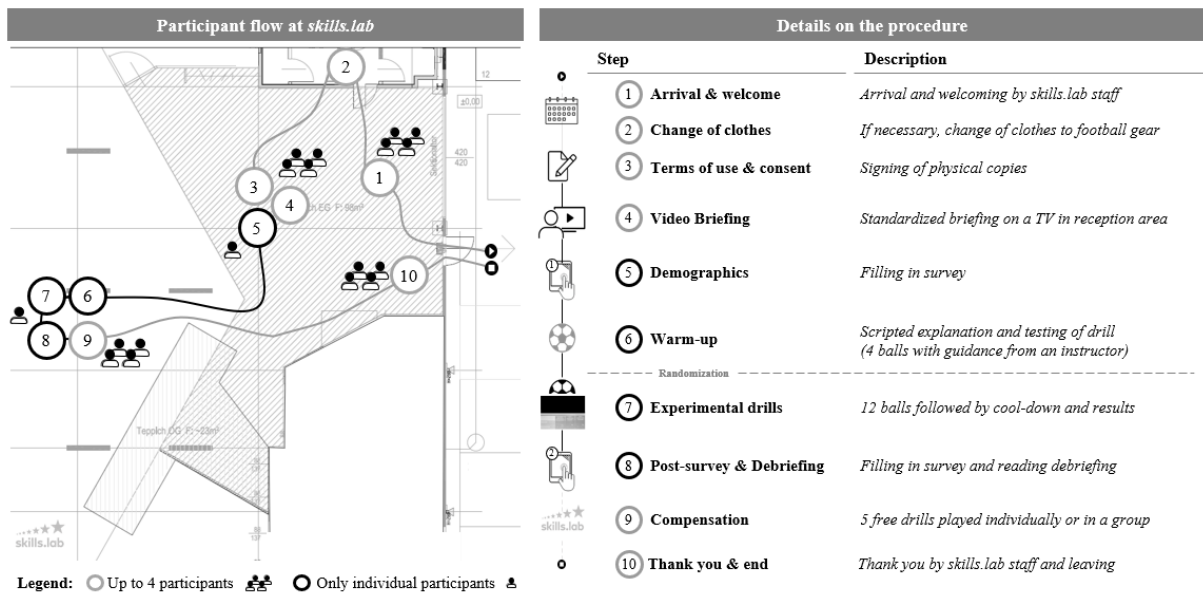
This exhibit shows the fully interacted 3x2 design employed in the experiment. The comparison treatment arm contains three treatments: *Absolute*, *Relative*, or *Absolute&Relative* performance information. The number of measures treatment arm contained two treatments: aggregate or detail.

Appendix 6. [Paper II] On-Site Procedure of the Field Experiment



This exhibit shows in detail the experimental procedure. In the first step of the procedure, participants registered for study participation with Anton Paar SportsTec. In the second step, participants read and signed the terms of use of skills.lab (if they visited skills.lab for the first time) and provided consent. At this step, participants were randomly assigned to one of the six treatment groups. Participants were not informed about the details of the experiment nor the specific treatment to which they were assigned. Participants next watched a standardized video briefing that explained the experimental task and procedure. In the fourth step, participants took part in a demographic survey. This survey was consistent with the regular data entry process employed by the site when collecting information from users. In the fifth step, participants performed a warm-up drill. After this, each participant completed the competitive passing drill, where each player faced a total of 12 balls. This was followed by a short cool-down period after which the player received their performance feedback. This feedback was displayed on to the walls of the simulator. After receiving their feedback, participants again completed the competitive passing drill and received a second round of PF. Once the drills were completed, participants exited the simulator and individually completed the post-experimental survey. Each participant was then debriefed. As compensation for taking part in the study, participants could play five trainings free of charge at skills.lab.

Appendix 7. [Paper II] Participant Flow During the Field Experiment



This exhibit shows the flow of participants through the training facility during the field experiment. The experiment procedure was designed such that participants could not interact with each other between steps 5 and 8 of the experiment when the field site collected demographic information, delivered the intervention, measured performance, and collected post-survey responses.

Appendix 8. [Paper II] Exemplary Performance Feedback Treatments

Absolute x Aggregate treatment:

	Result
🎯 Target accuracy	
Hit rate (on target)	58.33 %
Average distance (to center of the target)	0.93 m
🕒 Speed of execution	
Fastest time (ball handling)	2.45 s
Average time (ball handling)	2.70 s

Relative x Detail treatment:

	Result summary: Turning over the left & right shoulder	Result: Turning over the left shoulder	Result: Turning over the right shoulder
🎯 Target accuracy			
Hit rate (on target)	Place 19 of 32	Place 22 of 32	Place 16 of 32
Average distance (to center of the target)	15 of 32	19 of 32	11 of 32
🕒 Speed of execution			
Fastest time (ball handling)	Place 10 of 32	Place 6 of 32	Place 13 of 32
Average time (ball handling)	9 of 32	7 of 32	10 of 32

Absolute&Relative x Detail treatment:

	Result summary: Turning over the left & right shoulder	Result: Turning over the left shoulder	Result: Turning over the right shoulder
🎯 Target accuracy			
Hit rate (on target)	58.33 % Place 19 of 32	50.00 % Place 22 of 32	66.67 % Place 16 of 32
Average distance (to center of the target)	0.93 m 15 of 32	0.99 m 19 of 32	0.86 m 11 of 32
🕒 Speed of execution			
Fastest time (ball handling)	2.45 s Place 10 of 32	2.10 s Place 6 of 32	2.80 s Place 13 of 32
Average time (ball handling)	2.70 s 9 of 32	2.40 s 7 of 32	3.00 s 10 of 32

This exhibit shows examples for three out of the six treatment conditions in detail. The full six treatment conditions were shown on a smaller scale in Appendix 5. All three examples here are shown in English, while they were provided in German during the study. The *Absolute x Aggregate* PF display features four measures. The *Relative x Detail* PF display contains 12 measures. The *Absolute&Relative x Detail* PF display contains 24 measures.

Appendix 9. [Paper II] Variable Definitions

Variables	Description
Performance measures	
<i>Hitrate</i>	The percentage of the total amount of 12 balls played in one round of the task where a participant has hit the target area.
<i>Distance</i>	The mean distance from a participant's passes to the center of the target for all 12 balls played in one round of the task, measured in meters.
<i>Avg_Time</i>	The mean time that it took a player to complete a pass averaged across all 12 balls played in one round of the task.
<i>Fast_Time</i>	The fastest time that it took a player to complete a pass among the 12 balls played in one round of the task.
Treatment variables	
<i>Absolute</i>	An indicator variable equal to one if the participant is assigned to the treatment group that receives absolute performance information on passing accuracy and speed of execution in the experiment.
<i>Relative</i>	An indicator variable equal to one if the participant is assigned to the treatment group that receives performance ranks relative to a reference group on passing accuracy and speed of execution in the experiment.
<i>Absolute&Relative</i>	An indicator variable equal to one if the participant is assigned to the treatment group that receives both absolute performance information and performance ranks relative to a reference group on passing accuracy and speed of execution in the experiment.
<i>Aggregate</i>	An indicator variable equal to one if the participant is assigned to the treatment group that receives a summary result for all 12 balls played in one round of the real-effort task in the experiment.
<i>Detail</i>	An indicator variable equal to one if the participant is assigned to the treatment group that receives a summary result for all 12 balls and two subcategories (turn to pass left and right) for six balls each played in one round of the task in the experiment.
Post-experiment survey	
<i>Too many measures</i>	Response to survey question, "Based on the training you just played, please answer the following question: I would prefer to see less results." 1 = Strongly disagree; 4 = Neither agree or disagree; 7 = Strongly agree.
<i>Social comparison</i>	Result of a factor analysis of two survey questions (based on Tafkov (2013)). "Please indicate to what extent the following statements describe your experience during the training: 'I often thought about how my performance in the passing drill ranked relative to those of the other participants.'" 1 = Never; 4 = Sometimes; 7 = Very often. "I found that thoughts about performance comparisons interfered with my ability to concentrate on the passing drill." 1 = Not at all; 4 = To a moderate extent; 7 = To a great extent.
<i>Conscientiousness</i>	Mean value for responses to two survey questions from the Big Five Inventory-11 (based on Rammstedt & John (2007)), "I see myself as someone who ... tends to be lazy." (reverse-coded) and "I see myself as someone who ... does a thorough job." 1 = Disagree strongly; 3 = Neither agree nor disagree; 5 = Agree strongly.
<i>Interest</i>	Response to survey question, "Based on the training you just played, please answer the following question: I like to use feedback to help me identify areas that I can improve on." 1 = Strongly disagree; 4 = Neither agree or disagree; 7 = Strongly agree.

Demographic variables

<i>Age</i>	The individual age of the participant at the time of participation.
<i>Gender</i>	An indicator variable equal to one if the participant is female.
<i>Height</i>	Response to survey question, "Please state your current height in centimeters (cm)".
<i>Weight</i>	Response to survey question, "Please state your current weight in kilogram (kg)".
<i>League level</i>	An indicator variable based on the response to a survey question, "What is your current performance level?", equal to one if the participant is currently a professional player in the Bundesliga (performance level I and II in Austria).
<i>% Games entered</i>	Response to survey question, "How many games did you play for your team in the last season (both in the starting line-up and as a substitute)?", slider on a line from 0 % to 100% or field to enter percentage, 0% = No games; 25% = Few games; 50% = Half of all games; 75% = Many games; 100% = All games.
<i>Tenure</i>	Response to survey question, "How many full years have you played for your current club?".
<i>Simulator experience</i>	A categorical variable based on responses to two survey questions, "Have you trained before at skills.lab?", "If yes, how many times did you train at skills.lab?" 0 = 0 times; 1 = 1 time; 2 = 2-5 times; 3 = more than 5 times.
<i>Language</i>	A categorical variable representing the participant's first language. This is gathered using the survey question, "What language do you speak most often at home? Please select the most commonly spoken language."
<i>Training</i>	Response to survey question, "How many days a week do you usually play football (including training and matches)?" 0 = 0 days; 1 = 1 day; 2 = 2 days; 3 = 3 days; 4 = 4 days; 5 = 5 days; 6 = 6 days; 7 = 7 days.
<i>Position</i>	Response to survey question, "Which position do you play most often for your team?" 1 = Goalkeeper; 2 = Defender; 3 = Midfielder; 4 = Striker.
